Rustam Tagiew, Dmitry I. Ignatov, Alexey A. Neznanov, Jonas Poelmans (Eds.)

EEML 2012 – Experimental Economics in Machine Learning

Workshop co-located with the 10th International Conference on Formal Concept Analysis (ICFCA 2012)
May 2012, Leuven, Belgium

Volume Editors

Rustam Tagiew
Institute for Computer Science
Technische Universität Freiberg, Germany

Dmitry I. Ignatov
School of Applied Mathematics and Information Science
National Research University Higher School of Economics, Moscow, Russia

Alexey A. Neznanov
School of Applied Mathematics and Information Science
National Research University Higher School of Economics, Moscow, Russia

Jonas Poelmans
Faculty of Business and Economics
Katholieke Universiteit Leuven, Belgium

Printed in Belgium by the Katholieke Universiteit Leuven with ISBN 978-9-08-140992-6.

The proceedings are also published online on the CEUR-Workshop web site in a series with ISSN 1613-0073.

Preface

In Experimental Economics, laboratory and field experiments are conducted on subjects in order to improve theoretical knowledge about human behavior in interactions. Although paying different amounts of money restricts the preferences of the subjects in experiments, the exclusive application of analytical game theory does not suffice to explain the recorded data. It exacts the development and evaluation of more sophisticated models. In some experiments, human subjects are involved into an interaction with automated agents and these agents are used for simulating human interactions. The more data is used for the evaluation, the more of statistical significance can be achieved. Since huge amounts of behavioral data are required to be scanned for regularities and automated agents are required to simulate and to intervene human interactions, Machine Learning is the tool of choice for the research in Experimental Economics. Moreover modern economics extensively involves network structures, which can be modeled as graphs or more complicated relational structures.

This volume contains the papers presented at the inaugural International Workshop on Experimental Economics and Machine Learning (EEML 2012) held on May 9, 2012 at the Katholieke Universiteit Leuven, Belgium. This year the committee decided to accept 8 full papers for publication in the proceedings and two abstracts for presentation at the conference. Each submission was reviewed by on average 3 program committee members. R. Tagiew proposes a new method for mining determinism in human strategic behavior. N. Buzun et al. present a comparison of methods and measures for overlapping community detection. A. Fishkov et al. discuss a new click model for relevance prediction in Web search. A. Drutsa et al. applied novel data visualisation techniques to socio-semantic network data. Gilabert et al. made an experimental study on the relationship between trust and budgetary slack. O. Barinova et al. proposed using online random forest for interactive image segmentation. A. Bezzubtseva et al. built a new typology of collaboration platform users. V. Zaharchuk et al. proposed a new recommender system for interactive radio network services. D. Ignatov et al. designed a prototype system for collaborative platform data analysis.

We would like to express our gratitude to all contributing authors and reviewers, especially to Malay Bhattacharyya, Hoang Thanh Lam, Olga Barinova and Alexandra Kaminskaya for their enormous efforts. We also want to thank our sponsors Amsterdam-Amstelland police, IBM Belgium, Research Foundation Flanders, Vlerick Management School, OpenConnect Systems and Higher School of Economics.

May, 2012
Leuven

Rustam Tagiew
Dmitry I. Ignatov
Alexey A. Neznanov
Jonas Poelmans

Organization

The inaugural International Workshop on Experimental Economics and Machine Learning (EEML 2012) was held on May 9, 2012 at the Katholieke Universiteit Leuven, Belgium. The workshop was co-located with the 10th International Conference on Formal Concept Analysis (ICFCA-2012).

Program Chairs

Rustam Tagiew	Technische Universität Freiberg, Germany
Dmitry I. Ignatov	National Research University Higher School of Economics, Russia
Alexey A. Neznanov	National Research University Higher School of Economics, Russia
Jonas Poelmans	Katholieke Universiteit Leuven, Belgium

Program Committee

Olga Barinova	Moscow State University, Russia
Elvina Bayburina	National Research University Higher School of Economics, Russia
Malay Bhattacharyya	Indian Statistical Institute, India
Guido Dedene	Katholieke Universiteit Leuven, Belgium
Irina Efimenko	National Research University Higher School of Economics, Russia
Boris Galitsky	University of Girona, Spain
Hoang Thanh Lam	Eindhoven Technical University, The Netherlands
Daniel Karabekyan	National Research University Higher School of Economics, Russia
Aleksandr Karpov	National Research University Higher School of Economics, Russia
Mikhail Khachay	Institute of Mathematics and Mechanics of Russian Academy of Sciences, Russia
Vladimir Khoroshevsky	Dorodnicyn Computer Centre of Russian Academy of Sciences, Russia
Vlado Menkovski	Eindhoven Technical University, The Netherlands
Xenia Naidenova	Military Medical Academy, Russia
Sergey Nikolenko	Steklov Mathematical Institute of Russian Academy of Sciences, Russia
Mykola Pechenizkiy	Eindhoven Technical University, The Netherlands
Artem Revenko	Technische Universität Dresden, Germany

Stijn Viaene	Katholieke Universiteit Leuven, Belgium
Nicola Vitucci	Politecnico di Milano, Italy
Rostislav Yavorsky	Witology, Russia
Leonid Zhukov	National Research University Higher School of Economics, Russia
Sofia Kiselgof	National Research University Higher School of Economics, Russia

Additional Reviewers

Tapas Bhadra	Indian Statistical Institute, India
Saurav Mallik	Indian Statistical Institute, India
Dmitry Zhivotvorev	Yandex and NRU HSE, Russia

Sponsoring Institutions

Amsterdam-Amstelland police, The Netherlands
IBM, Belgium
OpenConnect Systems, USA
Research Foundation Flanders, Belgium
Vlerick Management School, Belgium
National Research University Higher of School Economics, Russia

Table of Contents

Regular Papers

Abstracts

Online Random Forest for Interactive Image Segmentation

Olga Barinova[1], Roman Shapovalov[1], Sergey Sudakov[2], Alexander Velizhev[1]

[1] Lomonosov Moscow State University
{obarinova,shapovalov,avelizhev}@graphics.cs.msu.su
[2] EligoVision Ltd.
svsudakov@gmail.com

Abstract. Many real-world applications require accurate segmentation of images into semantically-meaningful regions. In many cases one needs to obtain accurate segment maps for a large dataset of images that depict objects of certain semantic categories. As current state-of-the art methods for semantic image segmentation do not yet achieve the accuracy required for their use in real-world applications, they are not applicable in this case. The standard solution would be to apply interactive segmentation methods, however their use for a large number of images would be laborious and time-consuming. In this work we present an online learning framework for interactive semantic image segmentation that simplifies processing of such image datasets. This framework learns to recognize and segment user-defined target categories using the ground truth segmentations provided by user. While the user is working on ground truth image segmentation, our framework combines online-learned category models with the standard stroke-propagation mechanisms that are typically used in interactive segmentation methods. Our implementation of this framework in a software system has specific interface features that minimize the required amount of user input. We evaluate the implementation on several datasets from completely different domains (*Sowerby* dataset containing 7 different semantic categories, *sheep & cows* dataset containing 3 categories, and 6 different *flower* datasets with 2 categories each). Usage of our system requires substantially less user effort compared to the traditional interactive segmentation methods.

1 Introduction

Many applications, such as aerial and space image processing, defect detection, and medical imaging require accurate segmentation of large image datasets into some semantically-meaningful zones. Despite the substantial progress made, current state-of-the art methods for automatic semantic image segmentation [1] do not yet achieve the accuracy required for their use in real-world applications. The standard solution to obtain accurate segmentations for a dataset of images would be to apply interactive segmentation methods. Moreover, in some cases interactivity and providing feedback for the computational algorithms to perform segmentation, can not only overcome the inherent difficulties of automatic

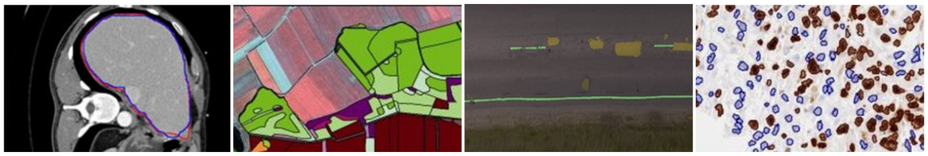


Fig. 1. Image segmentation is used in medical imaging, processing aerial, space images, and detection of road defects. In many applications one needs to accurately segment objects of the some target semantic categories from a large dataset of images. In this work we present a framework that automates this process and substantially reduces the user effort

semantic segmentation, but may also be desirable because the user may want to be able to control the segmentation process and review the results. However applying standard interactive segmentation software for large image datasets would be an extremely laborious and time-consuming task.

In this work we consider the case when one needs to perform segmentation of a large image dataset into a number of semantically-meaningful categories. We aim at developing a general framework that would work with any user-defined target categories and learn these categories from the user input as semantic segmentation methods do. On the other hand, we want to give the user as much control on the segmentation results as interactive image segmentation methods provide. Our main goal is to minimize user effort while allowing her to produce accurate image segmentations. Most existing methods for interactive image segmentation work with a single image [2–4], which limits their power. For example, segmentation of an image from *MSRC* dataset with our system takes less than a minute compared to 15–60 minutes for manual annotation [5].

Related task of inducing segmentation from example was looked at [23]. In this approach a non-parametric model of the provided training pair is constructed by selecting a set of patch-based representatives inside each labeled region in the training image. These representatives are used to quantify the degree of resemblance between small regions in the input image and the labeled regions in the training set.

Adaptive learning of object detection was considered in [6]. The models for new categories may benefit from the detectors built previously for other categories. [7] presented a framework for dynamic visual category learning using incremental support vector machine. That method exploits a previously built classifier to learn the optimal parameters for the current set of training images more efficiently, which is faster than batch retraining. In contrast to those works, we consider a problem of interactive semantic segmentation and our framework enables both adding new categories and incremental learning of the existing ones.

The paper is organized as follows. Next section describes the general workflow of our system and our semantic segmentation algorithm. In Section 3 we describe the online random forest. Section 4 describes the experiments, and the last section is left for conclusions.

2 Interactive Semantic Segmentation Framework

2.1 Interactive Semantic Segmentation Workflow

Suppose one needs to process a large image dataset and obtain accurate segmentation of the objects of certain categories. In framework, a user examines images from the dataset in sequence. Each image is presented as a set of superpixels (Section 2.2). The first image is segmented manually: a user should label each superpixel with one of the category labels. The newly-obtained labelling is used to update the appearance model. When the user opens one of the consequent images, segmentation is performed automatically using the current appearance model. Then the user may correct mistakes of the automatic method by changing superpixel labels. Each time the user approves the (possibly corrected) segmentation result, the system learns from the newly obtained examples of object categories and background. As training goes, user time spent on correction of category map reduces, thus the rate of image labelling increases.

Our **implementation** provides a set of tools to simplify the process of error correction for the user. The brush tool is used to modify superpixel labels. It is possible to change labels of groups of neighbouring superpixels by choosing the appropriate brush size. Each time the user applies it, the system also updates the global labelling using the newly observed labels as context. Therefore, one brush stroke usually changes labels of a large number of pixels. We use hierarchical clustering to obtain superpixels, so the user can switch between different scales of superpixels and choose appropriate scale for correction of errors in the segmentation. We also provide a user with a set of sliders, each one controls the trade-off between false positive and false negative rates of one object category. The sliders help to significantly reduce the amount of manual work in the beginning of the image set processing, when few images have been seen, and the classifiers are likely to be biased towards some categories.

The output of the most time-consuming operations (such as over-segmentation and feature extraction) can be cached, so in practice those operations are performed offline, before the user starts working with the system. We use efficient methods for inference of the optimal segmentation and learning the appearance models of object categories (see Sections 3 and 2.2), thus a user gets immediate response from the system.

2.2 Semantic segmentation algorithm

We obtain pixelwise object segmentation by assigning category labels to a set of superpixels obtained by clustering the joint color and coordinate space with mean-shift algorithm [8]. Usage of superpixels improves computational efficiency as well as makes segmentation more robust. Texture and color features are computed from the image by applying a filter bank [9]. We use texton histograms over superpixels generated by mean shift similarly to [10]. To take the geometric information into account we use simple geometric features like variance, elongation, orientation and area of a superpixel.

We use a simple pairwise conditional random field (CRF) that allows efficient inference. The vector of superpixel labels $\mathbf{c} = \{c_i\}$ is determined as the one that minimizes the following energy function:

$$E(\mathbf{c}) = -\sum_{i} \Psi_i \left(c_i | I\right) - \sum_{(i,j)} \Phi_{ij} \left(c_i, c_j | I\right), \tag{1}$$

where the first term sums the appearance potentials of individual superpixels, the second sum is over the neighbouring pairs of superpixels.

The unary potential for assigning the object category c_i to the i-th superpixel in the image I is computed as $\Psi_i \left(c_i | I\right) = \log p \left(c_i | S_i, I\right) + \eta(c_i)$, where $p \left(c_i | S_i, I\right)$ is the probabilistic output of the online random forests (Section 3) for c_i-th class on the superpixel S_i. The second term $\eta(c_i)$ is the slider value that can be treated as the prior that prefers some categories over the others.

Pairwise potentials consist of the two terms: $\Phi_{ij} \left(c_i, c_j | I\right) = \theta \left(c_i, c_j | I\right) + \tau \left(c_i, c_j\right)$. The first term is the inverse of the boundary strength provided by the mean-shift segmentation. The second term corresponds to a fraction of neighbouring superpixels of the classes c_i and c_j among all neighbouring superpixels in the images seen so far. There are efficient algorithms for minimization of the energy (1), so inference can be performed every time when the user moves a slider to change the trade-off between false positives and false negative rates.

3 Online Random Forest

Our variant of online random forest[1] builds a set of Hoeffding trees [11]. This method is proven to produce the trees asymptotically arbitrarily close to the ones produced by a batch learner. Therefore the incremental nature of our version of Online Random Forest algorithm does not significantly affect the quality of the model it produces.

In Breiman's Random Forest [12] the training set of each tree is obtained by random resampling. This means that the probability that each of N instances is sampled exactly K times for each tree is binomially distributed:

$$p(K = k) = \binom{N}{k} \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{N-k}. \tag{2}$$

As N goes to infinity, the distribution of K converges to the Poisson distribution with the parameter equal to 1: $K \sim \frac{\exp(-1)}{k!}$. Therefore online bootstrapping can be performed as follows: for each base model, choose each example $K \sim \text{Pois}(1)$ times and update the base model accordingly. To diversify the trees in our variant of online random forest each tree operates with a random subset of features.

[1] http://graphics.cs.msu.ru/en/science/research/machinelearning/bolt

```
**Online Random Forest**
**Input:** Example $(x, y)$
For each base model $h_m = h_1, \ldots, h_M$
    Set $k = Poisson(1)$
    Do $k$ times
        $h_m = Update\_tree\,(x, y)$
**Return** updated $\{h_1, \ldots, h_M\}$
```

As long as Hoeffding trees can handle multiclass classification, our Online Random Forest naturally performs multiclass classification without any change in the algorithm.

Handling imbalanced classes. It was proven [13] that error balancing can be achieved by resampling the training set. As the expectation of Poisson random variable $p(K = k) = \frac{\lambda^k}{k!} \exp(-\lambda)$ equals to the parameter λ of Poisson distribution, we can balance the errors by introducing various parameters of Poisson distribution for different classes. In this work we use the balanced version of the Online Random Forest and allow the user to control false positive and false negative rates with the same sliders that were discussed in section .

4 Experiments

Image datasets. In the first experiment we used *Sowerby* dataset that contains 100 images of urban scenes. The goal in this experiment was to perform accurate multi-zone segmentation into 7 object categories provided in the ground truth annotation of this dataset.

In the second experiment we used a subset of the *MSRC* dataset[2] composed of 60 images of cows and sheep. In this experiment we considered a 3-zone segmentation problem where the goal was to segment cows and sheep from background.

In the third experiment we used a subset of *17-flower* dataset[3]. We considered 6 different flowers (daffodil, tigerlily, daisy, fritillary, pansy, sunflower), and 80 images of each flower. The goal in this experiment was to segment each flower from the background.

Measuring usability of the system. The typical sequence of user actions to segment an image in our system is the following. The user starts with tuning the sliders to adjust the false positive vs. false negative rates, and then corrects the segmentation errors using brush tool. In most cases the optimal strategy for error correction is to start with fixing the errors in the coarsest scale of superpixels and then proceed to more detailed scales.

To quantify the user input we have implemented a robot-user that emulates the actions of a human user working with the system. Given the initial image

[2] `http://research.microsoft.com/en-us/projects/ObjectClassRecognition/`

[3] `http://www.robots.ox.ac.uk/~vgg/data/flowers/`

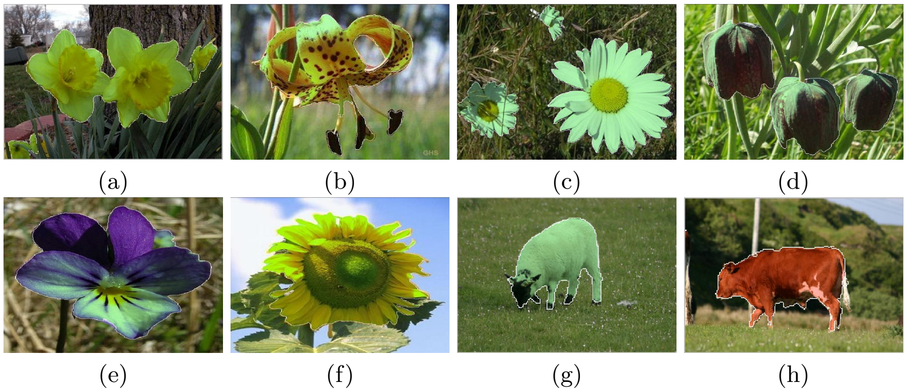


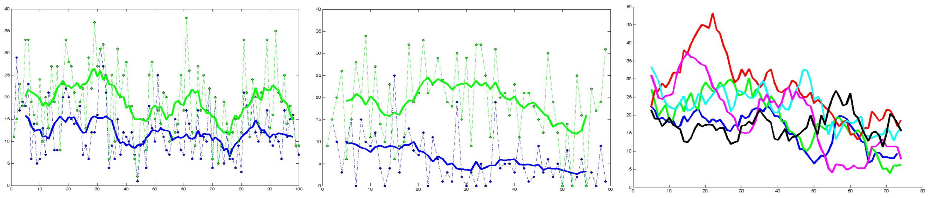
Fig. 2. Example segmentations created using our framework: (a) daffodil, (b) tigerlily, (c) daisy, (d) fritillary, (e) pansy, (f) sunflower; (g),(h) images from our *sheep and cows* dataset. Object is shown with blending, the boundary of an object is marked with white

segmentation, the robot first finds the optimal values of $\eta(c_i)$ for all object classes (i.e. optimal position of the sliders). For that we minimize the total area of misclassified superpixels using Nelder–Mead algorithm. Then we count the number superpixels that need to change labels in order to obtain correct segmentation result. We start by correcting the errors at the coarsest scale and proceed to more detailed scales of superpixels. The resulting metric characterizes overall amount of user input required to obtain correct result using our system, and we refer to it as *usability metric*.

To measure the gain provided by learning the appearance models of object categories, we compared two values of usability metric. First we computed the usability metric for the case of fully manual image segmentation using our brush tool, i.e assuming that all superpixels are initially labelled as background. Second, we computed the usability metric for our semantic segmentation framework.

The results of this experiment for *Sowerby* and *sheep & cows* datasets are shown in Figure 3 (a, b). The green lines show the results in fully manual case, and the blue lines show the results for our framework. The use of automatic segmentation helps to significantly reduce required amount of user input compared to performing fully manual segmentation.

To measure the gain of online learning we looked at the behaviour of the plots of usability metric with respect to the the total number of images processed. As the values of usability metric vary significantly for each particular image, we computed the average over 9 subsequent images to estimate the long-term trends of usability metric. The effect of online learning is most clearly visible for the flowers image datasets (Figure 3 (c)), where the total number of superpixels that require relabelling tends to decrease over time. For *Sowerby* and *sheep & cows* image datasets this metric decreases also decreases, but more slowly.



(a) (b) (c)

Fig. 3. The number of superpixels that have to change their label subject to the number of images processed: (a) *Sowerby* dataset, (b) *sheep & cows* dataset. Performance of our system is shown in blue, total number of clicks required to label an image from scratch is shown in green. Thin lines represent automatically calculated number of superpixels that need to change their labels as described in text, thick lines show the average value over 9 subsequent images. (c) averaged values of usability metric for the *6-flowers* datasets: blue — daffodil, green — fritillary, red — pansy, cyan — sunflower, magenta — tigerlily, black — daisy

Measuring the time. We compared the time required from a human user to obtain high-quality image segmentation with our system and with GrowCut interactive segmentation tool [4] on the *6-flowers* image datasets. The user had practical experience with both systems. The time required for producing high-quality image segmentation for a set of 80 images of the same flower varied from 14 min to 46 min. GrowCut took about twice more time to produce the segmentation of similar quality.

5 Conclusions

We have presented a framework for interactive semantic image segmentation that is based on online learning. The experiments show that online learning of object appearance models helps to significantly reduce user input required to obtain accurate image segmentation.

References

1. Yao, J., Fidler, S., Urtasun, R.: Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation. In: CVPR. (2012)
2. Blake, A., Rother, C., Brown, M., Perez, P., Torr, P.: Interactive image segmentation using an adaptive gmmrf model. In: ECCV. (2004)
3. Grady, L.: Random walks for image segmentatio. Transaction on Pattern Analysis and Machine Intelligence (2006)
4. Vezhnevets, V., Konouchine, V.: "grow-cut" - interactive multi-label n-d image segmentation. In: Graphicon. (2005) 150–156
5. Kohli, P., Ladicky, L., Torr, P.: Robust higher order potentials for enforcing label consistency. In: CVPR. (2008)

6. Opelt, A., Pinz, A., Zisserman, A.: Incremental learning of object detectors using a visual shape alphabet. In: CVPR. (2006)
7. Yeh, T., Darrell, T.: Dynamic visual category learning. In: CVPR. (2008)
8. Paris, S., Durand, F.: A topological approach to hierarchical segmentation using mean shift. In: CVPR. (2007)
9. Shotton, J., Winn, J., Rother, C., Criminisi, A.: Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: ECCV. (2006) 1–15
10. Yang, L., Meer, P., Foran, D.: Multiple class segmentation using a unified framework over mean-shift patches. In: CVPR. (2007)
11. Domingos, P.., Hulten, G.: Mining high-speed data streams. In: Knowledge Discovery and Data Mining. (2000)
12. Breiman, L.: Random forests. Machine Learning Journal **45**(1) (2001) 532
13. Elkan, C..: The foundations of cost-sensitive learning. In: IJCAI. (2001)

A Typology of Collaboration Platform Users

Anastasia Bezzubtseva[1,2], Dmitry Ignatov[1]

[1]Higher School of Economics, Moscow, Russia
[2]Witology, Moscow, Russia

nstbezz@gmail.com, dignatov@hse.ru

Abstract. In this paper we present a review of the existing typologies of Internet service users. We zoom in on social networking services including blogs and crowdsourcing websites. Based on the results of the analysis of the considered typologies obtained by means of FCA we developed a new user typology of a certain class of Internet services, namely a collaboration innovation platform. Cluster analysis of data extracted from the collaboration platform Witology was used to divide more than 500 participants into 6 groups based on 3 activity indicators: idea generation, commenting, and evaluation (assigning marks) The obtained groups and their percentages appear to follow the "90 – 9 – 1" rule.

Keywords. Crowdsourcing, typology classification, collaborative platform, innovation, social network, community, blog.

1 Introduction

Collaboration innovation platforms are relatively young and less common than blogs or social networks (e.g., compare [1] and [2]), yet interest in their organization and audience is not decreasing. The existing studies of consumer or media behavior of Internet users cannot be fully applied to collaboration platform participants, while general psychological or sociological typologies of people miss many important features, inherent only to networking and crowdsourcing.

For a certain type of social network services, i.e. the collaboration innovation platforms, finding user types pursues also some other objectives. Understanding user types could make a major contribution in the platform effectiveness. For instance, dynamic participant type detection and displaying are useful as a motivational game component, and the type itself will probably supplement or refine the exiting rating systems. Also, information about the amount of users of different groups could help platform moderators turn community life to a beneficial for invention direction.

In this study we present a review of the existing Internet service user classifications. Based on examined materials we attempted to develop a new typology of collaboration platform participants using data of one of the projects of Russian innovation platform Witology [10].

2 Terminology

In this paper we analyze not only collaboration platforms, but also all other kinds of social networking services and Internet services, as typologies of their users can be applied to platform participants. There is no fixed terminology in this area yet, but we will try to give some definitions of the important concepts used in the research in order to clarify its subject.

By *Internet service* we mean any website that provides any kind of service (e.g. blogs, file-sharing networks, chats, multiplayer games, online shops). Internet services which provide human interaction are referred to as *Social Networking Services* (SNS). They include social networks (Facebook, MySpace, last.fm, LinkedIn, Orkut), blogs (LiveJournal, Tumblr, Twitter), wiki (e.g., Wikipedia), media hosting sites (Flickr, Picasa, YouTube), etc. [3], [4]. Social networking services often generate *online communities, i.e.* groups of people, who share similar interests and communicate via a certain Internet service. Some scientists [5], [6], [7] understand community in a wider sense as the entire audience of some social networking service, which is wrong, according to Michael Wu [8]. We kept the original author vocabularies when describing the typologies, in other cases the first definition of community was used.

Crowdsourcing platforms are social networking services which are used to obtain the necessary services, ideas or content from platform participants, i.e. platform community, as opposed to regular staff or vendors [9]. *Crowdsourcing (collaboration) innovation platforms* are the ones which focus on idea generation. Activities on collaboration platforms often include message (idea or comment) posting, message reading and message evaluation. The winning solutions and true experts are identified on the basis of the amount and quality of such activities. Work on the platform usually goes as a certain time-limited project, devoted to some company's problem. Witology [10], Imaginatik [11], BrightIdea [12] and some other platforms are organized this way; though, there are many collaboration sites which are not alike (see list [13]).

3 Research objectives

To begin a classification of collaboration innovation platform users, we plan to perform the following tasks:

1. *Study of the existing Internet service user typologies.* The discovered user types and data mining techniques might be helpful in developing another typology.
2. *Developing of a new typology of collaboration innovation platform.* By means of mathematical methods we plan to analyze data of one of the collaboration platform project and identify distinct user types.
3. *Comparison of the obtained percentages with the ones from existing studies.* This might help to understand whether the community under analysis is typical and to find out, whether it can be improved (for example, by calculating community health index [14]).

4 Review of the existing typologies

Despite the fact that the online community being a relatively young phenomenon, tens of attempts in classifying internet users have been undertaken. Some of the studies [15], [16], [17] explore only children's media-behavior, others [18] investigate behavior in terms of online shopping. A significant part of early typologies (e.g. [19], [20]) is developed based on frequency and variety of web and new gadgets use, which resulted in rather trivial and similar typologies (generally people were divided into "advanced", "average" and "non-users", the three types were occasionally interspersed with "entertainment" and "functional" users).

Almost half of the encountered researches used cluster analysis as means of extracting user types, factor analysis appeared to be the second most popular method. Much more uncommon were regression analysis, qualitative in-depth analysis, graph mining, statistical analysis, etc.

Very few authors based on some sociological or psychological theories or referred to the existing typologies when classifying internet service users (it can be explained by their desire to take a new look on the differences in human behavior). One of the studies (Nielsen, 2006) [7] is not only descriptive, but is considered informal, and in spite of that the classification and the "90 – 9 – 1" rule are highly respected and popular.

As for the user typologies of the communities, which organization is close to that of innovation platforms, a notable part of papers is devoted to social network user behavior analysis, but there are also some studies of behavior of blog and forum visitors. Since information concerning behavior of collaboration platform participants has not been found yet, several of social network and blog studies might be interesting and useful as a basis for development of an original classification of collaboration platform users. Further we describe those relevant typologies.

4.1 Describing user typologies

Brandtzæg and Heim (2010). The study [5] is a descriptive one, though the list of existing theories and research papers is given in one of its sections. The results of online survey of 4 Norway social networks users were subjected to cluster analysis.

- *Sporadics* visit social network from time to time, mainly to check if somebody contacted them.
- *Lurkers* is the largest group, they do not create any content, but consume and spread the content created by other groups. They are also notable for a propensity to time-killing.
- *Socializers* use social networks to communicate, make new friends, comment on photos of the old ones, post congratulation messages on walls etc.
- *Debaters* are a more mature and educated version of socializers. Besides communication, less shallow than in the previous case, they are interested in consumption and discussion of news and other information available in social networks.

- *Actives* are engaged with all possible types of activity: communication, reading, creating, watching, establishing groups.

Budak, Agrawal, Abbadi (2010). This paper [13] describes the three types of people (presented in 2002 by Malcolm Gladwell [21]) in terms of graph theory in context of modern online communities (especially blogs). The presence of those people, in Gladwell's opinion, is the main cause of the resounding popularity of some innovations. Authors also introduce a new type (the Translators), which, along with the Sellers, more than other groups influences idea spread and success.

- *Connectors* are people who easily make friends and, thus, have a lot of them.
- *Mavens* are very informed due to their curiosity and like to share their knowledge.
- *Salesmen,* – it is natural for them to convince people and establish an emotional contact with them.
- *Translators* are "bridges" between different interest groups. They have the ability to interpret ideas in a different way, so that more people could understand and accept them.

Li, Bernoff, Fiorentino, and Glass (2007) present another classification [25] without theoretical basis. Groups were extracted with the help of cluster analysis of the poll values.

- *Creators* blog, publish video, maintain their own web-sites; usually belong to the young generation.
- *Critics* select and choose useful media content; typically older than the previous group.
- *Collectors* are known for their addiction to saving bookmarks on special services.
- *Joiners* spend much time in social networks; the youngest group.
- *Spectators* read blogs, watch video, listen to podcasts; main consumers of user-generated content.
- *Inactives* are not active in social services.

Nielsen (2006). In the study [7] it is assumed that active members of large communities are very few. No special mathematical instruments were used to develop the typology, although the author mentions that user activity follows Power law (in the Zipf curve variant).

- *Lurkers* (90%) are those who only consume.
- *Intermittent/sporadic contributors* (9%) are those who contribute rarely, occasionally.
- *Heavy contributors/active participants* (1%) are responsible for up to 90% of community materials.

Jepsen (2006). This is one of the few classifications [23] with a theoretical foundation (Kozinetz, 1999) [22]). The members of Danish newsgroups were classified according to mean and median survey values.

- *Tourists* are not very interested in community content.
- *Minglers* are sociable people, who prefer not to consume the site's content, but to communicate with other members.
- *Devotees* are compared to minglers more interested in newsgroup materials than in communication.
- *Insiders* both communicate and consume information.

Golder and Donath (2004). This is one more descriptive study [24] which examined 16 unmoderated Usenet newsgroups. The taxonomy was built after in-depth analysis of the message posting frequency and message content.

- *Celebrities* are central community figures, contribute more than others.
- *Newbies* are new members, which ask many questions and do not know how to act and communicate appropriately.
- *Lurkers* are those who read discussions, but do not take part in them.
- *Flamers, Trolls, Ranters* – three subgroups, members of which are notable for their negative behavior and love to conversation spoiling.

4.2 Comparing user typologies

Analysis of the mentioned typologies resulted in an assumption that, despite some significant differences in social networking services, there is a universal set of user types. Though, some sources claim that there could be no such a meta-typology [23], when others [6] make attempts in developing one.

The resemblance of user types can be seen more clearly from table 1. Also some insights could be provided by a formal concept lattice, derived from the table (fig. 1). Rows of the table represent the user types described previously (objects), columns are the relevant typologies (attributes). Similar classes were merged: thus, class Actives of the table includes Actives (Brandtzaeg & Heim, 2010), Active participants (Nielsen, 2006), Insiders (Jepsen, 2006), and Celebrities (Golder & Donath, 2004).

Table 1. Formal context (types as objects, typologies as attributes)

	Brandtzaeg & Heim (2010)	*Budak et al. (2010)*	*Li et al. (2007)*	*Nielsen (2006)*	*Jepsen (2006)*	*Golder & Donath (2004)*
Inactives	1	0	1	0	1	0
Lurkers	1	0	1	1	1	1
Socializers	1	1	1	0	1	0
Debators	1	0	1	0	0	0
Actives	1	0	0	1	1	1
Salesmen	0	1	0	0	0	0
Translators	0	1	0	0	0	0

Collectors	0	0	1	0	0	0
Creators	0	1	1	1	0	0
Newbies	0	0	0	0	0	1
Negatives	0	0	0	0	0	1

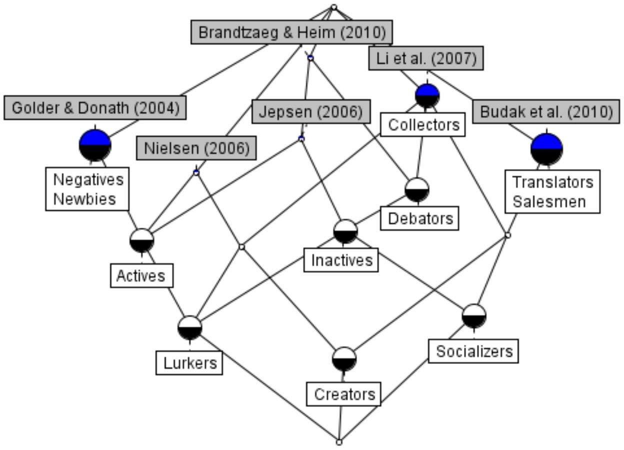


Fig. 1. Formal concept lattice of user typologies (built in ConExp [24])

It can be assumed from the picture that the three general classes of users at the bottom (Lurkers, Creators and Socializers) and, perhaps, two or three important, but less general classes (concepts) above (Actives, Inactives, Debators) form a universal classification of social networking service users. It can also be seen that three studies introduced five original user classes (Negatives, Newbies, Collectors, Translators, Salesmen), which are less likely to be found in a community. As for the typologies, the one of Brandtzaeg & Heim (2010) appears to be the most common.

We built Duquenne-Guigues base for the context and selected the implications with support greater than 4:

1. supp = 4, Actives ==> Lurkers;
2. supp = 3, Inactives ==> Lurkers Socializers;
3. supp = 3, Lurkers Socializers ==> Inactives;
4. supp = 2, Debators ==> Inactives Lurkers Socializers.

E.g., implication 1 can be read as “Each user typology which contains Actives also contains Lurkers and it is valid in 4 cases out of 6”.

5 Typology construction and analysis

5.1 Data sample

We used data obtained in one of the projects [25] of the collaboration platform Witology. It includes quantitative indicators of each of participants' activity: the number of generated ideas, the number of posted comments and the number of submitted evaluations.There were also some other types of activities on the platform, but the mentioned ones are the most basic and easy to interpret.

The project administrators and moderators were not considered as a part of a crowdsourcing community, so only 504 of all 519 registered platform users were sampled.

5.2 Analysis

Initially we detected those participants, who never commented, evaluated or generated ideas. These 248 users were clearly not interested in the project (165 of them never logged on the platform after the third day of its work); thus, they could be excluded from the further analysis.

Then we used clustering algorithm (k-means [26]) to divide the sample based on several parameters. The results of cluster analysis of 256 objects are presented in fig.1 (we used XLSTAT 2011 [27] for the analysis, and XLSTAT-3DPlot package for visualization).

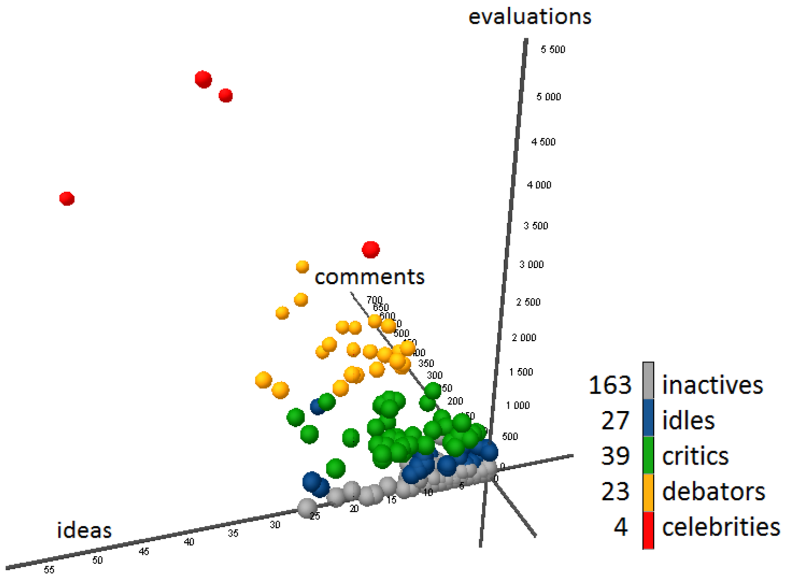


Fig. 2. Sample clustering (the number of objects in each cluster is displayed to the left to the color scale)

The first cluster (grey) represents the participants, who did not show much activity in evaluation and commenting. Because of the difference in orders of numbers of created ideas, comments and evaluations, the participants who seem to be prominent idea generators (created more than 10 ideas) ended up in this group.

The second cluster (blue) differs from the first with slightly higher evaluation activity of users. It can be assumed that those people were interested in project, but lacked motivation for message posting. It is reasonable to merge a certain part of this cluster with the previous one.

The third cluster (green) as a whole is hard to characterize. Its members are less passive: they may skip idea generation or comment posting, but they always evaluate something.

The fourth cluster (yellow) is not far from the previous one in terms of evaluation activity, but the number of comments is quite different.

The last cluster (red) is the smallest one. It consists of four absolute project leaders, who together with some of the yellow participants turned out to be winners or winning ideas authors.

For greater classification veracity the obtained clusters were modified: some of the grey, blue and green balls formed a new class of creators, the rest of the blue joined the grey cluster; also, some minor rearrangements were made.

6 Results

Table 2 represents the resulting user types, their percentages, descriptions and equivalents in other studies.

Table 2. Types of collaboration platform Witology participants

User type	*Number / % of objects*		*Description*	*User types of previous studies*
Celebrities	4	1%	Outstanding users, champions.	Actives [5], mavens [13], active participants [7], insiders [23], celebrities [24]
Debators	21	4%	Those who comment and evaluate actively.	Debators/socializers [5], connectors/salesmen [13], active participants [7], minglers [23]
Creators	20	4%	Idea generators. Could be divided into two groups: energetic creators (6 users), who not only create, and sociopathic ones (14 users),	Mavens [13], creators [25], active participants/sporadic contributors [7], insiders/devotees [23]

			who comment or evaluate many times less.	
Critics	34	7%	Those who evaluate but don't meddle in discussions.	Critics/spectators [25], sporadic contributors [7], lurkers [24]
Tourists	177	35%	Those who rarely make attempts to participate.	Sporadics/lurkers [5], spectators [25], lurkers [7], tourists [23], newbies/lurkers [24]
Inactives	248	49%	Those who do absolutely nothing.	Sporadics/lurkers [5], inactives[25], lurkers [7], tourists [23]

The developed typology and type percentages can be compared with two rather general typologies from the top of the lattice (fig. 1). Table 3 shows how the six classes of this research correspond to their classes.

Table 3. Comparison of different typologies class percentages

Nielsen	*%*	*Brandtzæg*	*%*	*This study*	*%*
Active participants	1%	Actives	18%	Celebrity Debators	5%
Sporadic contributors	9%	Debators Socializers	36%	Creators Critics	11%
Lurkers	90%	Lurkers Sporadics	46%	Tourists Inactives	84%

Interestingly, the percentages in the obtained typology are very close to the ones in Nielsen typology. Brandtzæg explains the discrepancy with the "90 – 9 – 1" rule by a relatively low popularity of Norway social networks compared to YouTube or Wikipedia and by smaller content creation barriers, but such an explanation is not likely to be relevant for the given collaboration project. Nearly 90% of lurkers could be accounted for by initially a small interest of participants to the work itself and a great curiosity to a new for Russia phenomenon, crowdsourcing, as means of some company's growth and development. Other reasons may also take place, but it seems to be difficult to identify them without several projects or platforms comparison.

7 Conclusions

During the process of literature exploration it appeared that there is no generally accepted SNS user classification or any specific collaboration platform participant typology. Based on the existing relevant typologies of social networks, blogs, newsgroups users by means of cluster analysis we developed an original collaboration platform typology. The six classes are so far not expected to be suitable for other crowdsourcing communities. The percentages of classes follow the rule “90 – 9 – 1”, according to which only a minor part of the community is really active.

Thus, all the research objectives were mainly attained.

7.1 Future Work

The developed typology is far from being complete and final. Only a small sample of one of the project was analyzed, while different projects data comparison is expected to specify the classification greatly. Possible future work also includes the following:

- Involving more diverse information on the project (e.g. logs, qualitative values of user evaluations).
- Using other methods (factor analysis, graph mining, mean analysis) of group detection or other clustering algorithms.
- Finding special users (e. g. trolls, flamers, flooders [24]).
- Developing a classification algorithm.
- Testing connection between group membership and demographical factors (age, sex) or psychological tests results.
- Using special metrics to determine community health [14].

Judging by the number of possible work improvement directions it can be concluded that this paper is only a small test sally into the investigation of collaboration platform participants’ behavior, which describes only a static snapshot of one project and does not claim to be indisputable and fundamental.

Acknowledgements. This work was partially done during the mutual research project between Witology and Higher School of Economics (Project and studying group “Data Mining algorithms for analysing Web forums of innovation projects discussion”). We would like to thank Jonas Poelmans for his suggestions for improving the paper.

References

1. Prediction Markets, http://wiki.witology.com/index.php/ Рынки_предсказаний (in Russian)
2. The Growth of Social Media: An Infographic, http://www.searchenginejournal.com/the-growth-of-social-media-an-infographic/32788/
3. Kelsey, T.: Social Networking Spaces: From Facebook to Twitter and Everything In Between. Springer-Verlag, 2010.
4. Boyd, D. M., Ellison, N. B.: Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication, 13 (1) (2007) (http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html)

5. Brandtzæg, P. B., Heim, J. A Typology of Social Networking Sites Users. Interna-tional Journal of Web Based Communities 2011, 7 (1).
6. Brandtzæg, P. B.: Towards a unified media-user typology (MUT): a meta-analysis and review of the research literature on media-user typologies. Computers in Human Behaviour, 26 (5), 2010.
7. Nielsen, J.: Participation Inequality: Encouraging More Users to Contribute. Jakob Nielsen's Alertbox, 9 October 2006 (2006) (http://www.useit.com/alertbox/participation_inequality.html)
8. Community vs. Social Network, http://lithosphere.lithium.com/t5/Building-Community-the-Platform/Community-vs-Social-Network/ba-p/5283
9. Crowdsourcing, http://wiki.witology.com/index.php/Краудсорсинг
10. Witology, http://witology.com/en (in Russian)
11. Imaginatik, http://www.imaginatik.com/
12. BrightIdea, http://www.brightidea.com/
13. Open Innovation Crowdsourcing Examples, http://www.openinnovators.net/list-open-innovation-crowdsourcing-examples/
14. Measuring Community Health for Online Communities. Community Health Index White Paper. Lithium (2011) (http://pages.lithium.com/community-health-index.html)
15. Johnson, G. M., Kulpa, A.: Dimensions of online behavior: Toward a user typology. CyberPsychology & Behavior, 10 (6) (2007)
16. Heim, J., Brandtzæg, P. B., Endestad, T., Kaare, B. H., Torgersen, L.: Children's us-age of media technologies and psychosocial factors. New Media & Society, 9(3) (2007)
17. Livingstone, S., Helsper, E.: Gradations in digital inclusion: Children, young people and the digital divide. New Media & Society, 9(4) (2007)
18. Barnes, S. J., Bauer, H., Neumann, M., and Huber, F.: Segmenting cyberspace: A customer typology for the Internet. European Journal of Marketing, 41(1) (2007)
19. Selwyn, N., Gorard, S., Furlong, J.: Whose Internet is it anyway? Exploring adults (non)use of the internet in everyday life. European Journal of Communication, 20(1) (2005)
20. Heim, J., Brandtzæg, P. B.: Patterns of Media Usage and the Non-Professional Users. In Proc. of the SIGCHI Conference on Human factors in computing systems (CHI 2007), San Jose, California, USA (2007)
21. Gladwell, M. The Tipping Point: How Little Things Can Make a Big Difference. — Back Bay Books (2002)
22. Kozinets, R. V.: E-Tribalized Marketing? The Strategic Implications of Virtual Communities of Consumption. European Management Journal, 17 (3) (1999)
23. Angeletou, S., Rowe, M., Alani, H.: Modelling and Analysis of User Behaviour in Online Communities. In.: Proc. of International Semantic Web Conf. (ISWC 2011), Bonn, Germany (2011)
24. Concept Explorer, http://conexp.sourceforge.net/index.html
25. Sberbank-21, http://sberbank21.ru/
26. K-Means Clustering, http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html
27. XLSTAT, http://www.xlstat.com/en/

Innovative Methods and Measures in Overlapping Community Detection

Nazar Buzun and Anton Korshunov

Institute for System Programming Russian Academy of Sciences
Moscow 109004, Russia
{nazar,korshunov}@ispras.ru
http://www.ispras.ru

Abstract. Cluster structure is one of the main features of social graphs. Many algorithms have been proposed in recent years that are capable of revealing fuzzy communities. But a lot of them tend to degrade in some special cases, for example when nodes assigned to more than two groups. Taking into account that such highly overlapping membership is rather common for many social networks, it becomes obvious that there is a need for flexible techniques and detecting the scope of their effective applicability for various network configuration parameters. This article focuses on the resistance to cluster's growth intersection with emphasis on local fitness function's optimization. The testing of the modern fuzzy clustering methods and generalized classical approaches is performed. Depending on the scale of fuzziness the conclusion is provided about the applicability of certain algorithm classes with common methodology and their representatives.

Keywords: community detection, fuzzy clustering, social networks, social graph mining, local optimization

1 Introduction

Networks are natural representations of various complex systems from society, biology, engineering and other fields. The set of networks is characterized by mesoscopic organisational level inside groups of vertices, which comprise units with a big number of links. Such units are referred to as clusters (or communities or modules).

The universal definition of community partition is stated here only in a qualitative form. It is due to a big variety of formal community detection problem statements and different final goals in particular applications. So far, the problem of partition quality estimation appears to be non-trivial.

In the recent years this research domain has been focused on social and natural networks, whose internal structure cannot be detected by classical clustering algorithms. In these areas analogues of communities are the lists of friends and subscribers, friends circles in Google+ and some social interest groups.

One can figure out several applications of useful information obtained from the network partitioning into communities: system functional units detection;

identification of the community vertices similarity; vertices from a community can be classified in accordance with their position (leaders, linking ones and so on); convenient method of system visualization; vertices' attributes learning on the basis of general attributes of communities which include them. Furthermore, one can specify several methods of machine learning: classification, recommendation, prediction, filtration of non-typical elements (where case division into modular units is a sub-problem). In addition, it's appropriate to mention issues of optimal storage, placement and compression of data; analysis of information distribution; influence inside the global networks.

In spite of the applied problems variety, let us sort out the most general requirements to the methods. Here we also regard some important features of social networks structure.

- The vertex could be found in more than one communities with various degrees of belonging (*fuzzy clusters*) [2-10,24,27]
- Communities may have a *hierarchical structure* [4,6,8,11,22,24] that is required for the efficient management in large-scale organizations, and its presence stresses the stability of the system [12].
- In addition high density of edges doesn't indicate the cluster. Therefore, in order to cut-off "pseudo-communities" a probability of a particular subgraph configuration (*"statistical significance"*) is calculated, under assumption of random edges distribution hypothesis (for the given values of vertices degree) [9,13]. For this purpose it makes sense to look for "significant" subgraphs by taking into account *weak* links [8]. The link (edge) between nodes assumed to be weak if it is not a part of a triangle.
- In some cases (such as for defining attributes of vertices) one need to manipulate vertices and edges with *additional parameters* [1,2,14]. But the majority of current algorithms take only one input parameter like weight of a link.
- While searching for implicit individual user communities (circles of friends, *egomunities*) the execution time and access to graph structure are often limited. Usually in such a case only second friends' neighborhood is known.
- One may also put an additional problem of studying the *community dynamics* [15].

This article focuses on identifying overlapping communities in large networks ($n = 10^8, m \sim n$) with a high coefficient of intersection ($r \sim 10$). Here P is a set of communities $G = (E, V), m = |E|, n = |V|, r = \sum |P_i|/n$. These characteristics are inherent for many real networks. In the Fig.1 one could find more specific communities settings as a vertexes degree function in social Facebook's subgraphs.

We are going to discuss a variety of modern algorithms that is initially characterized by the ability to identify fuzzy communities. In addition, several universal generalizations of classical algorithms will be proposed for the case of graphs with overlapping clusters. The first purpose of this research is to determine of algorithm classes in accordance with their basic ideas. The second aim is to identify the most relevant methods of fuzzy (overlapping) clustering and ways to assess the quality of graph partition.

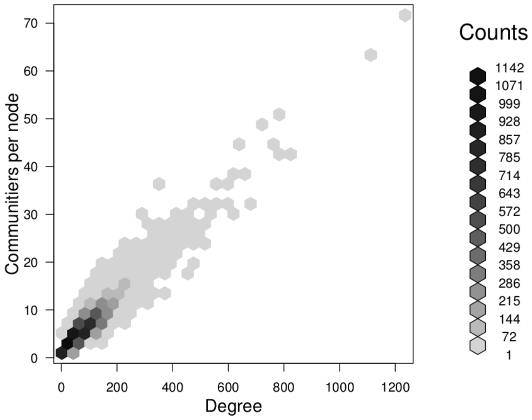


Fig. 1. Social network properties. Coefficient of intersection in Facebook's subgraphs. Shows dependence of communities count that includes node and number of its neighbors. Right column illustrates amount of vertices in the point with certain color.

2 Graph clustering methods overview

Considered clustering methods is divided into classes according to the generality of underlying principles and specification of community definition.

2.1 Null graph model

In methods within this class the given configuration of edges is compared with their uniform distribution for each vertex in graph. For this reason one should use a probabilistic graph model (null model). In its definition the expectation of the nodes degree is fixed. Follows the edge existence probability *(i,j)* is defined as a composition of node's degrees divided by the doubled count of edges: $P_{ij} = \frac{k_i k_j}{2m}$. The classic variant here is to maximize the target **modularity** function and its modifications [16-21], that characterize sum of differences between the total number of edges in the community and its mathematical expectation:

$$Q = \frac{1}{2m} \sum_{c \in P} \sum_{i,j \in c} [A_{ij} - Pr(A_{ij} = 1)],$$

where P - communities set, A - adjacency matrix.

Similarly, instead of edges triangles and more larger cliques could be taken into account. Considering that a link between vertices is weak if it is not a triangle's edge, the community is optimized to increase amount of internal triangles. At the same time the count of adjacent triangles that have exactly two nodes inside the community should be reduced [8,20].

Originally modularity measure was introduced to describe disjoint partitions, but there are some generalizations for the case of overlapped communities [17,21]. Additionally it is worth mentioning its quantum-mechanical modification [18,19], that allows to improve the *resolution limit* and to give it an energetic sense. So it becomes a hamiltonian for a set of particles with various spin values (*spinglass* [19]).

A more common approach is the detection of "significant" clusters. In this case algorithms tend to include in each module those nodes that are most strongly connected to each other. Such cluster type should have a low probability of gathering better interacted users according to random graph model. But due to correlations it is rather complicated to calculate the statistics of the internal connections. Really it is more practical to fix inner community structure and calculate the statistics for the external vertices. This inform us of how much of users for some group are compatible with the null model distribution (*oslom*) [9].

Alternatively one could define probability of link existence to be proportional to the number of communities to which the link belongs (*moses*) [5] and then find the maximum likelihood. This model doesn't account for node degree distribution.So it leads to worse results in some cases but it is rather stable in implementations with high fuzziness.

2.2 Random walks

Here we have three most common methods.

infomap [10,22]: In this case, the clusters are formed to minimize the description length of a random walk in the graph. One of the code length's estimators is entropy that is widely used in various information theory branches. Based on it, [22] propose to consider the following function as a partition quality measure:

$$L(P) = qH(Q) + \sum_i p_i H(P_i),$$

where q - probability that the random walker switches module, p_i- fraction of within module movements, H(Q)- entropy of module names, H(P_i)- entropy of inner module movements including its exit code, i - module number.

walktrap [23]: Here the formation of communities is based on the following proposition: Let the vertices i, j belong to the same cluster, then

$Pr(k \to i, t) \approx Pr(k \to j, t)$ for all $k \in V$, where Pr - transition matrix from a random walk process.

betweenness [9]: Using the measure called "betweenness" on the set of edges (the higher runs count along the edge during a random walk, the greater is measure value). Edges with a high "betweenness" are naturally considered as links between communities (*conga, GN*) [3].

2.3 Local expansion

In a local study and formation of the cluster is generally considered the ratio between the amount of interior edges or triangles and the exterior ones (*cohesion*

[8], GCE [7]). In some approaches link density is additionally compared with its possible maximum. And all such optimizations are usually done disregarding the rest of the graph structure. So the distinguishing feature of this class is an iterative addition of new nodes to the cluster and removal of the existing ones independently of any other clusters. Communities can also be formed on the basis of similarity to a complete graph or a set of connected cliques with different sizes (*CFinder, GCE*) [7]. Besides that, above-mentioned *"statistical significance"* can be used as a local characteristic of similarity between a subgraph and real community. There is also a set of methods in this section which allow independent subgraphs detection to provide high-value influence of vertices within the module (*moduland* [6]). For this class methods a selection of intersecting communities is rather natural, but at the other hand there are some difficulties with the subsequent formation of the final partition in the graph.

2.4 Agent based model

In this case an epidemic process is generated that usually represent a speakers-listener model (*copra* [3], *slpa* [27]). During the execution we should fix a listener node and start gathering information from each of its neighbours. So every such node could save recommendations per each module from received messages. After that it could give an advice to others basing on obtained experience. Here we don't have to define any functional for community, we only spread labels between nodes according to pairwise interaction rules.

There are two types of execution of such epidemic process: synchronous and asynchronous. Synchronous type is more preferable because it prevents monster communities and is easily parallelized. But at the other side it may trigger oscillation phenomenon which should be calmed down with colouring phase (linked nodes get different colours and aren't handled synchronously). .

2.5 R_n metric space

Another elegant approach is to assign coordinates to vertices in the graph [26]. Such coordinates are components of the eigenvectors for the normalized Laplacian matrix L.

$$L_{ij} = \begin{cases} 1, & i = j \\ -\frac{1}{\sqrt{k_i k_j}}, & i - edge - j \\ 0, & else \end{cases}$$

This method of clustering is very useful if one wants to take in account some additional attributes of the vertices.

Summing up the review we can distinguish such methods as spinglass, infomap, wolktrap that have the highest rates of *Normalized Mutual Information* [24] (for the case of **disjoint** communities) with a relatively short execution time and the possibility of parallel execution [25].

3 Generalization methods in the case of overlapping communities

Unfortunately, not all algorithms from the considered classes support fuzzy clusterization. That is why methods of their generalization are required.

3.1 Static

Using the measure of "betweenness" on the set of nodes, one can divide each vertex with high value into two ones connected by an edge. Thus after clustering of the modified graph some user parts could be included into different modules and consequently perform overlapping communities.

The alternative is a generation of line graphs (where edges are turned to vertices and vertices are turned to zero or several edges) and successive edge clustering.

3.2 Dynamic

Introducing membership coefficients for the vertices (which are equal to probabilities of being the member of the particular community), one then assigns a vertex to several classes simultaneously during the algorithm's run. As a first approximation for the membership coefficient one can use the following functions:

- Individual contribution to increase of objective function:

$$Pr(V_i \in P_k) \sim Q(V_i \in P_k) - Q(P_k \backslash V_i) = \triangle Q_{ik}$$

- Probability of being at the particular energy level:

$$Pr(V_i \in P_k) = e^{-\beta Q(V_i \in P_k)} / \sum_S e^{-\beta Q(V_i \in P_S)},$$

where Q - objective function, - value that is inversely proportional to the overlapping coefficient.

It worth noticing that introduction of membership coefficients often improves partition into non-overlapping communities. The main idea here is that by setting probabilities of vertex transition to other communities (staying with some probability in the original one) we let other vertices know about their behaviour tactics. Thereby the following expression can be used in order to set up the coefficients in this case:

$$Pr(V_i \in P_k) \sim \triangle Q_{ik} - \underset{h}{min}(\triangle Q_{ih}),\ V_i \in P_h$$

$$Pr(V_i \in P_{hmax}) \sim 0.1$$

4 Implementation methods

Let us try sort out several implementation ways without binding ourselves to a particular community detection algorithm.

1. Greedy algorithm (used by the majority of the algorithms mentioned above): Originally each vertex is a community itself. Then at the each step of the algorithm every vertex selects the communities to be appended to by comparing relative increases of objective function. A completion phase in this implementation is a clustering of obtained modules, whose unions improve final graph partition.
2. Central vertices: In this method one sets several central users. So the others are gradually attached to them by selecting the closest cluster.
3. Recursive graph partition into two or more parts: In the beginning vertices are randomly partitioned. Then those of them which give the maximal objective function increase are relocated.
4. In the case of local optimization one can recommend to apply the following scheme:

   Single-cluster analysis → Internal structure validation → Clusters consolidation → Membership coefficient computation $\circlearrowright$

   At single-cluster analysis stage each community either gets new nodes or loses those nodes weakly connected with the rest of vertices in the community. So to reach an extremum in this process we should define a function $F(n_{in}, m_{in}, m_{ext}, k_{in})$ depended on internal nodes count, internal and external edges set, links between the community, and the considered node.

   Also, one could use order statistics to work with ranks defined vertex-community closeness. Then to optimize cluster structure one search for a minimum of rank distribution value: $min[F_q(r_q)]$, where q - order number of rank. For this purpose one of null graph models should be chosen (Girvan and Newman [16] or Molloy and Reed [28], for example). If the first stage has a probabilistic character it repeats several times. The final cluster contains those vertices that appear to be included into the group more than fixed times. The considered subgraph is significant cluster if the single-cluster analysis yields a non-empty subgraph in more than definite percent of iterations.

   At the following step of clusters consolidation we may unite some closely located modules or divide them to more small parts. In this case the following measures are usually used:

   common nodes fraction: $\frac{|P_i \cap P_j|}{\min(P_i, P_j)}$

   edges density: $dQ = (\sum A_{ij} - Exp(\sum A_{ij}))$,

   where Q - objective function, P - communities set, A - adjacency matrix.

   Another way is to run a single-cluster analysis on the subgraph of two modules that are to be united or separated.

5 Testing

LFM [1] benchmark algorithm is used as a generator of networks with overlapping cluster structure. In order to investigate the algorithms performance for various degree of community overlapping, two sets of test graphs with predefined partition were generated. The following variables were given to the generator as input parameters: n - number of nodes, k - vertex degree average, k_{max} - maximal value of vertex degree, $|Pi|$ - number of nodes in a cluster, τ_1 - value of the exponent of power law distribution for vertex degree, τ_2 - value of the exponent of power law distribution $|Pi|$, μ - averaged normalized vertex degree inside parent community, on - number of vertices owned by more than one community, om - number of communities containing fixed vertex. Parameters of the graphs from the first community differ by the 'om' value, from the second community - by the 'on' value.

For a comparison of partitions obtained by different methods (Fig 2: fig.1, fig.2, fig.4), let us introduce measure Normalized Mutual Information (I_{norm}) [24] based on the following assumption: if two graph partitions are similar then there is a little information required to obtain the first partition when the second one is known.

$$I(X,Y) = H(X) - H(X|Y)$$

$$I_{norm}(X,Y) = \tfrac{2I(X,Y)}{H(X)+H(Y)},$$

where H - Shannon entropy

In addition to graph partitions from the first set, let us calculate modularity generalized for fuzzy clustering.

From the results, one can conclude that for the case of considerable overlapping only following methods show acceptable performance: **oslom** [9], **moses** [5], **gce** [7], which are representatives of the local optimization class. In particular, first two of them prove the effectiveness of exploiting statistical significance as an individual (local) characteristic of cluster structure. It is also worth to notice the effectiveness of overlapping edge clusterization, that can be applied to the networks of small and moderate size. For the networks of large size with insignificant overlapping one can exploit methods of complexity no more than $O(n^\alpha), \alpha \in [1,2]$ - fuzzy infomap [10], gce [7], spinglass [18] generalization, slpa [27].

Besides this, after analyzing plots of modularity values (Fig 2: fig3), its worth to emphasize a discrepancy of NMI partition quality while increasing 'on'. Therefore, modularity provides impartial partition estimate only if overlapping coefficient 'r' is small.

We also have tested the same methods in local community detection task. Our purpose here was to identify fractions of global network clusters that are the friend circles. We deals only with the second area of the fixed central user, that is

[1] http://sites.google.com/site/andrealancichinetti/files

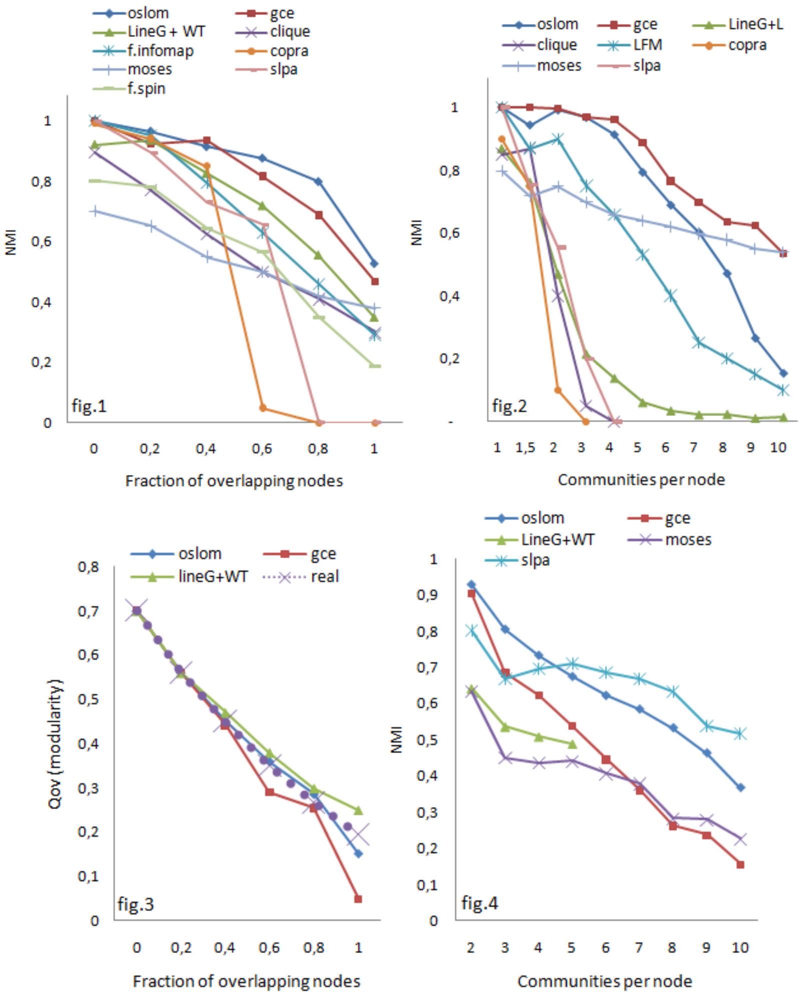


Fig. 2. Testing algorithms of overlapping community detection.
fig.1,fig.3: $n = 2000, k = 15om, kmax = 45om, |P_i| \in [15, 60], \tau_1 = 2, \tau_2 = 0, \mu = 0.2, on \in \{0, 1000, 2000\}$, $om \in \{1, 1.5, 2, 3, ..., 9, 10\}$
fig.2: $n = 1000, k = 20, kmax = 50, |P_i| \in [20, 100], \tau_1 = 2, \tau_2 = 1, \mu = 0.3, on \in \{0, 200, 400, ..., 1000\}$, $om = 2$
fig.4: $n = 4000, k = \max(3om, 10), kmax = 3k, |P_i| \in [20, 80], \tau_1 = 2, \tau_2 = 1, \mu = 0.3, on = 800$, $om \in \{2, 3, ..., 10\}$

the graph information about friends and connections between them. Such kind of local communities (egomunities) could be obtained from the corresponding global ones generated the same way as mentioned above.

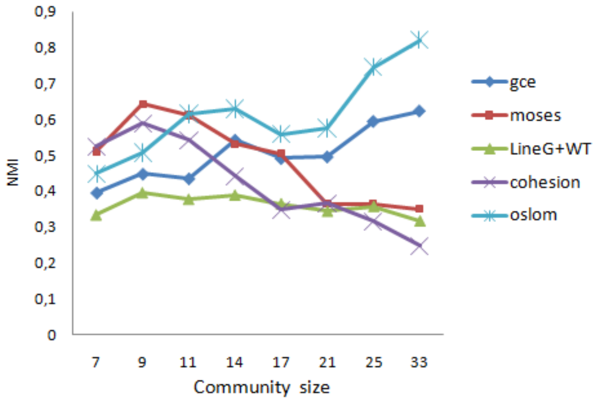


Fig. 3. Testing algorithms of user's second neighborhood egomunity(local community) detection. $n \in [30, 250], |P_i| \in [7, 33], \tau_1 = 2, \tau_2 = 0, \mu = 0.2, om = 6$

In case of local tests attention is drawn to the instability of using "statistical significance" [9] with the small circles of friends. So here in some situations (Fig.3) algorithms that do not use null graph model [16,28] work more efficiently: *cohesion* [8]. On the other hand *moses* [5] utilizing alternative random graph model is quite suitable in such cases. But if the size of the friends neighborhood is rather large the methods similar to *oslom* [9] have higher NMI scores.

6 Conclusion

In summary, several basic features of social and natural graphs were pointed out and algorithms were divided into five classes. Also several different types of their generalization were proposed, and main variants of their implementation were provided. Artificially created networks were used to compare an applicability of the most modern methods. We tested the methods with various network generator parameters. The most effective ones were identified for the particular overlapping coefficient values.

One of the plausible directions of further research is an investigation of weak and strong features of the discussed algorithm classes depending on graph properties and application goals. Herewith all features of social networks mentioned in the beginning of the paper will also be taken into account. In particular, among considerable enough problems are: hierarchical structure detection and methods of its assessment, clustering of graphs with attributes (ordered graphs [14]) on a set of vertices and edges. The last is a task of the highest priority for unknown attributes prediction. Also accumulation of the results of the conducted experiments may possibly result in the development of supervised graph analyser. It will determine at which parts of a graph it would be possible to effectively apply a particular method.

References

1. Lei Tang. 2010. Learning with Large-Scale Social Media Networks. Ph.D. Dissertation. Arizona State University, Tempe, AZ, USA. Advisor(s) Huan Liu. AAI3425805
2. Zhang S, Wang RS, Zhang XS. 2007. Identification of overlapping community structure in complex networks using fuzzy c-means clustering. Physica A 374: 483490.
3. Gregory S. 2007. An algorithm to find overlapping community structure in networks. Berlin, Germany: Springer-Verlag. pp 91102. https://www.cs.bris.ac.uk/~steve
4. Y Ahn, JP Bagrow, S Lehmann. 2010. Link communities reveal multi-scale complexity in networks. Nature 466, 761764.
5. AF McDaid, NJ Hurley. 2010. Using Model-based Overlapping Seed Expansion to detect highly overlapping community structure. In: ASONAM 2010. http://sites.google.com/site/aaronmcdaid/moses
6. Kovacs IA, Palotai R, Szalay MS, Csermely P. 2010. Community landscapes: An integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. PLoS ONE 5: e12528
7. Lee C, Reid F, McDaid A, Hurley N. 2010. Detecting highly overlapping community structure by greedy clique expansion. Poster at KDD 2010.
8. A. Friggeri, G. Chelius, and E. Fleury. 2011. Egomunities, Exploring Socially Cohesive Person-based Communities. NRIA, Research Report RR-7535, 02 2011
9. A. Lancichinetti, F. Radicchi, J. Ramasco, S. Fortunato. 2011. Finding Statistically Significant Communities in Networks. PLoS ONE 6(4): e18961. http://santo.fortunato.googlepages.com/inthepress2
10. AV Esquivel, M Rosvall. 2011. Compression of flow can reveal overlapping modular organization in networks. Phys. Rev. X 1, 021025 (2011). https://sites.google.com/site/alcidesve82

11. Clauset A, Moore C, Newman MEJ. 2008. Hierarchical structure and the prediction of missing links in networks. Nature 453: 98101.
12. Simon H. 1962. The architecture of complexity. Proc Am Phil Soc 106: 467482.
13. Lancichinetti A, Radicchi F, Ramasco JJ. 2010. Statistical significance of communities in networks. Phys Rev E 81: 046110
14. Gregory S. 2011. Ordered community structure in networks. Physica A: Statistical Mechanics and its Applications (December 2011)
15. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela J. 2010. Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. Science 328: 876.
16. M. E. J. Newman. 2006. Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E 74, 036104 (2006)
17. V. Nicosia,G. Mangioni,V. Carchiolo,M. Malgeri. 2008. Extending modularity definition for directed graphs with overlapping communities. J. Stat. Mech. P03024 (2009).
18. J. Reichardt, S. Bornholdt. 2008. Statistical Mechanics of Community Detection. Phys. Rev. E 74 (1) (2006) 016110
19. P. Ronhovde, Z. Nussinov. 2009. Multiresolution community detection for megascale networks by information-based replica correlations. Phys. Rev. E 80 (1) (2009) 016109
20. A. Arenas, A. Fernandez, S. Fortunato, S. Gomez. 2008. Motif-based communities in complex networks. J. Phys. A 41 (22) (2008) 224001.
21. A Lazar, D Abel, T Vicsek. 2009. Modularity Measure of Networks With Overlapping Modules. IOP Publishing, Pages: 18001
22. M Rosvall, CT Bergstrom. 2008. Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci USA 105: 11181123. `http://www.tp.umu.se/~rosvall/code.html`
23. P Pons, M Latapy. 2005. Computing communities in large networks using random walks. Sci. 3733 (2005) 284293.
24. A. Lancichinetti, S. Fortunato, J. Kertesz. 2009. Detecting the overlapping and hierarchical community structure in complex networks. New J. Phys. 11, 033015, 2009
25. Santo Fortunato. 2009 Community detection in graphs. Physics Reports , 486, 75 174
26. L Donetti, M.A. Mutoz. 2004. Detecting network communities: a new systematic and efficient algorithm. J. Stat. Mech. P10012 (2004).
27. Xie, J., Szymanski, B. K., and Liu, X. 2011. Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. In IEEE ICDM 2011 Workshop on DMCCI
28. M. Molloy, B. Reed. 1995. A critical point for random graphs with a given degree sequence. Random Structures and Algorithms, Vol. 6, no. 2 and 3 161-179.

Socio-Semantic Network Data Visualization

Alexey Drutsa[1,2], Konstantin Yavorskiy[1]

[1] Witology
alexey.drutsa@witology.com, konstantin.yavorskiy@witology.com
http://www.witology.com
[2] Moscow State University, Dep. of mech. & math.
http://www.math.msu.su

Abstract. The paper is devoted to some information visualization problems arising in the course of the development of the software package *WitoAnalytics* that enables to analyze and visualize data resulting from the socio-semantic network of the Witology web-platform. The work on the software is in progress. The article contains a short overview of the first software capabilities to visualize some types of socio-semantic network subgraphs.

Keywords: information visualization, data visualization, socio-semantic network, graph

1 Introduction

The Witology company is engaged in solution of some specific real-world problems by constructing active human community, while developing collective mind of participants. In order to achieve the goal a collaborative software platform has been developed and used in the company. Its essential difference from other similar systems consists in direct involvement of specifically trained facilitators in the community. In connection with a possible large number of community members, there is a need of a visual representation of data on their activity on the platform. The data can be used by facilitators as analytical material allowing them to quickly make the right decisions.

At the present time, there is a large number of software designed for analysis and visualization of social networks data (Social Network Analysis, SNA). They include both wide-section programs to analyze all kinds of graphs such as, for example, UCINet [3], Pajek [4] and Cytoscape [5], and programs for the text analysis, for example, Discourse Network Analyzer [6] and AutoMap [7]. Furthermore the class of SNA programs includes specialized software for the analysis of social networks, for example, NodeXL [8], which allows you to retrieve, analyze and visualize data from networks such as Twitter and Facebook. Since the Witology platform is a socio-semantic network [2], then it requires a special analysis software package, adjusted to analysis and visualization of this type of network. Note that the main focus of the research is the scientific field named *information visualization* [9, 10], rather than technological problems of implementation of various methods.

2 Problem statement

In the paper [2] a general model of socio-semantic network is defined as a triplet $\mathbb{G} = (G, C, A)$, such that

- $G = \{V, E_1, \ldots, E_k; \pi, \delta_1, \ldots, \delta_k\}$ is a social network — weighted oriented multi-graph, where V is a set of network members, $E_1, \ldots, E_k \subset V \times V$ are different relations between the members, $\pi : V \to \mathit{\Pi}$ is a *user profile* function and $\delta_i : E_i \to \mathit{\Delta}_i$ $(i \in \{1, \ldots, k\})$ denotes parameters of corresponding relation;
- $C = \{T, R_1, \ldots, R_m; \theta, \gamma_1, \ldots, \gamma_m\}$ is a content multi-graph, where T is a set of all generated content elements (texts, media, evaluations, tags etc), $R_1, \ldots, R_m \subset T \times T$ are different relations between the content elements, $\theta : T \to \mathit{\Theta}$ denotes a function that corresponds to content element parameters and $\gamma_i : R_i \to \mathit{\Gamma}_i$ $(i \in \{1, \ldots, k\})$ denotes parameters of corresponding relation;
- $A \subset V \times T$ is a authorship relation between the social graph and the content.

For such graph analysis the following task is posed: to develop a series of visualizations for the most significant activity of the participants in the platform that would convey the activity in the most informative manner. For instance it could be user evaluations, text generation and etc. Such visualizations should demonstrate both time slices of the database and data change over time.

3 Results

In order to solve the task a specialized software package (hereinafter referred to *WitoAnalytics*) was developed. As mentioned above, the software developed by paper authors could be regarded as one of many SNA softwares, but adjusted to the analysis and visualization of a particular type of graph — socio-semantic network of the Witology platform. The network presented in the current article has more than 500 members and, but the visualizations contains around 200 major network members. At the moment the package allows you to build multiple WitoAnalytics monocot graph visualization and visualization of a bipartite graph.

3.1 User estimation graph

Consider the following oriented weighted subgraph of socio-semantic network: $G_e = \{V_e, E_e, \delta_e\}$, where $\delta_e : E_e \to [-k, k] \times \mathbb{N}$ is a bidimensional edge weight, the first component corresponding to the average value of vertex estimates (in some range $[-k, k]$) and the second component corresponding to the number of the estimates. Hereinafter the subgraph will be called as *user estimation graph.* Such graph could result from user content estimation data taking into account the author relations for the estimations and for the content which is estimated (like texts, etc).

Fig. 1. (A) — The visualization "Elka" (spruce, rus.), (B) — visualization of a local user neighborhood

Fig. 2. Scaled-up part of the visualization "Elka"

The two following visualizations of user estimation graph are proposed. The first is a bipartite representation, where each element from V_e is associated with two nodes situated on a plain, their vertical coordinates being equal. In this case the direction of edges coincides with direction of horizontal axis. The visualization is named as "Elka" (spruce, rus.) and its example is presented in figure 1 (A). Here the edge thickness corresponds to the number of estimates between nodes, and the edge color corresponds to the average value, diagonal edges being marked out with special color. Histograms of out-estimate distribution (on the left) and in-estimate distribution (on the right) are displayed near the nodes. A local user neighborhood of the user estimation graph is presented in figure 1 (B), that is only the edges connected with a fixed user are displayed and the nodes without visible edges are removed. Figure 2 contains a scaled-up part of the visualization "Elka" presented in figure 1 (A).

The second variant of user estimation graph visualization is a monocot representation, where each element from V_e is associated with only one node situated on a circumference. In order to distinguish in-edges and out-edges for a node all the in-edges have same joining angle to the node and all the out-edges have another same joining angle, in-angle and out-angle being not coincided and defining directions, that are symmetric with respect to the radius connected the node. The visualization is named as "Solntse" (sun, rus.) and its example is presented on the figure 3.

Fig. 3. The visualization "Solntse" (sun, rus.)

Fig. 4. Scaled-up part of the visualization "Solntse"

The visualization "Elka" allows us to quickly and accurately provide overall picture of estimations between users, and to identify the nature of evaluations of individual users stood out against a background of other users. Thus, for instance, one can see in figure 1, that all users on average have neutrally estimated each other. At the same time, some nodes stand out among them, their estimates are almost completely negative, or, conversely, are positive. Such users, for example, may be taken under special control by facilitators. In addition, such visualization could be used in order to instantly find a negative evaluation conspiracy of a user group against an individual node. This would be expressed in several broad red lines, leading to one of the nodes in the right column, and other its in-edges on average would not have red color.

Unfortunately, the visualization "Elka" cannot identify so-called "mark up" groups, in which an agreement between users on mutual positive estimation exists. Thus, even a group with two members must be a kind of thick green intersecting edges in the visualization, their symmetry check is quite time-consuming process for a large amount of nodes. To solve this problem the visualization "Solntse" can be very suitable, because in this case incoming and outgoing edge ends of a node coincide.

3.2 Idea support graph

Let's consider a restriction of socio-semantic graph $\bar{\mathbb{G}} = (\bar{G}, \bar{C}, A)$, where content $\bar{C}$ contains only one relation $\bar{R}$, which is strict partial order relation on the set $\bar{T}$, and $\bar{G}$ contains also only one relation $\bar{E}$ induced by the ratio of A as follows:

$$v\bar{E}w \Longleftrightarrow \exists t, \tau \in \bar{T} \mid vA\tau \ \wedge \ wAt \ \wedge \ t\bar{R}\tau \ \wedge \ \tau \in \bar{T}',$$

where $\bar{T}'$ — the set of all maximum elements from $\bar{T}$ relatively $\bar{R}$. Then such subgraph $\bar{\mathbb{G}}$ will be called as *idea support graph.* Idea support graph is visualized by WitoAnalytics as follows. The nodes $\bar{V}$ are allocated on an outer concentric circumference, and the nodes $\bar{T}'$ are allocated on an inner concentric circumference. Size of the nodes and their deviation from the line of the circumference corresponds to the number of edges. The visualization is named as "Glaz" (eye, rus.) and its example is presented in figure 5.

3.3 Short review of current WitoAnalytics capabilities

In the current state WitoAnalytics has the following list of capabilities:

- visualizations of user text estimation (5 types, that include both general view of the graph, and individual user view);
- visualizations of user actions like "content creation", "content evaluation", "content commenting" and etc;
- visualization of user group diversity (dendogram visualizations, adjacency matrix visualizations, histograms and densities);
- valued graph clusterization (3 methods, that include random max-clique search algorithm);
- N-gram and word extracting from user content.

Fig. 5. The visualization ”Glaz” (eye, rus.)

4 Prospect

Since Witology is a relatively young company the work on the analysis and visualization of socio-semantic network data of the platform is the unfinished project, in the framework of which one has to solve many analytical problems and problems of visualization known as information visualization problems [10]. They include the following questions:

- what data to visualize, for example, to detect collusion and ”mark up” groups of participants for many different subgraphs of the platform;
- how to place nodes and edges;
- which thresholds and for which the parameters of nodes and edges should be set.

References

1. Drutsa, A., Yavorskiy, K.: *Visualizatsia dannikh sociosemanticheskoy seti*, Lecture Notes in Computer Science and Information Technologies, National Open University ”INTUIT”, Moscow, 1, (2012) pp.112-118 (in russian).
2. Yavorskiy, R.: *Research Challenges of Dynamic Socio-Semantic Networks*, `http://www.witology.com`.
3. Borgatti, S., Everett, M., Freeman, L.: UCINET, Analytic Technologies, `http://www.analytictech.com/ucinet/`.
4. Pajek, `http://vlado.fmf.uni-lj.si/pub/networks/pajek/`.
5. Cytoscape, `http://www.cytoscape.org/`.

6. Philip Leifeld, Discourse Network Analyzer, http://www.philipleifeld.de/discourse-network-analyzer/.
7. Auto Map, Casos, http://www.casos.cs.cmu.edu/projects/automap/.
8. NodeXL, CodePlex, http://nodexl.codeplex.com/.
9. Apanovich, Z. V.: *Method of information vizualization: scientif field of IT*, Komp'uternie instrumenti v obrazovanii, No 2, (2010) (in russian).
10. Apanovich, Z. V.: *From graph drawing to information visualization*, (preprint) Novosibirsk, 27 p., (2007) (http://www.iis.nsk.su/files/preprints/148.pdf) (in russian).

A New Click Model for Relevance Prediction in Web Search

Alexander Fishkov[1] and Sergey Nikolenko[2,3]

[1] St. Petersburg State Polytechnical University jetsnguns@gmail.com
[2] Steklov Mathematical Institute, St. Petersburg, Russia sergey@logic.pdmi.ras.ru
[3] St. Petersburg Academic University, St. Petersburg, Russia

Abstract. We present a new click model for processing click logs and predicting relevance and appeal for query–document pairs in search results. Our model is a simplified version of the task-centric click model but outperforms it in an experimental comparison.

Keywords: web search, click models, relevance prediction

1 Introduction

Search engines process huge amounts of information: the text of billions of web pages and hyperlinks between them that form the structure of the World Wide Web. Obviously, this information, usually provided by web crawlers, lies in the foundation of a successful search engine [1]. However, as a search engine accumulates active users, information about their behaviour begins to weigh in: *click logs* accumulate first-hand information on user behaviour, i.e., which search results for a certain query users *actually click*. Obviously, the best possible relevance estimates come from the humans themselves; thus, click log information represents an invaluable resource on which search engines would like to draw.

In this work, we propose a new model for processing click logs which is simpler for inference than an existing task-centric click model (TCM) but produces better results. In Section 2, we review existing click models and introduce basic definitions and problem setting; a separate Section 3 is devoted a detailed description of TCM. In Section 4, we present our modified click model and describe the inference procedure. Section 5 describes our experimental setup and results produced on a publicly available large-scale dataset, and Section 6 concludes the paper.

2 Related work

Recent years have seen a proliferation of click models for modeling user behaviour. This line of research began in studying the *position bias* effect: user behaviour studies have shown [2] that not only higher positions in search results rankings attract more attention and are more likely to be clicked on, but also that lower positions are often not even examined at all by the user. Ensuing probabilistic modeling confirmed these results and formalized them in the *examination hypothesis* [3] that captures this reasoning by specifying probabilities of the event C_i that the user clicks on document at position i as conditional probabilities on the event E_i that the user actually examines the document at position i; the examination hypothesis states that $p(C_i = 1 \mid E_i = 0) = 0$.

Latest research has built upon this assumption and has incorporated various additional assumptions and new pieces of information that could be used to predict the click event. Early models tried to capture position bias directly: the *clicks over expected clicks* model [4] estimates the number of expected clicks for each position, the *examination model* learns position bias with an EM algorithm [5], and logistic regression has also been used to estimate position biases [3].

However, it is actually true that in a good search engine, top results are generally more relevant than bottom results, so position bias is not just a feature of the user's perception as position models presuppose but also has sound underlying causes. Thus, emphasis shifted to more complex probabilistic graphical models that attempt to more accurately model actual user behaviour. They are usually based on the *cascade hypothesis* [3]: a user examines documents from top to bottom, so a document at position $i + 1$ can be examined only if the document at position i has been examined: $p(E_{i+1} = 1 \mid E_i = 0) = 0$. A notable model that does not use the cascade hypothesis is the *user browsing model* (UBM) proposed in [5]. UBM assumes that the user "jumps" from the previously clicked position i_c to one of the subsequent positions i with constant probabilities: $p(E_i = 1 \mid C_{i_c}) = \beta_{i_c, i-i_c}$.

Several graphical click models with varying complexity have been proposed under the cascade hypothesis [6–13]. Starting from the *dynamic Bayesian network* (DBN) model [6], click models usually draw a distinction between *appeal* and *relevance*, or, in terms of [6], *appeal*, *perceived relevance* and *intrinsic relevance*. Appeal shows how relevant the document looks for the user; it is directly responsible for user clicks. Perceived relevance shows how relevant the user has felt the document to be after the user has clicked on it and looked at it; perceived relevance is responsible for user satisfaction and, therefore, for the fact whether the user comes back and examines subsequent documents after this one. Intrinsic relevance is usually an auxiliary feature derived from appeal and perceived relevance: appeal and perceived relevance are usually normalized to lie between 0 and 1, and intrinsic relevance is computed as their product.

One of the latest click models is the *task-centric click model* (TCM) proposed by Zhang et al. [7]. TCM steps back and considers whole *sessions* of queries submitted by the same user, assuming that the user has a certain purpose in mind (hence *task-centric*), and various queries are intended to carry out that purpose. Our model, presented in the next section, is close to TCM in essence but turns out to be better in experimental studies and simpler for inference.

3 The TCM model

The main characteristic feature of the task-centric click model is a broader look at search process: interaction between user and search engine is viewed as a sequential process of submitting and reformulating queries. TCM assumes that the user has a specific informational need, a search intent which is assumed to be fixed during the entire session. The user enters a query, examines the result, and then decides whether to click on some documents or enter another query and so on. There are two main assumptions about user behavior in TCM:

(1) if a query does not match the user's underlying intent, he will perform no clicks but learn from search results to pose a new, refined query;
(2) when a document has been examined before in the same session, it will have a lower probability to be clicked when the user examines it again.

For the i^{th} query in a session and for the j^{th} document in the search results, TCM introduces the following variables:

$$\begin{aligned}
M_i&, \text{ whether the } i^{\text{th}} \text{ query matches the user's intent;}\\
N_i&, \text{ whether the user submits another query after the } i^{\text{th}}(\textit{observed});\\
E_{i,j}&, \text{ whether the user examines the document at } (i,j);\\
H_{i,j}&, \text{ whether the document at } (i,j) \text{ has already been shown}\\
&\quad \text{during the current session ;}\\
F_{i,j}&, \text{ whether the document is considered fresh by the user;}\\
C_{i,j}&, \text{ whether the document is clicked } (\textit{observed});\\
R_{i,j}&, \text{ whether the document is relevant;}\\
(i',j')&, \text{ previous position of document at } (i,j) \text{ if this document}\\
&\quad \text{has already been shown during the current session .}
\end{aligned}$$

The following formulas complete the definition of TCM:

$$\begin{aligned}
p(M_i = 1) &= \alpha_1, & p(R_{i,j} = 1) &= r_{i,j},\\
p(N_i = 1|M_i = 1) &= \alpha_2, & p(E_{i,j} = 1) &= \beta_j,\\
p(N_i = 1|M_i = 0) &= 1, & H_{i,j} = 0 &\Leftrightarrow H_{i',j'} = E_{i',j'} = 0,\\
p(F_{i,j} = 1|H_{i,j} = 1) &= \alpha_3, & C_{i,j} = 1 &\Leftrightarrow M_i = E_{i,j} = A_{i,j} = F_{i,j} = 1,\\
p(F_{i,j} = 1|H_{i,j} = 0) &= 1.
\end{aligned}$$

The TCM model is presented as a Bayesian network on Fig. 1. Variables F and H represent the second assumption on user behavior: the probability to click on the current dicument is affected by probabilities of its previous examinations. Looking back at the formulas, one can find out that additional edges are added to the network based on what documents are shown to the user on each

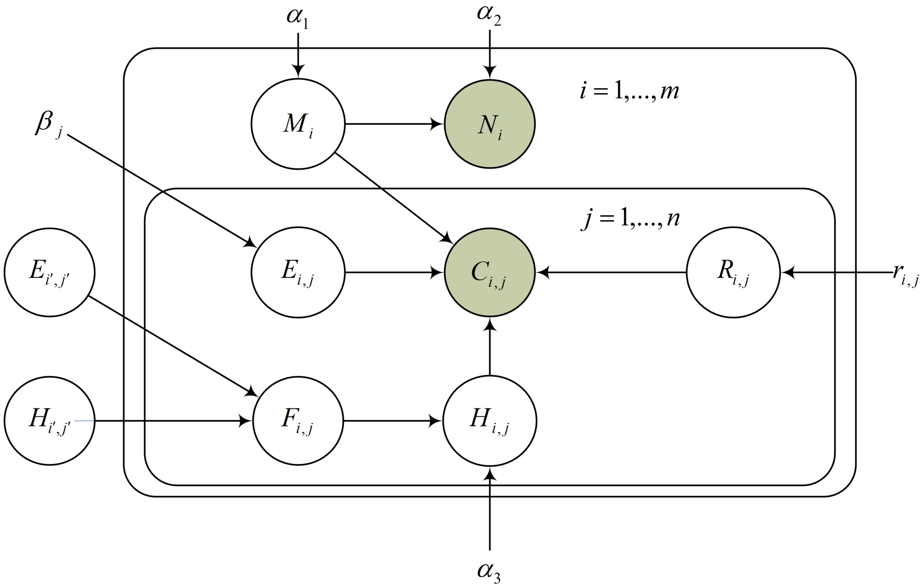


Fig. 1. The TCM model. Shaded nodes represent observed variables.

query result page. Our experiments have shown that large sessions with similar queries (some documents are shown many times during session) significantly slow down inference. It would be much better for the model's efficiency to have each query self-contained. Another way to improve the model is to introduce a more complete query-level structure. We tried to adress these points in our model presented in the next section.

4 The SCM model

In this work, we propose a new model that we call the *session click model* (SCM). It is essentially a simplification of TCM: we break down some of the connections in the factor graph of TCM in order to ease and speed up Bayesian inference. However, as we will see below, our model actually outperforms TCM on real-world data.

For the i^{th} query in a session and for the j^{th} document in the search results, SCM introduces the following variables:

M_i, whether the i^{th} query matches the user's intent;

N_i, whether the user submits another query after the i^{th}(*observed*);

$E_{i,j}$, whether the user examines the document at (i, j);

$H_{i,j}$, whether the document at (i, j) has already been shown during the current session (*observed*);

$F_{i,j}$, whether the document is considered fresh by the user;

$C_{i,j}$, whether the document is clicked (*observed*);

$A_{i,j}$, whether the document appeals to the user;

$S_{i,j}$, whether the document satisfies the user.

Unlike TCM, our model uses $H_{i,j}$ as indication of prior appearance of a document so it becomes a new observed variable. In TCM, these variables make up additional connections between different queries in a session; in SCM, they are observed so different queries become dependent only via M_i.

Formally, SCM is defined as follows (we write some conditional probabilities as logical formulas for brevity and clarity):

$$\begin{array}{rl}
p(M_i = 1) = \alpha_1, & E_{i,j-1} = 0 \Rightarrow E_{i,j} = 0, \\
p(N_i = 1|M_i = 1) = \alpha_2, & S_{i,j-1} = 1 \Rightarrow E_{i,j} = 0, \\
p(N_i = 1|M_i = 0) = 1, & E_{i,j} = 1 \Rightarrow E_{i,j-1} = 1 \,\&\, S_{i,j-1} = 0, \\
p(F_{i,j} = 1|H_{i,j} = 1) = \alpha_3, & p(S_{i,j} = 1 \mid C_{i,j} = 1) = s_{i,j}, \\
p(F_{i,j} = 1|H_{i,j} = 0) = 1, & p(S_{i,j} = 1 \mid C_{i,j} = 0) = 0, \\
p(A_{i,j} = 1) = a_{i,j}, & C_{i,j} = 1 \Leftrightarrow M_i = E_{i,j} = A_{i,j} = F_{i,j} = 1.
\end{array}$$

As a Bayesian network, the model is presented on Fig. 2; the conditional probability tables are shown above, so Fig. 2 together with the above formulas represent a complete specification of the joint probability distribution in SCM. We perform Bayesian inference in SCM via loopy belief propagation. Our model has three global parameters α_1, α_2, α_3. They represent various conditional probabilities. To estimate their values from data, we first convert our model in the form of a factor graph. For each parameter, we add another variable node and connect it to the corresponding factor. Then we iterate through the click log as follows:

(1) assign a uniform Beta prior for each α_i;
(2) process a single session from click log and get posteriors for α_1, α_2, and α_3;
(3) for the next session, set priors for them to posteriors from previous session;
(4) return to step 2.

This would be a very lengthy process for the entire click log, but posterior estimates do converge relatively quickly, so we can stop when the variance becomes small enough. Then we use these estimates as parameter values for the SCM.

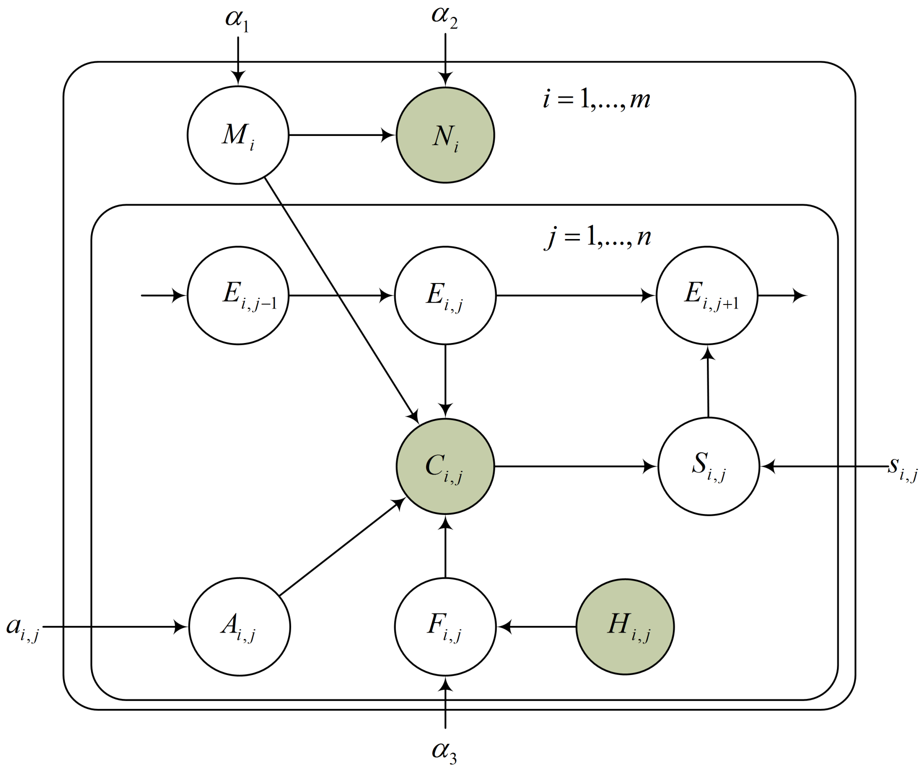


Fig. 2. The SCM model. Shaded nodes represent observed variables.

5 Experiments

5.1 Dataset

We evaluate our results on (a representative subset of) the Yandex click log dataset that was made available for the "Internet Mathematics" competition of 2011 [14]. The data is divided into user sessions that consist of queries, search results, and user clicks for these results; the logs are anonymized, and no user information is provided (we do not know which sessions come from the same user). Click logs also contain time delays between clicks, but neither our model nor any of the competitive click models we compare it with makes use of temporal information; this is a very interesting subject for further study.

5.2 Experimental setup

In general, to evaluate the results we use the *area under curve* (AUC) metric [15] computed on a test set with relevances evaluated by experts. AUC is a popular quality metric for classifiers; it represents the probability that for a uniformly selected pair consisting of a relevant and an irrelevant document the classifier ranks the relevant one higher. Thus, the optimal AUC is 1 (all relevant documents come before irrelevant ones), and a completely random classifier will get, on average, an AUC of 0.5.

However, this is not the whole story. While they do aim to capture user behaviour, click models cannot produce cutting edge results by themselves. To get reasonable prediction accuracy (i.e., a competitive AUC score), click log analysis must also take into account other features that can be inferred from the click logs. For example, one strikingly useful feature is the actual ranking of a document in the original search results: the search engine has processed a lot of additional information which is not available from click logs, and this information has been succinctly represented in the search results rankings, so why not use it. There are many other important features, too.

To simulate this real-world application of click models, we set up our experiments as follows. We have computed 60 *static features* for every query-document pair; these features comprise the base set, and in our experiments, they are augmented by various *dynamic features* that come from click models. We have implemented click models as probabilistic graphical models in the Infer.NET framework [16]; the Infer.NET suite provides readily available inference algorithms, including loopy belief propagation.

To combine all features, static and dynamic, we have used boosting techniques. Following [17, 18], we construct a ranking function by transforming the data into pairwise preferences and considering the resulting problem as a regression problem. To do so, we break every list of search results in the training set into pairwise comparisons between documents. Every ordered pair of documents specified by their features, f_1 and f_2, is represented by the concatenation of their feature vectors in the same order, $\langle f_1, f_2 \rangle$. Then, the target value is chosen to be 1 if the first document is relevant and the second one is not, -1 in the opposite case, and 0 if they are incomparable (both relevant or both irrelevant). These feature vectors together with their training values are fed into a regression boosting algorithm.

In static features, we aimed for simplicity; besides, we did not use the best known boosting techniques [19–21], again choosing a readily available Matlab implementation of least squares regression boosting [22] for simplicity (the learning procedure is stochastic, so we have run it five times and averaged the results). Thus, even the best of our results do not match the top AUC scores obtained in the “Internet Mathematics” competition. For a report of the winners see [23]; the winners did not invent new click models but did everything right with rank boosting and feature generation for other features; see also [24] for a report of a team who used random forests rather than boosting techniques. Nevertheless, we believe that our results do provide a fair comparison in a situation representative of real-life applications.

Model	Features	AUC
DBN	appeal	0.6252
	perceived relevance	**0.6254**
	intrinsic relevance	0.6253
	without static features	0.5944
TCM	appeal	0.6255
	perceived relevance	0.6278
	intrinsic relevance	**0.6279**
	without static features	0.5963
SCM	appeal	0.6265
	perceived relevance	0.6294
	intrinsic relevance	**0.6296**
	without static features	0.5964
All models and features together		**0.6313**

Table 1. Experimental results.

5.3 Results

The results are summarized in Table 1. Each click model in our comparison provides three features: appeal, perceived relevance, and intrinsic relevance that we compute as the product of appeal and perceived relevance (an idea first presented in [6]). In SCM, appeal is estimated as the maximum a posteriori estimate of $a_{i,j}$, perceived relevance is estimated as the maximum a posteriori estimate of $s_{i,j}$, and intrinsic relevance is computed as $a_{i,j}s_{i,j}$. We provide results for the ranking resulting from the three features from a single dynamic model alone, without static features, and results of least squares boosting learning on static features together with each of the three dynamic features from a certain model. As we can see, SCM outperforms both DBN and TCM in terms of AUC, both with static features and without them (although different variables come out ahead in different models).

In the last row of Table 1, we also provide the results for all static and all dynamic features from all three models thrown together. Improved AUC suggests that DBN and TCM do capture some aspects of user behaviour that SCM does not; in further work, we plan to investigate this further and bring other sides of user behaviour into our model, hopefully still leaving the model relatively simple.

6 Conclusion

In this work, we have proposed a new click log model which is in essence a simplification of the task-centric model but has outperformed it in our experiments. Further work may include extending the model to capture more different aspects of user behaviour (e.g., distinguishing between navigational and informational queries) and devising a large-scale highly parallel implementation of our click model.

Acknowledgements

We thank Yandex employees Igor Kuralenok and Andrey Strelkovskiy for fruitful discussions. This project has been supported by the Russian Fund for Basic Research grant 12-01-00450-a and by the EMC student grant. Work of the second author has also been supported by the Russian Presidential Grant Programme for Young Ph.D.'s, grant no. MK-6628.2012.1, for Leading Scientific Schools, grant no. NSh-3229.2012.1, and RFBR grants 11-01-12135-ofi-m-2011 and 11-01-00760-a.

References

1. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems **30**(1–7) (1998) 107–117
2. Granka, L.A., Joachims, T., Gay, G.: Eye-tracking analysis of user behavior in WWW search. In: Proceedings of the 27th Annual ACM SIGIR Conference. (2004) 478–479
3. Craswell, N., Zoeter, O., Taylor, M., Ramsey, B.: An experimental comparison of click position-bias models. In: Proceedings of the 1st ACM International Conference on Web Search and Data Mining. (2008) 87–94
4. Zhang, V., Jones, R.: Comparing click logs and editorial labels for training query rewriting. In: Query Log Analysis: Social And Technological Challenges, 16th WWW Conference workshop. (2007)
5. Dupret, G., Piwowarski, B.: A user browsing model to predict search engine click data from past observations. In: Proceedings of the 31st Annual ACM SIGIR Conference. (2008) 331–338
6. Chapelle, O., Zhang, Y.: A dynamic Bayesian network click model for web search ranking. In: Proceedings of the 18th International Conference on World Wide Web. (2009) 1–10
7. Zhang, Y., Chen, W., Wang, D., Yang, Q.: User-click modeling for understanding and predicting search-behavior. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA (2011) 1388–1396
8. Guo, F., Liu, C., Kannan, A., Minka, T., Taylor, M., Wan, Y., Faloutsos, C.: Click chain model in web search. In: Proceedings of the 18th International Conference on World Wide Web. (2009) 11–20
9. Hu, B., Zhang, Y., Chen, W., Wang, G., Yang, Q.: Characterize search intent diversity into click models. In: Proceedings of the 20th International Conference on World Wide Web. (2011) 17–26
10. Srikant, R., Basu, S., Wang, N., Pregibon, D.: User browsing models: relevance versus examination. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. (2010) 223–232
11. Zhu, Z., Chen, W., Minka, T., Zhu, C., Chen, C.: A novel click model and its applications to online advertising. In: Proceedings of the 3rd ACM International Conference on Web Search and Data Mining. (2010) 321–330
12. Zhang, Y., Wang, D., Wang, G., Chen, W., Zhang, Z., Hu, B., Zhang, L.: Learning click models via probit Bayesian inference. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management. (2010) 439–448

13. Dupret, G., Liao, C.: A model to estimate intrinsic document relevance from the clickthrough logs of a web search engine. In: Proceedings of the 3rd ACM International Conference on Web Search and Data Mining. (2010) 181–190
14. Yandex: Yandex Internet Mathematics competition. `http://imat-relpred.yandex.ru/` (2011)
15. Ling, C.X., Huang, J., Zhang, H.: Auc: a statistically consistent and more discriminating measure than accuracy. In: Proceedings of the International Joint Conference on Artificial Intelligence 2003. (2003) 519–526
16. Minka, T., Winn, J., Guiver, J., Knowles, D.: Infer.NET 2.4 (2010) Microsoft Research Cambridge. http://research.microsoft.com/infernet.
17. Zheng, Z., Chen, K., Sun, G., Zha, H.: A regression framework for learning ranking functions using relative relevance judgments. In: Proceedings of the 30th Annual ACM SIGIR Conference, ACM (2007) 287–294
18. Hastie, T., Tibshirani, R., Friedman, J.: Elements of Statistical Learning. Springer, New York (2008)
19. Donmez, P., Svore, K.M., Burges, C.J.C.: On the local optimality of lambdarank. In: Proceedings of the 32nd Annual ACM SIGIR Conference, ACM (2009) 460–467
20. Burges, C.J.C.: From RankNet to LambdaRank to LambdaMART : An overview. Technical report, Microsoft Research (2010)
21. Burges, C.J.C., Svore, K.M., Bennett, P.N., Pastusiak, A., Wu, Q.: Learning to rank using an ensemble of lambda-gradient models. Journal of Machine Learning Research **14** (2011) 25–35
22. Friedman, J.: Greedy function approximation: a gradient boosting machine. Annals of Statistics **29** (2001) 1180
23. Hu, B., Liu, N.N., Chen, W.: Learning from click model and latent factor model for relevance prediction challenge. In: Proceedings of the Workshop on Web Search Click Data, WSDM 2012. (2012)
24. Figurnov, M., Kirillov, A.: Linear combination of random forests for the relevance prediction challenge. In: Proceedings of the Workshop on Web Search Click Data, WSDM 2012. (2012)

The Relationship between Trust and Budgetary Slack: an Empirical Study

María Gilabert-Carreras[1], Susana Gago[2], and David Naranjo-Gil[1]

[1]Pablo de Olavide University, Spain
[2]Carlos III University, Spain

mgilcar@upo.es, Susana.gago@uc3m.es, dnargil@upo.es

Abstract. The budgetary slack has been studied extensively in the management and accounting literature, but results are inconclusive. This could be because the research has focused on economic factors rather than on psychological variables, such as trust. This paper tries to contrast psychological and economic causes in the creation of budgetary slack. Particularly, we examine whether a higher amount of subordinates' trust in their superiors with an economic incentive helps to reduce the tendency of subordinates to create budgetary slack. This study is based on a laboratory experiment conducted with 240 managers in order to investigate how trust, understood as a psychological and moral factor, contributes toward the reduction of slack in the absence and presence of pecuniary incentives. Subjects were divided into three groups: managers, executives and controller. Results support partially our hypotheses. This paper shows that non-monetary incentives could help managers to reduce budgetary slack in organizations.

Keywords: budgetary slack, trust, monetary and non-monetary incentives.

1 Introduction

The existence of slack leads harmful consequences for companies like lost business opportunities and/or inflated costs. The word "slack" is used to describe a circumstance in which the resources and effort employed in the development of an activity no longer contribute to the achievement of organizational objectives (Cyert and March, 1963). The budgetary slack has been studied extensively in the management and accounting literature. However, the results obtained are not concluding about the source of this slack and the way to reduce it (Fisher et al., 2007). One possible explanation for this is that experimental research has focused primarily on testing theory-based economic models, with no reference to the various psychological, social, and institutional issues that contribute to the behavior of slack in practice (Covalenski et al., 2003).

The goal of this paper is to develop new theories that integrate behavioral and economic factors, and thus we treat together a psychological factor (trust) and an economic factor (economic incentives).

The current dominant economic view of slack is based on agency models. Agency models theorize that providing agents (subordinates) with more information than principals (supervisors) need not result in greater efficiency; the reason for this is that agents may use this information to shirk. Assuming an agency perspective, numerous experimental studies have studied the effects of risk aversion, information asymmetry, and pay schemes on budgetary slack; the goal of this research is to understand the incentives that promote honesty in agents (e.g. Chow et al., 1988) and if the incentives that promote honesty are not in conflict with economic incentives. Recent experiments incorporate social, institutional, ethical, and psychological factors, under the assumption that they also influence agents' decisions about slack. Social pressure, identification with a group, personal integrity, and aversion to lying are examples of non-economic factors that can affect budgetary slack and an agent's level of effort. In this line, the seminal experiment of Young (1985) provides evidence that risk-averse subjects create more slack than non-risk averse subjects. In the absence of information asymmetry, social pressure to reveal truthful information mitigates the amount of slack. The experiment of Young et al. (1993) suggests that cooperativeness is a relevant factor. Although cooperativeness among subjects does not necessarily result in less slack than internal competition, it has an incremental effect. Evans et al. (2001) observe in their laboratory that subjects are prepared to surrender some payoff for reporting honestly, or honestly in part. This finding contradicts the assumption in experiments that firms can achieve honest reporting if they pay enough for it, i.e., the revelation principle. In addition, the experiment of Stevens (2002) indicates that ethical concerns are negatively correlated with slack under a slack-inducing pay scheme, and independent of information asymmetry. Hannan et al (2006) observe in their experiment that subjects are willing to sacrifice the benefits of misrepresentation for being (appearing) honest because they prefer to create a positive impression. Brügen and Moers (2007) find that ethical concerns and social incentives, stated as individual and social norms, respectively, mutually reinforce the behavior of subjects and mitigate agency problems.

In summary, results in prior experiments suggest that subjects with no economic incentives to cooperate (because they are paid with slack-inducing schemes) nonetheless reduce the amount of slack in the laboratory, and as a consequence their wealth. Not only do subjects create less budgetary slack than expected, but in addition honesty can prevail in the absence of pecuniary incentives. In other words, the introduction of explicit monetary incentives may weaken non-pecuniary incentives. The experiment conducted by Rankin et al. (2005) disentangles the preference for honesty from other non-pecuniary preferences, demonstrating that subjects who have final budget authority significantly prefer honesty. In addition, the slack generated in this experiment was less than the theory predicted.

2 Hypotheses development

Trust can be defined as the willingness of one party (trustor) to be vulnerable to the actions of another party based on the expectation that the other will perform in the way that trustor expect (Mayer et al. 1995). We can also characterize trust as the "undertaking of a risky course of action on the confident expectation that all persons involved in the action will act competently and dutifully" (Lewis & Weigert 1985). Similarly, Robinson (1996) defined trust as a person's expectations, assumptions, or beliefs about the likelihood that another's future actions will be beneficial, favorable, or at least not detrimental to one's interests. An important number of economic and accounting laboratory experiments have applied the trust game, which aims to determine how much cooperation develops among individuals when they may possibly gain from it. In these experiments subjects exhibit substantial trust and reciprocity (e.g. Berg et al., 1995; Fehr and Gächter, 1998; Evans al., 2001). These experiments see trust as a rational decision. However, trust does not always operate like the element of calculated risk that is ubiquitous in economic models. Trust is also seen by managers as a predilection to assume the best when interpreting another's motives, regardless of economic incentives (Coletti et al., 2005; Kramer, 1999; Uzzi, 1997). Hence, we view trust as a psychological and moral issue. This approach differs from the previous rational view, where trust arises in games when the economic incentives favor cooperative behavior. Furthermore, trust encompasses several different levels: trust, no trust, and distrust. Trust and distrust lie at the extremes of a continuum. While trust is based on confidence in another, distrust refers to the concern that another may act to do harm.

In summary, we expect that in the laboratory: a) subjects who distrust or don't trust but are economically encouraged, are prone to decrease slack; b) subjects who distrust or don't trust but are not economically encouraged, are prone to ever-increasing slack; c) subjects who trust but are economically encouraged, submit budgets with higher slack; and d) subjects who trust but are not economically encouraged, submit budgets with low slack. Thus, we formulate the following hypothesis:

Hypothesis 1: Subjects who trust in their superiors and are not economically encouraged choose budgets with less slack than subjects who evidence distrust or "no trust" and are economically encouraged.

The manipulation of the level of trust in the laboratory should have consequences for subjects' choice of budgetary slack. When suspicion about superiors arises, budgetary slack should increase. This results from the fact that trust is formed over time (Rousseau et al., 1998).

It is always feasible to move managers from their initial positions along the continuum of trust-distrust because trust is an induced mind-set. Trust is a non-personality factor, susceptible to change when individuals interact in laboratory experiments. Thus, an individual can change his or her level of trust (or mistrust) while attempting to solve a problem (Rowe, 2004; Zand, 1972). Trust can then be altered both with and without economic incentives (Zand, 1972). In particular, we are interested in the effects on slack that result from altering trust in the presence and absence of economic implications.

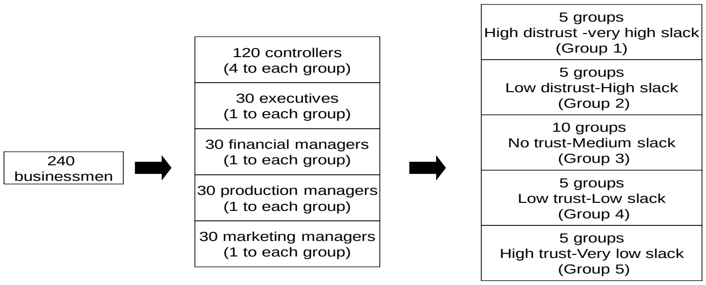
Having established a level of trust with another person, a perception that trust is one-sided leads to some diminution. When individuals begin to doubt that another person is operating in good faith, they manifest suspicion. Suspicion, in turn, results in a loss of trust. Similarly, individuals begin to distrust when they anticipate violations of trust in the future. The thought that unfulfilled expectations in one interpersonal exchange are likely to manifest in all other exchanges leads to distrust. Distrust emerges through negative expectations, assumptions, or beliefs about others' motives. Recurring abuses further increase distrust (Jones and George, 1998; Sitkin and Roth, 1993).

Hypothesis 2: A reduction in trust generates an increase in slack, independently of the presence and absence of economic incentives.

3 Experimental Design

The laboratory experiment employed a 5 (trust-slack levels) x 3 (information asymmetries on trust and slack) factorial design. We randomly assigned 240 participants to the roles of 30 executives, 90 managers, and 120 controllers, and in 30 groups (see Figure 1).

Fig. 1. Research experimental design


240 businessmen
120 controllers (4 to each group)
30 executives (1 to each group)
30 financial managers (1 to each group)
30 production managers (1 to each group)
30 marketing managers (1 to each group)
5 groups High distrust -very high slack (Group 1)
5 groups Low distrust-High slack (Group 2)
10 groups No trust-Medium slack (Group 3)
5 groups Low trust-Low slack (Group 4)
5 groups High trust-Very low slack (Group 5)

In order to design our experiment, we based on a previous one. That experiment tried to check if subjects who trust in their superiors choose budgets with less slack than subjects who evidence no trust or distrust without any monetary incentives. It was recruited a total of 240 businesspersons enrolled in postgraduate business courses to participate in an experiment about the effect of trust on budgetary slack. Subjects were pseudo-volunteers, as the experiment was part of a class assignment. The subjects did not receive payment for their participation in this experiment. Businesspersons were invited to participate in the experiment as a means of improving their knowledge of the budgetary process, consistent with the notion that classroom experiments have pedagogical value (Friedman and Sunder, 1988). The average managerial experience of a participant was 3.57 years. The percentages of males and females in the sample were 68% and 32%, respectively. 48% of the participants were currently dealing with budgets in their professional activities, while all of the subjects had experience dealing with budgets at some time.

The experiment consisted of a simulation study of a business game, where participants were assigned simulation tasks (DeJong et al., 1985; Lombardo and McCall, 1982). It was replicated a corporation: namely, the travel agency of an international holding company, whose primary business activity was tourism. The laboratory experiment employed a 5 (trust-slack levels) x 3 (information asymmetries on trust and slack) factorial design. It was randomly assigned 240 participants to the roles of 30 executives, 90 managers, and 120 controllers, and in 30 groups, where different combinations of trust and slack were present. Groups were of five types: high trust-very low slack, group 5; low trust-low slack, group 4; no trust-medium slack, group 3; low distrust-high slack group, group 2; and high distrust-very high slack, group 1. Each group was composed of three managers (production, marketing, and finance), one executive, and four controllers.

It was verbally informed participants regarding the general purpose of the experiment, the resource and information endowment, the set of actions available to them, and the moral and economic consequences of each action (Friedman and Sunder, 1988). Participants also received private written instructions, which they were not allowed to reveal at any time during the experiment. It was also provided all participants with written information about the nature of the budgets under discussion. In particular, they knew the global profitability underlying each budget: a) 5.35% (budget 1), b) 5.78% (budget 2), c) 6.31% (budget 3), d) 6.68% (budget 4), and, e) 7.09% (budget 5). Nonetheless, only subjects in the roles of managers knew the amount of budgetary slack, as they were told privately that 7.10% was the maximum attainable global profitability. The amounts of slack were 0.01% (budget 5), 0.42% (budget 2), 0.79% (budget 3), 1.32% (budget 2) and, 1.75% (budget 1). Thus, they were aware of the slack associated with each budget.

3.1 Variables Measurement

The endogenous variables are: the first budget proposed, which represents the earliest manifestation of slack (V1), and the final budget, which is the last manifestation of slack (V2) (Fisher et al., 2000, 2002). The exogenous variable group (V3) refers to the five types of groups. The five groups are based on the participation of the managers in previous conditions of high distrust-very high slack (group 1), low distrust-high slack (group 2), no trust-medium slack (group 3), low trust-low slack (group 4), and high trust-very low slack (group 5). As soon as the meeting was completed, we questioned all the participants about their evaluation of the final level of trust executives had in managers (V4). Final trust was measured from 1 (high distrust) to 5 (high trust).

Both executives and controllers were uninformed about slack conditions and the amount of slack. Hence, to identify how aware executives and controllers were of slack during discussion of the budget, they were asked about the effort that managers invested in their last budget proposal. A variable based on effort was built, which varies from 1 (very easy to attain) to 5 (required their maximum effort) (V5). It was also checked if executives and controllers were conscious of: a) managers' success in submitting budgets easily attainable (V6); b) if budgetary targets induce high managerial productivity (reverse code) (V7); c) if it was costly to manage budgets carefully (reverse code) (V8); and d) if they thought that budgets had motivated managers to be concerned with improving efficiency (V9). Executives' and controllers' responses were on a scale from 1 (definitely true) to 7 (definitely false). With regard to trust, executives gleaned some indirect information through the level of cooperation, whereas controllers knew nothing. To differentiate between these two situations, a binary control variable that we denote as the absence of information on trust was defined (V10); this provides a value of one for controllers and zero for executives. We also control for gender differences (V11), professional experience (years in the workplace as a manager) (V12), and previous knowledge of budgets (V13).

4 Results

To test our hypotheses, a multinomial logit model was specified (Hosmer and Lemeshow, 1989; Menard, 2002). The initial budget is the dependent variable; the group and control variables comprise the independent variables. The initial budget is the response variable in five categories. Four equations were derived. Each of the four equations comprises a multinomial logistic regression comparing the other budgets with budget 1 (slack=1.75%). The multinomial logistic regression model takes the form:

$$P\,(y_k = 1/\ ß_k, x) = \exp\,(\,ß_{kT}\,.x)\,/\sum \exp\,(\,ß_{kT}\,.x)\,, \qquad (1)$$

Where y is the class indicator for the *k*th budget; x is the predictor vector extended by one to be paired with the intercept parameter. Each βk is a vector of parameters, one for each class (the letter T means total). The initial budgets diverge. Subjects in the role of managers start the budgetary meeting with budget 3 (amount of slack: 0.79%) 28.9% of the time, followed by budget 1 (slack: 1.75%) 24.4% of the time, and budget 2 (slack: 1.32%) 21.1% of the time. Budgets 4 (slack: 0.42%) and 5 (slack: 0.01%) are chosen less frequently, 12.2% and 13.3% of the time, respectively. A Wald test permitted appraising the significance of the individual logistic regression coefficients for the variable group (V4) and the insignificance of the control variables (Table 3). Using the Wald statistic, group is significant with the exception of Budget 2. Likelihood ratio tests also corroborate the significance of group and the insignificance of the control variables (see Table 1).

The odds ratio, Exp (B), in Table 1 shows that as group increases by one unit, the odds ratios of budget 3 (slack= 0.79%), budget 4 (slack= 6.68%), and budget 5 (slack= 0.01%) increase by multiples of 4.05, 5.14, and 2.43, respectively, once the variables for sex (V11), years at work (V12), and budget experience (V13) were controlled. Thus, the parameter estimates confirm that when one-time prior conditions of subjects move from distrust-high slack to trust-low slack, the probability of a subject submitting initial budgets with low slack (0.42%), medium slack (0.79%), and very low slack (0.01%) increases. This result confirms, to some extent, Hypothesis 1: Subjects who previously trust create less slack than managers who distrust, i.e., they intend to invest more effort. We cannot show, however, that subjects who evidence low distrust in their superiors produce more (or less) sack than he ones who evidence high distrust.

Table 1. Initial Proposals of Budgets by Managers: Parameter Estimates and Likelihood Ratio Tests

Panel A: Parameter Estimates							
Initial Proposal of Budget by Managers	Independent variables	B	Std. error	Wald	D.f.	Sig.	Exp (B)
Budget 2 (Slack= 1.32%)	Intercept	–0.086	0.985	0.008	1	0.930	
	Group V3	0.117	0.303	0.149	1	0.699	1.124
	Sex V11	0.080	0.696	0.013	1	0.908	1.083
	Years at work V12	–0.162	0.137	1.396	1	0.237	0.851
	Budget experience V13	0.418	0.739	0.320	1	0.572	1.519
Budget 3 (Slack=0.79%)	Intercept	–4.620	1.372	11.330	1	0.001	
	Group V3	1.399	0.356	15.430	1	0.000	4.051
	Sex V11	–0.249	0.733	0.115	1	0.734	0.780

	Years at work	V12	0.040	0.122	0.106	1	0.745	1.041
	Budget experience	V13	1.515	0.758	3.996	1	0.046	4.550
Budget 4 (Slack=0.42%)	Intercept		–6.737	1.925	12.246	1	0.000	
	Group	V3	1.637	0.462	12.558	1	0.000	5.141
	Sex	V11	–0.482	0.911	0.280	1	0.597	0.618
	Years at work	V12	0.059	0.142	0.175	1	0.676	1.061
	Budget experience	V13	2.198	0.953	5.317	1	0.021	9.005
Budget 5 (Slack=0.01%)	Intercept		–4.738	10.593	8.844	1	0.003	
	Group	V3	0.888	0.377	5.562	1	0.018	2.430
	Sex	V11	1.003	0.973	1.062	1	0.303	2.726
	Years at work	V12	0.048	0.132	0.132	1	0.717	1.049
	Budget experience	V13	1.819	0.856	4.520	1	0.034	6.165

Panel B: Likelihood Ratio Tests

	Effect		–2 log likelihood of reduced model	Chi-square	D.f.	Sig.
	Intercept		246.767	34.360	4	0.000
	Group	V3	247.445	35.038	4	0.000
	Sex	V11	215.213	2.806	4	0.591
	Years at work	V12	215.176	2.769	4	0.597
	Budget experience	V13	221.381	8.974	4	0.062

The chi-square statistic is the difference in –2 log likelihoods between the final model and the reduced model. The reduced model is formed by omitting a variable from the final model. The null hypothesis is that all parameters of the effect are zero

A different multinomial logit model for closing budgets was constructed. Final budget is the dependent variable with five categories generating four equations. Each of the four equations is a binary logistic regression that contrasts other budgets with Budget 1 (very high slack). Multinomial logistic regression simultaneously estimates the four logits.

Final budgets show some discrepancy. A greater number of subjects (32.2%) finish the budgetary meeting agreeing to budget 3 (slack=0.79%). Smaller numbers of managers decide on other budgets: 14.4% are inclined to close the meeting with budget 1 (slack=0.79%), 20.0% with budget 2 (Slack=1.32%), 18.9% with budget 4 (slack=1.32%), and 14.4% with budget 5 (slack=0.01%). It was found that the amount of slack in final budgets is less than in initial budgets. Therefore, disagreement appears to reduce slack on average. 57 subjects adhere to their opening budget proposals, however, while 33 subjects change their final proposal from the opening offer. Using the Wald statistic, group (V3) is significant with the exception of budget 2 (Table 2), and as well as in the likelihood ratio tests (Table2). The odds ratio, Exp (B), bears out the preceding outcome. A one unit increase in group, i.e., subjects moving towards early high-trust and low-slack, brings about an increase of 1.998 in the odds ratio of budget 3 (slack=0.79%), and 2.152 in the odds ratio of budget 4 (slack=0.42%). The odds of budget 2 (slack=1.32%) and budget 5 (slack=0.01%) as final proposals by subjects in meetings, however, are not significantly explained by the initial group.

Table 2. Final Proposals of Budgets by Managers: Parameter Estimates and Likelihood Ratio Tests

Panel A: Parameter Estimates

Final Proposal of Budget by Managers	Independent variables		B	Std. error	Wald	D. f.	Sig.	Exp (B)
Budget 2	Intercept		3.252	1.600	4.132	1	0.042	
(Slack= 1.32%)	Group	V3	–0.057	0.333	0.030	1	0.864	0.944
	Final trust	V4	–0.918	0.414	4.904	1	0.027	0.399
	Sex	V11	–0.058	0.856	0.005	1	0.946	0.944
	Years at work	V12	0.098	0.143	0.466	1	0.495	1.102
	Budget experience	V13	–0.606	0.808	0.562	1	0.453	0.545
Budget 3	Intercept		1.392	1.478	0.887	1	0.346	
(Slack=0.79%)	Group	V3	0.692	0.306	5.125	1	0.024	1.998
	Final trust	V4	–0.711	0.403	3.115	1	0.078	0.491
	Sex	V11	–0.389	0.772	0.254	1	0.614	0.678
	Years at work	V12	0.067	0.130	0.268	1	0.605	1.070
	Budget experience	V13	–0.639	0.737	0.752	1	0.386	0.528
Budget 4	Intercept		1.714	1.634	1.100	1	0.294	
(Slack=0.42%)	Group	V3	0.767	0.357	4.601	1	0.032	2.152
	Final trust	V4	–0.952	0.446	4.562	1	0.033	0.386
	Sex	V11	–0.360	0.837	0.185	1	0.667	0.697
	Years at work	V12	–0.143	0.178	0.649	1	0.420	0.866
	Budget experience	V13	0.010	0.820	0.000	1	0.990	1.010
Budget 5	Intercept		–1.311	1.775	0.545	1	0.460	
(Slack=0.01%)	Group	V3	0.480	0.339	2.013	1	0.156	1.617
	Final trust	V4	–0.320	0.456	0.491	1	0.483	0.726
	Sex	V11	0.932	1.023	0.831	1	0.362	2.541
	Years at work	V12	0.100	0.136	0.545	1	0.460	1.105

Budget experience	V13	–0.177	0.841	0.045	1	0.833	0.837

Panel B: Likelihood Ratio Tests

Effect		–2 log likelihood of reduced model	Chi-square	D.f.	Sig.
Intercept		253.165	11.485	4	0.022
Group	V3	253.461	11.780	4	0.019
Final trust	V4	249.810	8.130	4	0.087
Sex	V11	244.551	2.871	4	0.580
Years at work	V12	245.542	3.862	4	0.425
Budget experience	V13	243.354	1.674	4	0.795

The chi-square statistic is the difference in –2 log likelihoods between the final model and the reduced model. The reduced model is formed by omitting a variable from the final model. The null hypothesis is that all parameters of the effect are zero

The Wald test also indicates that final trust (V4) explains the odds ratios of final budgets 2 and 4 (Table 2). The likelihood ratio tests, however, do not strongly support the significance of final trust (Table 2). If the ending trust increases by one unit (towards high trust), the odds ratios of budget 2 (slack=1.32%) and budget 4 (slack=0.42%) are less than one. Further units of final trust generate a reduction of 0.399 in the odds ratio of Budget 2, and 0.386 in the odds ratio of Budget 3 (Table 2). Accordingly, once final trust increases, the probability that subjects propose low and medium slack final budgets, instead of budgets with the maximum slack, is less. The exception is the odds ratio for budget 5, which is found to be insignificant. These findings validate to some extent Hypothesis 2: By taking into consideration the fact that final trust produces consequences for subjects' slack choices, follow-on slack is greater than before as one introduces suspicion. The initial trust and slack conditions, however, are determinants for most managers. For example, 57 managers do not adjust their original budget suggestions.

5 Discussion and Conclusions

The experiment shows that trust, understood as a moral and psychological factor, ameliorates the problem of slack in the absence of any explicit link between trust preferences and the distribution of wealth (as recommended by Rankin et al., 2005). The existence of subjects' distrust of their superiors stimulates higher levels of slack.

The budgets initially and finally proposed by subjects in the role of managers contain less slack than expected, which is consistent with previous findings that indicate that subjects tend to produce less budgetary slack than agency theory predicts (e.g. Stevens, 2002). The results also show that prior conditions of trust and slack facilitate the understanding of subjects' preferences for proposing initial budgets. This explains the likelihood of budgets with medium, low, and very low slack, but not budgets with high slack. An incremental effect on subjects' honesty, i.e., a reduction in slack, was found related to trust in a budgetary setting in which the superior has the final authority over budget approval. That seems to contradict the previous finding of Rankin et al. (2005) that suggests that there is no incremental effect on honesty when a superior has final authority over budget approval, while the opposite occurs when subordinates dictate the budget. Trust maybe acts as a moderator, positively motivating subjects to honesty when superiors dictate budgets. The trust levels on subjects in the role of managers were manipulated by introducing suspicion. Nevertheless, most of subjects held to their initial budgetary choices throughout the meeting. This finding demonstrates the weight of initial conditions of trust and slack in budgetary settings. In particular, the significance of the preceding trust-slack interaction in a trust-low slack environment, as preferences for medium and low slack budgets over very high slack budgets were moderated by group discussion.

Several subjects in the role of managers adjusted their budgetary choice. As soon as their final trust moved from distrust towards trust, subjects modified their budgets and thus their slack. In response, most subjects decided to reduce, rather than increase, slack. When suspicion appeared in the budgetary setting, and the managers' final trust shifted from a position of distrust to trust, the possibility that managers chose budgets with high, medium, and low slack, instead of very high slack, diminished as the final conditions depended more on trust. This is a key finding: Budgetary slack levels decrease in trust settings, even in the absence of any direct pecuniary incentive. This finding is relevant for management accounting researchers as trust, understood as a psychological and moral factor, has a positive effect on the amount of slack. But we ask: budgetary slack levels decrease more (in trust settings) in the presence of monetary incentive? Or conversely budgetary slack levels decrease less? We try to design an experiment that answers these questions.

6 References

1. Berg, J, Dickhaut, J. and K. McCabe, 1995, Trust, reciprocity, and social norms, *Games and Economic Behavior* 10 (1), 122-142.
2. Brüggen, A. and F. Moers, 2007, The role of financial incentives and social incentives in multi-task settings, Journal of Management Accounting Research; 2007 19, 25-50.
3. Chow, C. W., Cooper, J. C., and W. S. Waller, 1988, Participative budgeting: Effects of a truth-inducing pay scheme and information asymmetry on slack and performances, The Accounting Review 63 (1), 111–122.
4. Coletti, A. L., Sedatole, K. L., and K. L. Towry, 2005, The effect of control systems on trust and cooperation in collaborative environments, *The Accounting Review* 80 (2), 477–500.
5. Covaleski, M. A., Evans III, J. H., Luft, J. L., and M. D. Shields, 2003, Budgeting research: Three theoretical perspectives and criteria for selective integration, *Journal of Management Accounting Research* 15, pp. 3-49.
6. Cyert, R. M. and J. G. March, 1963, *A behavioral theory of the firm* (Prentice-Hall, Englewood Cliffs, NJ).
7. DeJong, D. V., Forsythe, R., Lundholm, R.J., and W. C. Uecker, 1985, A laboratory investigation of the moral hazard problem in an agency relationship, *Journal of Accounting Research* 23 (Supplement), 81-120.
8. Evans, J. H., Hannan, R. L., Krishnan, R., and D. V. Moser, 2001, Honesty in managerial reporting, *The Accounting Review* 76 (4), 537–559.

9. Fehr, E. and Gächter, S., 1998, How effective are trust-and reciprocity-based incentives?, in A. Ben-Ner and L. Putterman, eds., *Economics, values, and organization* (Cambridge University Press, United Kingdom) 337–363.
10. Fisher, J. G., Frederickson, J. R., and S. A. Peffer, 2000, Budgeting: An experimental investigation of the effects of negotiation, *The Accounting Review* 75 (1), 93–114.
11. Fisher, J. G., Frederickson, J. R., and S. A. Peffer, 2002, The effect of information asymmetry on negotiated budgets: An empirical investigation, *Accounting, Organizations and Society* 27 (1–2), 27–43.
12. Fisher, J. G., Sprinkle, G. B., and L. Walker, 2007, Experimental Budgeting Research: Implications for Practitioners, *The Journal of Corporate Accounting & Finance* 18 (6), 67–75.
13. Friedman, D., and S. Sunder, 1994, *Experimental methods. A primer for economist* (Cambridge University Press, Cambridge: NY).
14. Hannan, R., Rankin, F., Towry, K., Salterio, S., and A. Webb, 2006, The effect of information systems on honesty in managerial reporting: A behavioral perspective, Contemporary Accounting Research 23 (4), 885-932.
15. Hosmer D. W., Lemeshow S. *Applied Logistic Regression*. New York: Wiley 1989.
16. Jones, J. R. and J. M. George, 1998, The experience and evolution of trust: Implications for cooperation and teamwork, *Academy of Management Journal* 23 (3), 531–546.
17. Kramer, R. M., 1999, Trust and distrust in organizations: Emerging perspectives, enduring questions, *Annual Review of Psychology* 50 (1), 569–598.
18. Lewis JD, Weigert A. 1985. Trust as a social reality. Soc. Forces 63:967–85
19. Lombardo, M. M. and M. W. McCall, 1982, Leaders on line: observations from a simulation of managerial work, in J. G. Hunt, U. Sekaran, and C. A. Schriesheimn (eds.), *Leadership Beyond Established Views* (Southern Illinois University PresS: Carbondale), 50-67.
20. Mayer., R. C., Davis, J. H., and F. D. Schoorman, 1995, An integrative model of organizational trust, *The Academy of Management Review* 20(3), 709-734.
21. Rankin, F. W., S. Schwartz, and R. Young, 2005, The effect of honesty preferences and superior authority on budget proposals, *Working Paper*.
22. Robinson, Sandra L, 1996, Trust and breach of the psychological contract, Administrative Science Quarterly 41. 4: 574-599.
23. Rousseau, D. M., Sitkin, S. B., Burt, R. S., and C. Camerer, 1998, Not so different after all: A cross-discipline view of trust, *Academy of Management Journal* 23 (3), 393–404.
24. Rowe, C., 2004, The effect of accounting report structure and team structure on performance in cross-functional teams, *The Accounting Review* 79(4): 1153-1180.
25. Sitkin, S. and N. Roth, 1993, Explaining the limited effectiveness of legalistic remedies for trust/distrust, *Organization Science* 4 (3), 367–392.
26. Stevens, D. E., 2002, The effects of reputation and ethics on budgetary slack, *Journal of Management Accounting Research* 14, 153-171.
27. Uzzi, B., 1997, Social structure and competition in interfirm networks: The paradox of embeddedness, *Administrative Science Quarterly* 42(1), 35-67.
28. Young, S. M., 1985, Participative budgeting: The effects of risk aversion and asymmetric information on budgetary slack, Journal of Accounting Research 23 (2), 829–842.
29. Zand, D. E., 1972, Trust and managerial problem solving, *Administrative Science Quarterly* 17 (2), 229–239.

CrowDM: the System for Collaborative Platform Data Analysis

Dmitry I. Ignatov[1], Alexandra Yu. Kaminskaya[1,2], Anastasya A. Bezzubtseva[1,2], Ekaterina L. Chernyak[1,2], Konstantin N. Blinkin[1], Daniil R. Nedumov[1], Olga N. Chugunova[1], Andrey V. Konstantinov[1], Nikita S. Romashkin[1] Fedor V. Strok[1], Daria A. Goncharova[1,2], and Rostislav E.Yavorsky[2]

[1] National Research University Higher School of Economics, Russia, 101000, Moscow, Myasnitskaya str., 20
dignatov@hse.ru
http://www.hse.ru
[2] Witology
rostislav.yavorskiy@witology.com
http://www.witology.com

Abstract. The paper considers a data analysis system of the Witology company collaborative platform and mainly describes a methodology and results of the first experiments. The developed system is based on several models and methods of modern analysis of object-attribute and unstructured data (texts) such as Formal Concept Analysis, multimodal clustering, association rules mining and keywords and collocations extraction from texts.

Keywords: collaborative and crowdsourcing platforms, Data Mining, Formal Concept Analysis, multimodal clustering.

1 Introduction and related work

The success of modern collaborative technologies is marked by the appearance of many novel platforms for holding distributed brainstorming or carrying out so called "public examination". There are a lot of such crowdsourcing companies in USA (Spigit [1], BrightIdea [2], InnoCentive [3] etc.) and Europe (Imaginatik [4]). A couple of years ago Russian companies launched business in that area as well. Two most vivid examples of such Russian companies are Witology [5] and Wikivote [6]. The reality as yet is far away from technological breakthrough, though some all-Russian projects have already been finished successfully (for example, Sberbank-21, National Entrepreneurial Initiative-2012 [7] etc.). The core of such crowdsourcing systems is a socio-semantic network [8,9,10,11], which data requires new approaches to analyze. This paper is devoted to the new methodological base for the collaborative systems data analysis, which uses modern data mining and artificial intelligence models and methods. As a rule, while participating in a project, users of such crowdsourcing platforms [12] discuss and solve one common problem, propose their ideas, evaluate ideas of each other as experts. Finally, as a result of the discussion and ranking of users and their ideas we get the best ideas and users (their generators). For deeper understanding of users's behavior, developing the sufficient ranking criteria, dynamics and statistics analysis the special means are needed. Traditional methods of clustering, community detection and text mining need to be adapted or even fully redesigned. Moreover, these methods require ingenuity for their effective and efficient use (finding non-trivial results). We briefly describe models of data used in crowdsourcing projects in terms of Formal Concept Analysis (FCA) [13]. Furthermore, we present the collaborative platform data analysis system CrowDM (Crowd Data Mining), its architecture and methods underlying the key steps of data analysis.

The remainder of the paper is organized as follows. In section 2 we describe some key notions from FCA, our data and methods. In section 3 we discuss the analysis scheme of the developed system. In section 4 we present the results of our first experiments with the Sberbank-21 data. Section 5 concludes our paper and describes some possible directions for future research.

2 Mathematical models and methods

At the initial stage of collaborative platform data analysis two data types were identified: data without using keywords (links, evaluations, user actions) and data with keywords (all user-generated content). These two data types totally correspond with two components of a socio-semantic network. For the analysis of the 1st type of data (with keywords) we suggest to apply Social Network Analysis (SNA) methods, clustering (biclustering and triclustering [14,15,16], spectral clustering), FCA (concept lattices, implications, association rules) and its extensions for multimodal data, triadic, for instance [17]; recommender systems [18,19,20] and statistical methods of data analysis [21] (the analysis of distributions and average values).

Methods described in this paper are colored blue at the analysis scheme (see fig. 2). The protagonists of crowdsourcing projects (and corresponding collaborative platforms) are platform users (project participants). We consider them as *objects* for analysis. More than that, each object can (or cannot) possess a certain set of *attributes*. The user's attributes can be: topics which the user discussed, ideas which he generated or voted for, or even other users. The main instrument for analysis of such object-attribute data is FCA. Let us give formal definitions. *The formal context* in FCA is a triple $\mathbb{K} = (G, M, I)$, where G is a *set of objects*, M is a*set of attributes*, and the relation $I \subseteq G \times M$ shows which object which attribute possesses. For any $A \subseteq G$ and $B \subseteq M$ one can define *Galois operators*:

$$\begin{aligned} A' &= \{m \in M \mid gIm \text{ for all } g \in A\}, \qquad (1)\\ B' &= \{g \in G \mid gIm \text{ for all } m \in B\}. \end{aligned}$$

The operator $''$ (applying the operator $'$ twice) is a *closure operator*: it is idempotent ($A'''' = A''$), monotonous ($A \subseteq B$ implies $A'' \subseteq B''$) and extensive ($A \subseteq A''$). The set of objects $A \subseteq G$ such that $A'' = A$ is called closed. The same is for closed attributes sets, subsets of a set M. A couple (A, B) such that $A \subset G$, $B \subset M$, $A' = B$ and $B' = A$, is called *formal concept* of a context K. The sets A and B are closed and called *extent* and *intent* of a formal concept (A, B) correspondingly. For the set of objects A the set of their common attributes A' describes the similarity of objects of the set A, and the closed set A'' is a cluster of similar objects (with the set of common attributes A'). The relation "to be more general concept" is defined as follows: $(A, B) \geq (C, D)$ iff $A \subseteq C$. The concepts of a formal context $\mathbb{K} = (G, M, I)$ ordered by extensions inclusion form a lattice, which is called *concept lattice*. For its visualization the *line diagrams* (Hasse diagrams) can be used, i.e. cover graph of the relation "to be more general concept". In the worst case (Boolean lattice) the number of concepts is equal to $2^{\{\min |G|, |M|\}}$, thus, for large contexts, FCA can be used only if the data is sparse. Moreover, one can use different ways of reducing the number of formal concepts (choosing concepts by stability index or extent size). The alternative approach is a relaxation of the definition of formal concept as maximal rectangle in object-attribute matrix which elements belong to the incidence relation. One of such relaxations is a notion of object-attribute bicluster [15]. If $(g, m) \in I$, then (m', g') is called object-attribute bicluster with the density $\rho(m', g') = |I \cap (m' \times g')|/(|m'| \cdot |g'|)$.

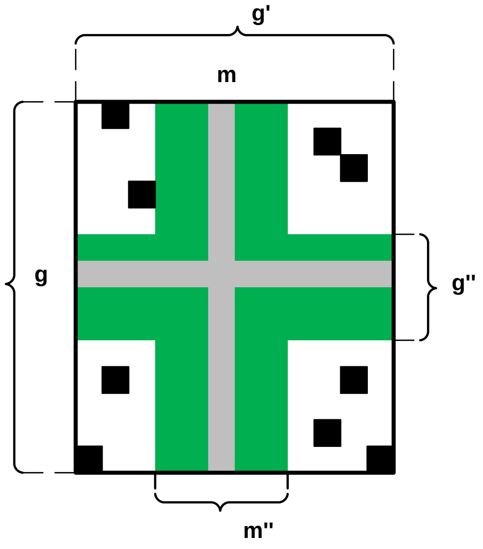


Fig. 1. OA-bicluster.

The main features of OA-biclusters are listed below:

1. For any bicluster $(A, B) \subseteq 2^G \times 2^M$ it is true that $0 \leq \rho(A, B) \leq 1$.
2. OA-bicluster (m', g') is a formal concept iff $\rho = 1$.
3. If (m', g') is a bicluster, then $(g'', g') \leq (m', m'')$.

Let $(A, B) \subseteq 2^G \times 2^M$ be a bicluster and ρ_{min} be a non-negative real number such that $0 \leq \rho_{\min} \leq 1$, then (A, B) is called *dense*, if it fits the constraint $\rho(A, B) \geq \rho_{\min}$. The above mentioned properties show that OA-biclusters differ from formal concepts since unit density is not required. Graphically it means that not all the cells of a bicluster must be filled by a cross (see fig. 1). Besides formal lattice construction and visualization by means of Hasse diagrams one can use implications and association rules for detecting attribute dependencies in data. Then, using the obtained results, it is easy to form recommendations (for example, offering users the most interesting discussions for them). Furthermore, structural analysis can be performed and then used for finding communities. Statistical methods are helpful for frequency analysis of the different users' activities. Almost all of the above mentioned methods can be applied to data containing users' keywords (in this case they become attributes of a user).

2.1 Keywords and keyphrases extraction

We consider *Keywords (keyphrases)* as a set of the most significant words (phrases) in a text document that can provide a compact description for the content and style of this document. In the remainder of this paper we do not always differentiate between keywords and keyphrases, assuming that a keyword is a particular case of a keyphrase. In our project two similar problems of keywords and keyphrases extraction arise:

1. Keywords and keyphrases of the whole Witology forum;
2. Keywords and keyphrases of one user, topic etc.

In the first case we concentrate on finding syntactically well associated keywords (keyphrases). In the second case specific words and phrases of a certain user or topic are the subject of interest. Hence, we have to use two different methods for each keywords (keyphrases) extraction problem. The first one is solved by using any statistical measure of association, such as Pointwise Mutual Information (PMI), T-Score or Chi-Square [22]. To solve the second problem we may use TF-IDF or Mutual Information (MI) measures that reflect how important the word or phrase is for the given subset of texts. All the above mentioned measures define the weight of a specific word or phrase in the text. The words and phrases of the highest weight then can be considered as keywords and keyphrases. We are more interested in the quality of extracted keywords and keyphrases than in the way we obtain them. To tokenize texts we use a basic principle of word separation: there should be either a spacee or a punctuation mark between two words. A hyphen between two sequences of symbols makes them one word. To lemmatize words we use Russian AOT lemmatizer [23], which is far from being ideal, but it is the only freely available one (even for commercial usage) for processing Russian texts. To normalize bi- and tri-grams we use one of our Python scripts that normalizes phrases according to their formal grammatical patterns. We are going to use formal contexts based on sets of extracted keyphrases and people who use them, the occurence of keyphrases in texts and so on. By analogy, keyphrases, texts and users all together form a tricontext for further analysis. Moreover, keyphrases are an essential part of a socio-semantic network model, where they are used for semantic representation of the network's nodes.

3 Analysis scheme

The data analysis scheme of CrowDM, which is developed now by the project and educational team of Witology and NRU HSE is presented in figure 2. As it was mentioned before, after downloading data from a platform database, we obtain formal contexts and text collections. In turn, the latter become formal contexts as well after keyword extraction. After that, the resulting contexts are analyzed.

4 First experiments results

For carrying out experiments we constructed formal concepts where objects are users of the platform and attributes are ideas which users proposed within one of 5 project topics ("Сбербанк и частный клиент" ("Sberbank and private client")). We selected only the ideas that reached the end or almost the end of the project. An object"user" has an attribute "idea" if this user somehow contributed to the discussion of this idea, i.e. he is an author of the idea, commented on the idea and evaluated the idea or comments which were added to the idea. Thus, the extracted formal concepts (U, I), where U is a set of users, I is a set of ideas, correspond to so called epistemic communities (communities of interests), i.e. the set of users U who are interested in the ideas of I. Figure 3 displays the diagram of the obtained concept lattice.

Each node of the diagram coincides with one formal concept (in total the lattice contains 198 concepts). A node is marked by the label of an object or an attribute if this object (moving bottom-up by diagram) or attribute (moving top-down) first appeared in this node. It is obvious that the obtained diagram is too awkward to be analyzed as a static image. Usually in such cases one can use order filters or diagrams of the sets of stable concepts or iceberg-lattices for visualization. We will showcase how to read a concept lattice using the lattice fragment in figure 4. The experiments were carried out using the program Concept Explorer (ConExp) which was developed for applying FCA algorithms to object-attribute data [24]. Clicking on a lattice node, one can see the objects and attributes corresponding to the concept which this node represents. Objects are accumulated from below (in the given example the set of objects contains User45 and User22), attributes come from above (we have only one attribute, "Микрокредиты от 1000 до 5000"("Microcredits from 1000 to 5000")). This means that User45 and User22 together took part in the discussion of the given idea and nobody else discussed it.

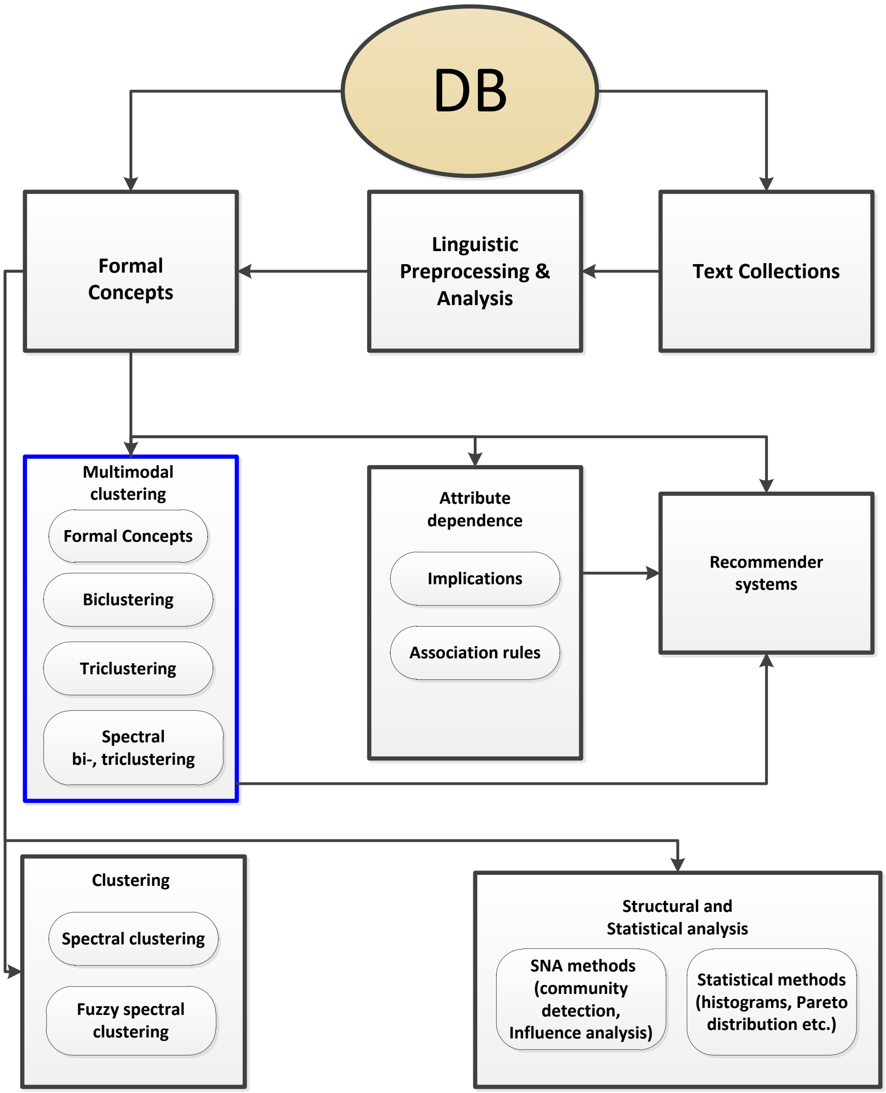

DB
Formal Concepts
Linguistic Preprocessing & Analysis
Text Collections
Multimodal clustering
Formal Concepts
Biclustering
Triclustering
Spectral bi-, triclustering
Attribute dependence
Implications
Association rules
Recommender systems
Clustering
Spectral clustering
Fuzzy spectral clustering
Structural and Statistical analysis
SNA methods (community detection, Influence analysis)
Statistical methods (histograms, Pareto distribution etc.)

Fig. 2. The data analysis scheme of CrowDM.

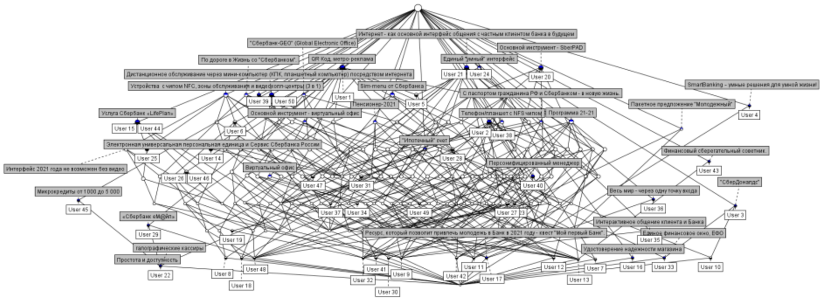
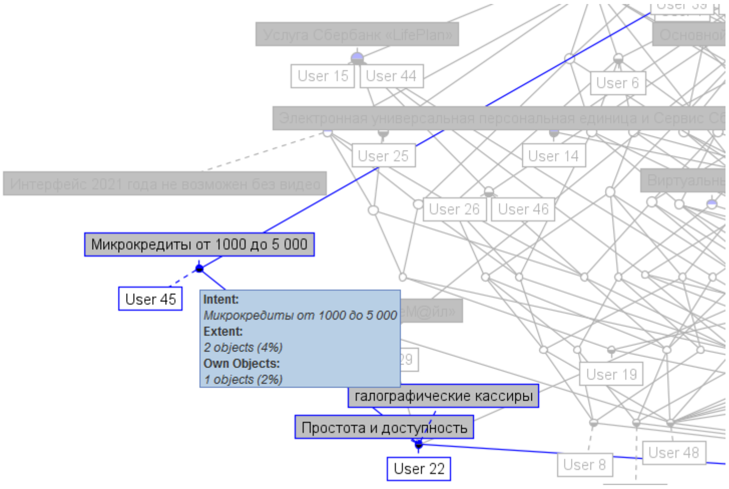

Интернет - как основной интерфейс общения с частным клиентом банка в будущем
"Сбербанк-GEO" (Global Electronic Office)
Основной инструмент - SberPAD
По дороге в Жизнь со "Сбербанком".
QR Код, метро-реклама
Единый "умный" интерфейс
Дистанционное обслуживание через мини-компьютер (КПК, планшетный компьютер) посредством интернета
User 21
User 24
User 20
Устройства с чипом NFC, зоны обслуживания и видеоколл-центры (3 в 1)
Sim-menu от Сбербанка
SmartBanking – умные решения для умной жизни!
User 39
User 50
User 1
Пенсионер-2021
User 5
С паспортом гражданина РФ и Сбербанком - в новую жизнь.
Пакетное предложение "Молодежный"
Услуга Сбербанк «LifePlan»
Основной инструмент - виртуальный офис
Телефон/планшет с NFS чипом
Программа 21-21
User 4
User 15
User 44
User 6
User 2
User 38
"Ипотечный" счет
Электронная универсальная персональная единица и Сервис Сбербанка России
User 25
User 14
User 28
Персонифицированный менеджер
Финансовый сберегательный советник.
Интерфейс 2021 года не возможен без видео
Виртуальный офис
User 43
User 26
User 46
User 47
User 31
User 40
"СберДоналдс"
Микрокредиты от 1000 до 5 000
Весь мир - через одну точку входа
User 36
User 45
User 37
User 34
User 49
User 27
User 3
«Сбербанк еМ@йл»
Интерактивное общение клиента и Банка
User 29
Ресурс, который позволит привлечь молодежь в Банк в 2021 году - квест "Мой первый Банк".
Единое финансовое окно, ЕФО
User 19
User 35
галографические кассиры
Удостоверение надежности магазина
Простота и доступность
User 48
User 41
User 11
User 12
User 7
User 16
User 33
User 10
User 22
User 8
User 9
User 42
User 17
User 32
User 13
User 18
User 30

Fig. 3. Concept lattice diagram for users-ideas context.


Услуга Сбербанк «LifePlan»
User 15
User 44
User 6
Основно
Электронная универсальная персональная единица и Сервис Сб
User 25
User 14
Интерфейс 2021 года не возможен без видео
Виртуальны
User 26
User 46
Микрокредиты от 1000 до 5 000
User 45
User 19
галографические кассиры
Простота и доступность
User 22
User 8
User 48

Intent:
Микрокредиты от 1000 до 5 000
Extent:
2 objects (4%)
Own Objects:
1 objects (2%)

Fig. 4. Fragment of concept lattice diagram

We demonstrate the results of applying biclustering algorithms on the same data below.

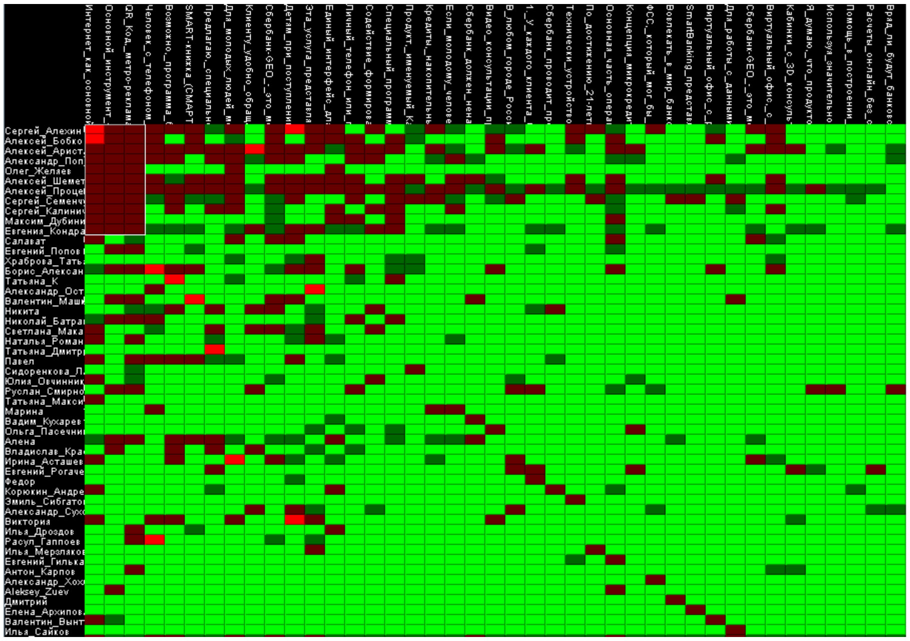


Fig. 5. Results of biclustering algorithm BiMax

Let us explain the figure 5. During experiments we used the system for gene expression data analysis BicAT [14]. Rows correspond to users, columns are ideas of a given topic ("Сбербанк и частный клиент" ("Sberbank and private client")), in the discussion of which users participated. The color of the cell of the corresponding row and column intersection depicts the contribution intensity of a given user to a given idea. The contribution is a weighted sum of the number of comments and evaluations to that idea and takes into account the fact whether this user is an author of this idea. The lightest cells coincide with zero contribution, the brightest ones (fig. 6, top left cell) show the maximum contribution. After data discretization (0 – zero contribution, 1 – otherwise) we applied the BiMax algorithm which found some biclusters (see fig. 6 for example). Since one of the important crowdsourcing project problems is the search of people with similar ideas, the presented bicluster with 11 users is most interesting while other found biclusters contained 4-5 users on average (we constrained the number of ideas in a bicluster to be strictly greater than 2).

Then, to gain a better understanding of the evaluation process in the project, evaluation distribution was plotted in several ways. One of them is presented in

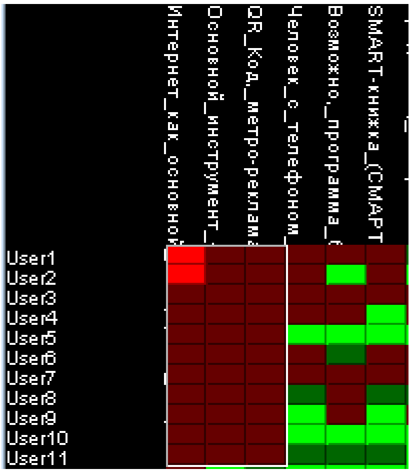


Fig. 6. Biclsuter with a large number of users

fig. 7; it shows the cumulative number of users, who made more than a certain amount of evaluations during the entire project.

The horizontal axis displays the amount of submitted evaluations. The vertical axis represents the number of users, who made more than a fixed amount of evaluations. For instance, there is only one participant, who produced more than 5000 evaluations, and one more person, who made more than 3000 but less than 5000 evaluations. Thus, the rightmost dot on the X-axis shows the first participant (the y-coordinate is 1), and the next dot shows both of them (the y-coordinate is 2). The total number of users, who have once evaluated something, is 167. The set of graph points is explicitly split into two parts: the long gentle line (from $x = 0$ to 544 inclusive) and the steep tail. The fact, that both lines seem almost straight in logarithmic scales, indicates that the evaluation activity on the project might follow a Pareto distribution. It is reasonable to seek the individual distribution functions for the main and the tail parts of the sample, as testing the whole sample for goodness of fit to a Pareto distribution results in strong rejection of the null hypothesis ($H0$: "The sample follows a Pareto distribution").

5 Conclusion

The results of our first experiments suggest that the developed methodology will be useful for analysis of collaborative systems data and resource-sharing systems. The most important directions for future work include the analysis of textual information generated by users, applying multimodal clustering methods and using them for developing recommender systems.

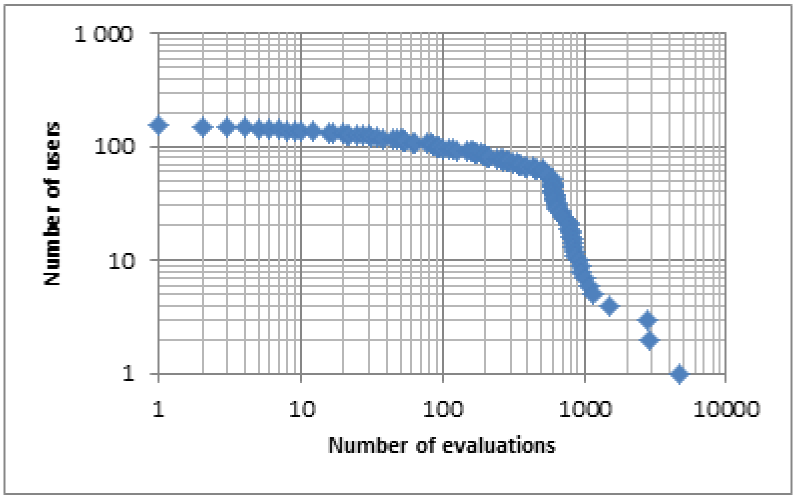


Fig. 7. Evaluation distribution

Acknowledgments. The work was performed by the project and educational group "Algorithms of Data Mining for Innovative Projects Internet Forum" (NRU HSE).

References

1. Spigit company, http://spigit.com/
2. Brightidea company, www.brightidea.com/
3. Innocentive comp, http://www.innocentive.com/
4. Imaginatik company, http://www.imaginatik.com/
5. Witology company, http://witology.com/
6. Wikivote company, http://www.wikivote.ru/
7. Sberbank-21, national entrepreneurial initiative-2012, http://sberbank21.ru/
8. Roth, C.: Generalized preferential attachment: Towards realistic socio-semantic network models. In: ISWC 4th Intl Semantic Web Conference, Workshop on Semantic Network Analysis, Galway, Ireland,. Volume 171 of CEUR-WS Series (ISSN 1613-0073). (2005) 29–42
9. Cointet, J.P., Roth, C.: Socio-semantic dynamics in a blog network. In: CSE (4), IEEE Computer Society (2009) 114–121
10. Roth, C., Cointet, J.P.: Social and semantic coevolution in knowledge networks. Social Networks **32** (2010) 16–29
11. Yavorsky, R.: Research Challenges of Dynamic Socio-Semantic Networks. In Ignatov, D., Poelmans, J., Kuznetsov, S., eds.: CEUR Workshop proceedings Vol-757, CDUD'11 - Concept Discovery in Unstructured Data. (2011) 119–122
12. Howe, J.: The rise of crowdsourcing. Wired (2006)
13. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. 1st edn. Springer-Verlag New York, Inc., Secaucus, NJ, USA (1999)

14. Barkow, S., Bleuler, S., Prelic, A., Zimmermann, P., Zitzler, E.: Bicat: a biclustering analysis toolbox. Bioinformatics **22**(10) (2006) 1282–1283
15. Igantov, D.I., Kaminskaya, A.Y., Kuznetsov, S., Magizov, R.A.: Method of Biclusterzation Based on Object and Attribute Closures. In: Proc. of 8-th international Conference on Intellectualization of Information Processing (IIP 2011). Cyprus, Paphos, October 17–24, MAKS Press (2010) 140–143 (in Russian).
16. Ignatov, D.I., Kuznetsov, S.O., Magizov, R.A., Zhukov, L.E.: From Triconcepts to Triclusters. In: Proceedings of the 13th international conference on Rough sets, fuzzy sets, data mining and granular computing. RSFDGrC'11, Berlin, Heidelberg, Springer-Verlag (2011) 257–264
17. Jäschke, R., Hotho, A., Schmitz, C., Ganter, B., Stumme, G.: TRIAS–An Algorithm for Mining Iceberg Tri-Lattices. In: Proceedings of the Sixth International Conference on Data Mining. ICDM '06, Washington, DC, USA, IEEE Computer Society (2006) 907–911
18. Ignatov, D.I., Kuznetsov, S.O.: Concept-based Recommendations for Internet Advertisement. In Belohlavek, R., Kuznetsov, S.O., eds.: Proc. CLA 2008. Volume Vol. 433 of CEUR WS., PalackΓS University, Olomouc, 2008 (2008) 157–166
19. Ignatov, D., Poelmans, J., Zaharchuk, V.: Recommender System Based on Algorithm of Bicluster Analysis RecBi. In Ignatov, D., Poelmans, J., Kuznetsov, S., eds.: CEUR Workshop proceedings Vol-757, CDUD'11 - Concept Discovery in Unstructured Data. (2011) pp. 122–126
20. Ignatov, D.I., Poelmans, J., Dedene, G., Viaene, S.: A New Cross-Validation Technique to Evaluate Quality of Recommender Systems. In Kundu, M.K., Mitra, S., Mazumdar, D., Pal, S.K., eds.: PerMIn. Volume 7143 of LNCS., Springer (2012) 195–202
21. Clauset, A., Shalizi, C.R., Newman, M.E.J.: Power-law distributions in empirical data. SIAM Rev. **51**(4) (November 2009) 661–703
22. Manning, C.D., Schütze, H.: Foundations of statistical natural language processing. MIT Press, Cambridge, MA, USA (1999)
23. Russian project on automatic text processing, www.aot.ru
24. Grigoriev, P., Yevtushenko, S.: Elements of an agile discovery environment. In Grieser, G., Tanaka, Y., Yamamoto, A., eds.: Discovery Science. Volume 2843 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2003) 311–319

A New Recommender System for the Interactive Radio Network FMhost

Vasily Zaharchuk[1], Dmitry Ignatov[1], Andrey Konstantinov[1], and Sergey Nikolenko[2,3]

[1] National Research University Higher School of Economics
dignatov@hse.ru
http://www.hse.ru
[2] Steklov Mathematical Institute, St. Petersburg, Russia
[3] St. Petersburg Academic University, St. Petersburg, Russia

Abstract. We describe a new recommender system for the Russian interactive radio network FMhost. The new recommender model combines collaborative and user-based approaches. The system extracts information from tags of listened tracks for matching user and radio station profiles and follows an adaptive online learning strategy based on user history. We also provide some basic examples and describe the quality of service evaluation methodology.

Keywords: music recommender systems, interactive radio network, e-commerce, quality of service

1 Introduction and related work

Music recommendation is an important topic in the field of recommender systems; see, e.g. the proceedings of International Society for Music Information Retrieval Conference (ISMIR) [1], Workshop on Music Recommendation and Discovery (WOMRAD) [2,3], and Recommender Systems conference (RecSys) [4]. Although such services as LastFm, Yahoo!LaunchCast and Pandora are well known, they work on a commercial basis and, moreover, the latter two of them do not broadcast for Russia. Despite many high-quality papers on different aspects of music recommendation, there are only few studies on radio station recommender systems for online services.

This work is devoted to the Russian online radio hosting service FMhost and, in particular, its new hybrid recommender subsystem. Recently, the focus of computer science research for the music industry has shifted from music information retrieval and exploration [5,6,7] to music recommender services [8,9]. The topic is not new (see, e.g., [10]); however, it is now inspired by new capabilities of large online services to provide not only millions of tracks for listening to, but even radio station hosting. Social tagging is also one of the important factors which allows to apply new tag-similarity based recommender algorithms to the domain [11,12].

Recently, a widely acclaimed public contest on music recommender algorithms, KDD Cup, was held by Yahoo! (http://kddcup.yahoo.com/). In KDD Cup, track 1 was devoted to learning to predict users' ratings of musical items (tracks, albums, artists and genres) in which items formed a taxonomy. Each track belonged to an album, albums belonged to artists, and together they were tagged by genres. Track 2 aimed at the developing learning algorithms for separating tracks scored highly by specific users from tracks not scored by them. It attracted a lot attention from the community to problems which are both typical for recommender systems and specific for music recommendation: scalability issues, capturing the dynamics and taxonomical properties of items [13]. The current trends of music recommender systems reflect advantages of hybrid approaches and show the need for user-centric quality measures [14]. For instance, in [15] an interesting approach based on "forgetting curve" to evaluate "freshness" of predictions was proposed. In [16], the authors posed an important question, namely how much metadata do we need in music recommendation, and after a subjective evaluation of 19 users the authors concluded that pure content-based methods can be drastically improved by using genres.

In [17], the authors proposed the music recommender system Starnet for social networking. It generates recommendations based either on positive ratings of friends (social recommendations), positive ratings of others in the network (nonsocial recommendations), and it also makes random recommendations. Another interesting online music recommendation system we can mention is Hotttabs [18], dedicated to guitar learning. Some authors aim at improving music recommender systems by using semantic extraction techniques [19,20]. Paper [21] describes a system of genre recommendation for music and TV programs, which can be considered as an alternative channel selector. The authors of paper [22] proposed a recommender system GroupFan which is able to aggregate preferences of group users to their mutual satisfaction.

Many online services (e.g., Last.fm or LaunchCast) call their audio streams "radio stations", but in reality they produce a playlist from a database of tracks based on a recommender system rather than actually recommend a radio channel. FMhost, on the other hand, provides users with online radio stations in the classical meaning of this term: there are human DJs who perform live, a radio station actually represents a strategy or mood of a certain person (DJ), they play their own tracks, perform contests etc. Thus, the problem we are solving differs from most of the work done in music recommendation, and some of the challenges are unique.

The paper is organized as follows. In Section 2, we describe our online service FMhost. In section 3, we propose our new recommender model, two basic recommender algorithms, and describe the recommender system architecture. Quality of Service (QoS) measurement for the system and some insights on FMhost user behaviour are discussed in Section 4. Section 5 concludes the paper.

2 Online service FMhost.me

2.1 A concise online broadcasting dictionary

Before we proceed, we need to shortly explain some basic domain terminology.

A *chart* is a radio station track rating; for example, the rock chart shows a certain number (say, 10) of most popular rock tracks, ranked from the most popular (rank 1) to the least popular (rank 10) according to the survey. A *live performance* (or just *live* for short) is a performance to which one or several *DJs* (*disk jockeys*) are assigned. They do it from their own PCs, and the audio stream is being redirected from them to the Icecast server and then everywhere. Also they may have their own blog for each live, where people may interact with DJs who perform live. *LiquidSoap* is a sound generator that broadcasts audio files (*.mp3, *.aac etc.) into an audio stream. *Icecast* is a retranslation server that redirects audio stream from one source, for example LiquidSoap, to many receivers.

2.2 The FMhost project

FMhost is an interactive radio network. This portal allows users to listen and broadcast their own radio stations. There are four user categories in the portal: (1) unauthorized user; (2) listener; (3) Disk Jockey (DJ); (4) radio station owner.

User capabilities vary upon their status. Unauthorized listeners can listen to any station, but they cannot vote or become DJs. They also cannot use the recommender system and the rating system.

Listeners, unlike unauthorized users, can vote for tracks, lives, and radio stations. They can use a recommender system or rating system. They can subscribe to lives, radio stations, or DJs. They also can be appointed to a live and become a DJ.

There are three types of broadcasting: (1) stream redirection from another server; (2) AutoDJ translation; (3) live performance.

Stream redirection applies when a radio station owner has its own server and wants to use FMhost as a broadcasting platform, but also wants to broadcast using his own sound generator, e.g., SamBroadcaster (`http://spacial.com/sam-broadcaster`), LiquidSoap (`http://savonet.sourceforge.net/`) etc. AutoDj is a special option that allows the users to play music directly from the FMhost server. Every radio owner gets some space where he can download as much tracks as he can, and then LiquidSoap will generate the audio stream and the Icecast (`http://www.icecast.org/`) server will redirect it to the listeners. Usually the owner sets a radio schedule which is being played.

Live performances are done by DJs. Everyone who has performed live at least once can be called a DJ. He can also be added to a radio station crew. Moreover, a DJ can perform lives at any station, not only on his own station where he is in a crew.

FMhost was the first project of its kind in Russia, starting in 2009. Nowadays, following FMhost's success, there exist several radio broadcasting portals, such as http://frodio.com/, http://myradio24.com/, http://myradio24.com/, http://www.radio-hoster.ru/, http://www.taghosting.ru/, http://www.economhost.com/, and even http://fmhosting.ru/. In late 2011, FMhost was taken down for a serious rewrite of the codebase and rethinking of the recommender system's architecture. In this paper, we describe the results of this upgrade.

The previous version of the recommender system experienced several problems, such as tag discrepancy or personal tracks without tags at all. A survey by FMhost with about a hundred respondents showed that more than half of them appreciated the previous version of our recommender system and more than 80% of the answers were positive or neutral (see Table 1); nevertheless, we hope that the new recommender model and algorithms provide even more accurate recommendations and make even less prediction mistakes.

Table 1. FMhost's recommender system satisfaction survey.

User opinion	Number of respondents (%)
I like it very much, all recommendations were relevant	54 (49%)
Good, I like most of the radio stations	22 (20%)
Sometimes there are interesting stations	16 (14%)
I like only few recommended radio stations	9(8%)
None of the recommended stations was satisfactory	10 (9%)

2.3 FMhost conceptual improvements

The new version features a more complex system of user interaction. Every radio station has an owner who is not just a name but also has the ability to assign DJs for lives, prepare radio schedule, and assign lives and programs. There will be a new broadcasting panel for DJs that will allow them to play tracks with additional features that were not available before, such as an equalizer or fading between tracks. A new algorithm for the recommender system, a new rating system, and a new chart system will be launched.

The rating system has been developed to rank radio stations and DJs according to their popularity and quality of work. A new core is being implemented and a new concept of LiquidSoap and Icecast is being designed. The system is designed so that to eliminate all problems that have surfaced in the previous version.

3 Models, algorithms and recommender architecture

3.1 Input data and general structure

Our model is based on three data matrices. The first matrix $A = (a_{ut})$ tracks the number of times user u visits radio stations with a certain tag t. Each radio station r broadcasts audio tracks with a certain set of tags T_r. The sets of all users, radio stations, and tags are denoted by U, R, and T respectively. The second matrix $B = (b_{rt})$ contains how many tracks with a tag t a radio station r has played. Finally, the third matrix $C = (c_{ur})$ contains the number of times a user u visits a radio station r. For each of these three matrices, we denote by v^A, v^B, and v^C the respective vectors containing sums of elements: $v^A = \sum\limits_{t \in T} a_{ut}$, $v^B = \sum\limits_{t \in T} b_{rt}$, and $v^C = \sum\limits_{r \in R} a_{ur}$. We also denote for each matrix A, B, C the corresponding frequency of visits matrix by A_f, B_f, and C_f; the frequency matrix is obtained by normalizing the matrix with the respective visits vector, e.g., $A_f = (a_{ut} \cdot (v^A_u)^{-1})$. Our model is not purely static; the matrices A, B, and C change after a user u visits a radio station r with a tag t, i.e., each value a_{ut}, b_{rt}, and c_{ur} is incremented by 1 after this visit.

The model consists of three main blocks: the Individual-Based Recommender System (IBRS) model, the Collaborative-Based Recommender System (CBRS) model, and the End Recommender Systems (ERS) that aggregates the results of the former two.

Each model has its own algorithmic implementation. Since both our previous works [23,24] and this work implicitly use biclustering ideas, we continue to name our general algorithms with the RecBi acronym; this time it is the RecBi3 family. We call the resulting algorithms for the three proposed models RecBi3.1, RecBi3.2, and RecBi3.3, respectively. Here we do not use notation from formal concept analysis, but refer to [25] for the basic notation used in our previous algorithms RecBi2.1 and RecBi2.2.

3.2 IBRS

The **IBRS** model uses matrices A_f and B_f and aims to provide a particular user $u_0 \in U$ with top N recommendations represented mathematically by a special structure $Top_N(u)$. Formally, $Top_N(u_0)$ is a triple $(R_{u_0}, \preceq_{u_0}, \mathrm{rank})$, where R_{u_0} is the set of at most N radio stations recommended to a particular user u_0, $\preceq_{u_0}$ is a well-defined quasiordering (reflexive, transitive, and complete) on the set R_{u_0}, and rank is a function which maps each radio station r from R_{u_o} to $[0, 1]$.

The **RecBi3.1** algorithm computes the 1-norm distance between a user u_0 and a radio station r, i.e.m $d(u_0, r) = \sum t \in T|a_{u_0t} - b_{rt}|$. Then all distances between the user u_0 and the radio stations $r \in R$ are calculated. Further the algorithm constructs the relation $\prec_{u_0}$ according to the following rule: $r_i \preceq r_j$ iff $d(u_0, r_i) \leq d(u_0, r_i)$. The function rank operates on R_{u_0} according to the following rule:

$$\mathrm{rank}(r_i) = 1 - d(u_0, r_i) / \max_{r_j \in R} d(u_0, r_j).$$

Finally, after selecting N radio stations for N greatest rank values in the set R_{u_0}, we have the structure $Top_N(u_0)$ which represents a ranked list of radio stations recommended to the user u_0.

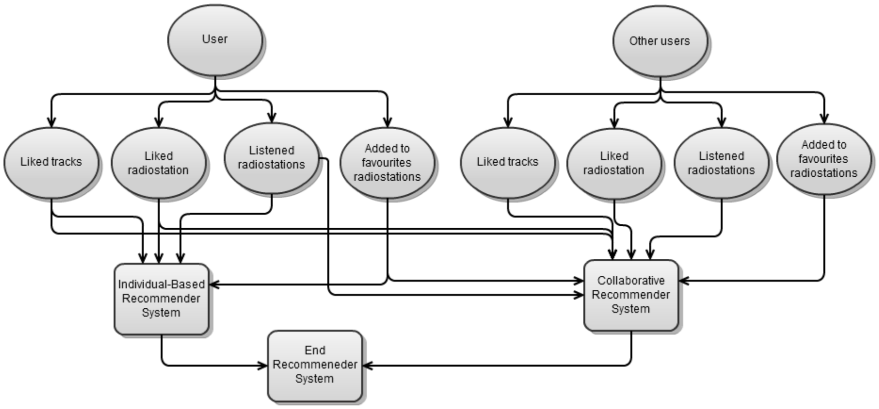


Fig. 1. The recommender system architecture

As shown in Fig. 1, our model takes into account not only listened tracks but also liked tracks, liked radio stations, and favorite radio stations. To refine the IBRS submodel we tune it with the SMARTS algorithm known from decision making theory [26]. According to the method and expert decisions, we should count each track tag of a listened radio station, liked radio station, liked track, and favorite radio station with a different weight. The SMARTS procedure provided us with the four weights for listened radio station, liked radio station, favorite radio station, and liked track according to our experts' assessment of mutual criterion importance, namely 0.07, 0.16, 0.3, and 0.47. In the SMARTS method, we consider each tag type as a criterion with two terminal values 0 and 100% on a real number scale. Some tag t may have some or even four of these types simultaneously; in this case, the algorithm adds to a_{ut} the total weight of the tag (i.e., the sum of weights) after a user u visits some radio station with this tag. In case there are several elements with the same rank so that $Top_N(u)$ is not uniquely defined, we simply choose the first elements according to some arbitrary ordering (e.g., the lexicographic ordering of station names).

3.3 CBRS

The **CBRS** model is based on the C_f matrix. The matrix also yields a vector n^C which stores the total number of listened stations for each user $u \in U$. This vector also changes over time, and this value is used as a threshold to transform matrix C_f to distance matrix D as follows:

$$d_{ij} = \begin{cases} |c_{fir} - c_{fjr}|, & if\ c_{fir} \geq n_i^{-1} \text{and } c_{fjr} \geq n_j^{-1} \\ |c_{fir} + c_{fjr}|, & if\ c_{fir} > n_i^{-1} \text{and } c_{fjr} < n_j^{-1} \text{or vice versa} \end{cases} \quad (1)$$

This distance takes into account the frequency n_u^C of all radio station visits for user u and considers its inverse value as a threshold to decide whether a particular station r should be considered as popular for this user. Thus, users with different signs of $c_{fir} - n_i^{-1}$ and $c_{fjr} - n_j^{-1}$ become more distant than for the conventional absolute distance. This distance d_{ij} actually serves as a sort of polarizing filter, and in Section 4 we compare it with common approaches.

After computing D, the algorithm **RecBi3.2** constructs the list $Top_k(u_0) = (U_{u_0}, \preceq_{u_0}, \text{sim})$ of k users similar to our target user u_0 who awaits recommendations, where $\text{sim}(u) = 1 - d_{uu_0} / \max\limits_{u' \in U} d_{u'u_0}$. We define the set of all radio stations user u_0 listened to as $L(u_0) = \{r | c_{fur} = 0\}$. In a similar way, we define

$$\begin{aligned} Top_N(u_0) &= (R_{u_0}, \preceq_{u_0}, \text{rank}), \text{ where} \\ \text{rank}(r) &= \text{sim}(u^*) \cdot c_{fu^*r} \text{ and} \\ u^* &= \arg \max_{u \in U_{u_0}, r \in U / L(u_0)} \text{sim}(u) \cdot c_{fur}. \end{aligned}$$

It is worth mentioning that rank : $r \mapsto [0, 1]$. The problem of choosing exactly N topmost stations is solved in the same way as in the IBRS submodel.

3.4 ERS

After IBRS and CBRS have finished, we are left with two ranked lists of recommended stations $Top_N^I(u_0)$ and $Top_N^C(u_0)$ for our target user u_0 from IBRS and CBRS respectively. The **ERS** submodel proposes a simple solution for aggregating these lists into the final recommendation structure $Top_N^E(u_0) = (R_{u_0}^E, \preceq_{u_0}^E, \text{rank}^E)$. For every $r \in R_{u_0}^C \cup R_{u_0}^I$, the function $\text{rank}^E(r)$ maps r to the weighted sum

$$\beta \cdot \text{rank}^C(r) + (1 - \beta) \cdot \text{rank}^I(r),$$

where we let $\beta \in$ [0,1], $\text{rank}^C(r) = 0$ for all $r \notin R^C$ and $\text{rank}^I(r) = 0$ for all $r \notin R^I$. The algorithm **RecBi3.3** adds the best N radio stations according to this criterion to the set $R_{u_0}^C$.

4 Quality of service assessment

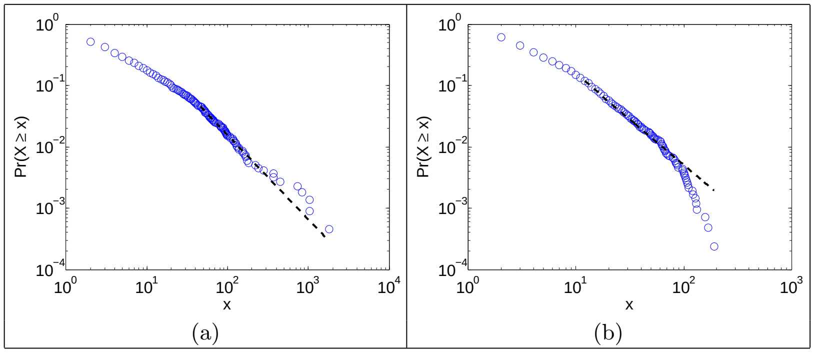
To evaluate the quality of the developed system, we propose a variant of the cross-validation technique [27]. Before we proceed to the detailed description of the procedure, we discuss some important analyses that we conducted on the FMhost data for the period from 2009 till 2011.

4.1 Basic statistics

It is a well-known fact that social networking data often follows the so called power law distribution [28]. To decide which amount of active users or radio stations we have to take into account for making recommendations, we performed a simple statistical analysis of user and radio station activity. Around 20% of the users (only registered ones) were analysed.

Table 2. Basic parameters of the user and radio visits datasets, along with their power-law fits and the corresponding *p-value* .

Dataset	n	$\langle x \rangle$	σ	x_{max}	$\hat{x}_{min}$	$\hat{\alpha}$	n_{tail}	*p-value*
User dataset	4187	5.86	12.9	191	12 ± 2	2.46(0.096)	117	**0.099**
Radio dataset	2209	11.22	60.05	1817	46 ± 11	2.37(0.22)	849	**0.629**



(a)

(b)

Fig. 2. Cumulative distribution functions $P(x)$ and their maximum likelihood power-law fits for the FMhost two empirical data sets. (a) The frequency distribution of radio station visits. (b) The frequency of visits of unique users.

Table 2 shows p-values of statistical tests, which were performed by means of Matlab tools from [28], show that the power law does fit the radio station dataset, and the probability to make an error by ruling out the null hypothesis (no power law) is about 0.1 for the user dataset. Thus, the radio station visits dataset is more likely to follow the power law than the user visits dataset, but we should take it into account for both datasets; Fig. 2 shows how the power law actually fits our data.

This analysis implies useful consequences according to the well-known "80:20" rule:

$$W = P^{(\alpha-2)/(\alpha-1)},$$

which means that the fraction W of the wealth is in the hands of the richest P of the population. In our case, 50% of users make 80% of all radio station visits, and 50% of radio stations have 83% of all visits. Thus, if the service tends to take into account only active stations and users, it can cover 80% of all visits by considering only 50% of their active audience. However, new radio stations still deserve to be recommended, so this rule can only be applied to the user database.

4.2 Quality assessment

To evaluate QoS for the IBRS subsystem (RecBi3.1 algorithm), we count average precision and recall on the set $R_N \subset R$, where N is a number of randomly "hidden" radiostations. We suppose that for all r in R_N and every user $u \in U$ the algorithm does not know whether the radio stations were liked, added to favorites, or even visited, and we change A_f and R accordingly. Then RecBi3.1 attempts to recommend Top-N radio stations for this modified matrix A_f.

Top-N average precision and recall are computed as follows:

$$\text{Precision} = \frac{\sum\limits_{u\in U} \frac{|R_u^I \cap L_u \cap R_N|}{|L_u \cap R_u^I|}}{|U|},$$

$$\text{Recall} = \frac{\sum\limits_{u\in U} \frac{|R_u^I \cap L_u \cap R_N|}{|L_u \cap R_N|}}{|U|}.$$

To deal with CBRS, we use a modification of the leave-one-out technique. At each step of the procedure for a particular user u, we "hide" all radio stations $r \in R_N$ by setting $c_{fur} = 0$. Then we perform RecBi3.2 assuming that $c_{fu'r}$ is unchanged for $u' \in U/u$. After that we compute

$$\text{Precision} = \frac{\sum\limits_{u\in U} \frac{|R_u^C \cap L_u \cap R_N|}{|L_u \cap R_u^C|}}{|U|},$$

$$\text{Recall} = \frac{\sum\limits_{u\in U} \frac{|R_u^C \cap L_u \cap R_N|}{|L_u \cap R_N|}}{|U|}.$$

To tune the ERS system, we can use a combination of these two procedures trying to find the optimal β as

$$\beta^* = \arg\max_{\beta} \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{(\text{Precision} + \text{Recall})}$$

.

We suppose that in one month of active operation we will have enough statistics to tune β and choose appropriate similarity and distance measures as well as thresholds. We suppose that the resulting system will provide reasonably accurate recommendations using only a single (last) month of user history and only 50% of the most active users. For quality assessment during the actual operation, we will compute Top-3, Top-5, and Top-10 Precision and Recall measures as well as whether the system provides a user only with Top-10 items with a highest rank. In addition, online surveys can be launched to assess user satisfaction with the new RS system.

5 Conclusion and further work

In this work, we have described the underlying models, algorithms, and the system architecture of the new improved FMhost service. We hope that the developed algorithms will help a user to find relevant radio stations to listen to. In future optimization and tuning, special attention should be paid to scalability issues and user-centric quality assessment. We consider matrix factorization techniques as a reasonable tool to increase scalability, but it has to be carefully adapted and assessed taking into account the folksonomic nature of tracks tags. Another attractive feature of the developed system is that it can serve as a kind of World of Music map built on track-to-track similarity matrices with tags [7]. Another important issue is dealing with the triadic relational nature of data (users, radio stations (tracks), and tags), which constitutes the so called *folksonomy* [29], a primary data structure in tagging resource-sharing systems. As shown in [30], this data can be successfully mined by means of triclustering, so we also plan to build a tag-based recommender system by means of triclustering.

Acknowledgments. We would like to thank Rustam Tagiew, Jonas Poelmans and Mykola Pechenizkiy for their comments, remarks and explicit and implicit help during paper preparations. Work of Sergey Nikolenko has been supported by the Russian Foundation for Basic Research grant 12-01-00450-a, the Russian Presidential Grant Programme for Young Ph.D.'s, grant no. MK-6628.2012.1, for Leading Scientific Schools, grant no. NSh-3229.2012.1, and RFBR grants 11-01-12135-ofi-m-2011 and 11-01-00760-a.

References

1. Klapuri, A., Leider, C., eds.: Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011, Miami, Florida, USA, October 24-28, 2011. In Klapuri, A., Leider, C., eds.: ISMIR, University of Miami (2011)
2. Anglade, A., Baccigalupo, C., Casagrande, N., Celma, Ò., Lamere, P.: Workshop report: Womrad 2010. In Amatriain, X., Torrens, M., Resnick, P., Zanker, M., eds.: RecSys, ACM (2010) 381–382
3. Anglade, A., Celma, O., Fields, B., Lamere, P., McFee, B.: Womrad: 2nd workshop on music recommendation and discovery. In: Proceedings of the fifth ACM conference on Recommender systems. RecSys '11, New York, NY, USA, ACM (2011) 381–382

4. RecSys '11: Proceedings of the fifth ACM conference on Recommender systems, New York, NY, USA, ACM (2011) 609116.
5. Hilliges, O., Holzer, P., Klüber, R., Butz, A.: Audioradar: A metaphorical visualization for the navigation of large music collections. In Butz, A., Fisher, B.D., Krüger, A., Olivier, P., eds.: Smart Graphics. Volume 4073 of Lecture Notes in Computer Science., Springer (2006) 82–92
6. Gleich, D.F., Rasmussen, M., Lang, K., Zhukov, L.: The world of music: User ratings; spectral and spherical embeddings; map projections. Online report (2006)
7. Gleich, D.F., Zhukov, L., Rasmussen, M., Lang, K.: The World of Music: SDP Embedding of High Dimensional data. In: Information Visualization 2005. (2005) Interactive Poster.
8. Brandenburg, K., Dittmar, C., Gruhne, M., Abeer, J., Lukashevich, H., Dunker, P., Grtner, D., Wolter, K., Grossmann, H.: Music search and recommendation. In Furht, B., ed.: Handbook of Multimedia for Digital Entertainment and Arts. Springer US (2009) 349–384
9. Celma, Ò.: Music Recommendation and Discovery - The Long Tail, Long Fail, and Long Play in the Digital Music Space. Springer (2010)
10. Avesani, P., Massa, P., Nori, M., Susi, A.: Collaborative radio community. In: Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems. AH '02, London, UK, UK, Springer-Verlag (2002) 462–465
11. Symeonidis, P., Ruxanda, M.M., Nanopoulos, A., Manolopoulos, Y.: Ternary semantic analysis of social tags for personalized music recommendation. In Bello, J.P., Chew, E., Turnbull, D., eds.: ISMIR. (2008) 219–224
12. Nanopoulos, A., Rafailidis, D., Symeonidis, P., Manolopoulos, Y.: Musicbox: Personalized music recommendation based on cubic analysis of social tags. IEEE Transactions on Audio, Speech & Language Processing **18**(2) (2010) 407–412
13. Koenigstein, N., Dror, G., Koren, Y.: Yahoo! music recommendations: modeling music ratings with temporal dynamics and item taxonomy. In: Proceedings of the fifth ACM conference on Recommender systems. RecSys '11, New York, NY, USA, ACM (2011) 165–172
14. Celma, O., Lamere, P.: Music recommendation and discovery revisited. In: Proceedings of the fifth ACM conference on Recommender systems. RecSys '11, New York, NY, USA, ACM (2011) 7–8
15. Hu, Y., Ogihara, M.: Nextone player: A music recommendation system based on user behavior. [1] 103–108
16. Bogdanov, D., Herrera, P.: How much metadata do we need in music recommendation? a subjective evaluation using preference sets. [1] 97–102
17. Mesnage, C.S., Rafiq, A., Dixon, S., Brixtel, R.P.: Music discovery with social networks. In: Workshop on Music Recommendation and Discovery 2011, Workshop on Music Recommendation and Discovery (October 2011) 1–6
18. Barthet, M., Anglade, A., Fazekas, G., Kolozali, Sefki Macrae, R.: Music recommendation for music learning: Hotttabs, a multimedia guitar tutor. In: Workshop on Music Recommendation and Discovery 2011, Workshop on Music Recommendation and Discovery (October 2011) 7–13
19. Tatl, I., Birturk, A.: Using semantic relations in context-based music recommendations. In: Workshop on Music Recommendation and Discovery 2011, Workshop on Music Recommendation and Discovery (October 2011) 14–17
20. Knees, P., Schedl, M.: Towards semantic music information extraction from the web using rule patterns and supervised learning. In: Workshop on Music Recommen-

dation and Discovery 2011, Workshop on Music Recommendation and Discovery (October 2011) 18–25

21. Knopke, I.: The importance of service and genre in recommendations for online radio and television programmes. In: Workshop on Music Recommendation and Discovery 2011, Workshop on Music Recommendation and Discovery (October 2011) 26–29
22. Popescu, G., Pu, P.: Probabilistic game theoretic algorithms for group recommender systems. In: Workshop on Music Recommendation and Discovery 2011, Workshop on Music Recommendation and Discovery (October 2011) 7–12
23. Ignatov, D., Poelmans, J., Zaharchuk, V.: Recommender System Based on Algorithm of Bicluster Analysis RecBi. In Ignatov, D., Poelmans, J., Kuznetsov, S., eds.: CEUR Workshop proceedings Vol-757, CDUD'11 - Concept Discovery in Unstructured Data. (2011) pp. 122–126
24. Ignatov, D.I., Kuznetsov, S.O.: Concept-based Recommendations for Internet Advertisement. In Belohlavek, R., Kuznetsov, S.O., eds.: Proc. CLA 2008. Volume Vol. 433 of CEUR WS., Palack University, Olomouc, 2008 (2008) 157–166
25. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. 1st edn. Springer-Verlag New York, Inc., Secaucus, NJ, USA (1999)
26. Edwards, W., Barron, F.: SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement. Organizational Behavior and Human Decision Processes **60**(3) (1994) 306 – 325
27. Ignatov, D.I., Poelmans, J., Dedene, G., Viaene, S.: A New Cross-Validation Technique to Evaluate Quality of Recommender Systems. In Kundu, M.K., Mitra, S., Mazumdar, D., Pal, S.K., eds.: PerMIn. Volume 7143 of LNCS., Springer (2012) 195–202
28. Clauset, A., Shalizi, C.R., Newman, M.E.J.: Power-law distributions in empirical data. SIAM Rev. **51**(4) (November 2009) 661–703
29. Vander Wal, T.: Folksonomy Coinage and Definition (2007) http://vanderwal.net/folksonomy.html (accessed on 12.03.2012).
30. Ignatov, D.I., Kuznetsov, S.O., Magizov, R.A., Zhukov, L.E.: From Triconcepts to Triclusters. In: Proceedings of the 13th international conference on Rough sets, fuzzy sets, data mining and granular computing. RSFDGrC'11, Berlin, Heidelberg, Springer-Verlag (2011) 257–264

Mining Determinism in Human Strategic Behavior

Rustam Tagiew

Institute for Computer Science of TU Bergakademie Freiberg, Germany
tagiew@informatik.tu-freiberg.de

Abstract. This work lies in the fusion of experimental economics and data mining. It continues author's previous work on mining behavior rules of human subjects from experimental data, where game-theoretic predictions partially fail to work. Game-theoretic predictions aka equilibria only tend to success with experienced subjects on specific games, what is rarely given. Apart from game theory, contemporary experimental economics offers a number of alternative models. In relevant literature, these models are always biased by psychological and near-psychological theories and are claimed to be proven by the data. This work introduces a data mining approach to the problem without using vast psychological background. Apart from determinism, no other biases are regarded. Two datasets from different human subject experiments are taken for evaluation. The first one is a repeated mixed strategy zero sum game and the second – repeated ultimatum game. As result, the way of mining deterministic regularities in human strategic behavior is described and evaluated. As future work, the design of a new representation formalism is discussed.

Key words: Game Theory, Psychology, Data Mining, Artificial Intelligence, Domain-Specific Languages

1 Introduction

Game theory is one of many scientific disciplines predicting outcomes of social, economical and competitive interactions among humans on the granularity level of individual decisions [1, p.4]. People are assumed to be autonomous and intelligent, and to decide according to their preferences. People can be regarded as rational, if they always make decisions, whose execution has according to their subjective estimation the most preferred consequences [2,3]. The correctness of subjective estimation depends on the level of intelligence. Rationality can justify own decisions and predictions of other people's decisions. If interacting people satisfy the concept of rationality and apply mutually and even recursively this concept, then the interaction is called strategic interaction (SI). Further, game is a notion for the formal structure of a concrete SI [4]. A definition of a game consists of a number of players, their legal actions and players' preferences. The preferences can be replaced by a payoff function under assumed payoff maximization. The payoff function defines each player's outcome depending on his actions, other players' actions and random events in the environment. The game-theoretic solution of a game is a prediction about the behavior of the players aka an equilibrium. The assumption of rationality is the basis for an equilibrium. Deviating from an equilibrium is beyond rationality, because it does not maximize the payoff. Not every game has an equilibrium. However, there is at least one mixed strategies equilibrium (MSE) in finite games [5].

The notion of game is commonly used for pleasant time spending activities like board games, but can also be extended to all social, economical and competitive interactions among humans. A board game can have the same game structure as a war. Some board games are even developed to train people, like Prussian army war game ”Kriegspiel Chess” [6] for officers. We like it to train ourselves in order to perform better in games [7]. In most cases, common human behavior in games deviates from game-theoretic predictions [8,9]. One can say without any doubt that if a human player is trained in a concrete game, he will perform close to equilibrium. But, a chess master does not also play poker perfectly and vice versa. On the other side, a game-theorist can find a way to compute an equilibrium for a game, but it does not make a successful player out of him. There are many games we can play; for most of them, we are not trained. That is why it is more important to investigate our behavior while playing general games than playing a concrete game on expert level. Conducting experiments for gathering data of human game playing is called experimental economics.

Although general human preferences are a subject of philosophical discussions [10], game theory assumes that they can be captured as required for modeling rationality. Regarding people as rational agents is disputed at least in psychology, where even a scientifically accessible argumentation exposes the existence of stable and consistent human preferences as a myth [11]. The problems of human rationality can not be explained by bounded cognitive abilities only. ”British people argue that it is worth spending billions of pounds to improve the safety of the rail system. However, the same people habitually travel by car rather than by train, even though traveling by car is approximately 30 times more dangerous than by train!”[12, p.527–530] Since the last six decades nevertheless, the common scientific standards for econometric experiments are that subjects’ preferences over outcomes can be insured by paying differing amounts of money [13]. However, insuring preferences by money is criticized by the term ”Homo Economicus” as well.

The ability of modeling other people’s rationality and reasoning as well corresponds with the psychological term ”Theory of Mind” (ToM) [14], which lacks almost only in the cases of autism. For experimental economics, subjects as well as researchers, who both are supposed to be non-autistic people, may fail in modeling of others’ minds anyway. In Wason task at least, subjects’ reasoning does not match the researchers’ one [15]. Human rationality is not restricted to capability for science-grade logical reasoning – rational people may use no logic at all [16]. However, people also mistake seriously in the calculus of probabilities [17]. In mixed strategy games, the required sequence of random decisions can not be properly generated by people [18]. Due to bounded cognitive abilities, every ”random” decision depends on previous ones and is predictable in this way. In ultimatum games [9, S. 43ff], the economists’ misconception of human preferences is revealed – people’s minds value fairness additionally to personal enrichment. Our minds originated from the time, when private property had not been invented and social values like fairness were essential for survival.

This work concentrates on human playing of general games and continues author’s previous work [19]. It is about the common human deviations from predicted equilibria in games, for which they are not trained or experienced. The two examples introduced in this work are a repeated mixed strategy zero sum game and a repeated ultimatum game from responders' perspective. The only assumption is the existence of deterministic rules in human behavior. Under this assumption, diverse data mining algorithms are evaluated. Apart from mining deterministic regularities, modeling human behavior in general games needs a representation formalism which is not specific to a concrete game. Representing human behavior models in such a formalism would increase their comparability. Therefore, this paper includes a general formalism discussion, where results from the evaluation are involved.

The next section summarizes related work on a formalism for human behavior in games. Then, the data mining approach on datasets is presented afterwards. Summary and discussion conclude this paper.

2 Related Work

A very comprehensive gathering of works in experimental psychology and economics on human behavior in general games can be found in [9]. This work inspired research in artificial intelligence [20], which led to the creation of network of influence diagrams (NID) as a representation formalism. NID is a formalism similar to the possible worlds semantics of Kripke models [21] and is a super-set of Bayesian games. The main idea of NID is modeling human reasoning patterns in diverse SIs. Every node of a NID is a multi-agent influence diagram (MAID) representing a model of SI of an agent. MAID is an influence diagram (ID), where every decision node is associated with an agent. ID is a Bayesian network (BN), where one has ordinary nodes, decision nodes and utility nodes. In summary, this approach assumes that human decision making can be modeled using BN – human reasoning is assumed to have a non-deterministic structure. This formalism is already applied for modeling reciprocity in a repeated ultimatum game called ”Colored Trails” (CT) [22]. The result of this work is that models of adaptation to human behavior based on BNs perform better than standard game theoretical algorithms.

Another independent work is an application of a cognitive architecture from psychology to games [23]. A cognitive architecture is a formalism concerned to represent general human reasoning [24] in order to compare different models. Today's most popular cognitive architecture is ACT-R (Adaptive Control of Thought Rational) [25]. In comparison to NID, ACT-R is used for a number of psychological studies. ACT-R consists of two tiers – symbolic and sub-symbolic. On the symbolic tier, there are chunks – facts and ”If-Then”-rules. On the sub-symbolic tier, there are exponential functions, which determine activation levels of chunks, delays in reasoning and priorities between rules. Based on ACT-R, an almost deterministic model for a mixed strategy zero sum game ”Rock Paper Scissors” (RPS) is designed. The only case, in which the designed model predicts random behavior is the beginning of a game sequence. The model was successfully evaluated as a base for an artificial player, which won against human subjects.

Whether deterministic or not, both works follow the same approach. First, they construct a model, which is based on theoretical considerations. Second, they adjust the parameters of this model to the experimental data. This makes the human behavior explainable using the concepts from the model, but needs a priori knowledge to construct the model.

3 Used Datasets

The first dataset chosen for our data mining approach has already been mentioned in our previous work [19]. It is the game RPS played over a computer network. This game is easy to explain and most people do not train to play it on expert level; it is symmetric, zero sum and two player. The study was conducted on threads of 30 one-shot games. A player had a delay for consideration of 6 sec for every shot. If he did not react, the last or default gesture was chosen. A thread lasted $30 * 6$ sec $= 3$ min. This game has one mixed strategy equilibrium (MSE), which is an equal probability distribution between the three gestures. At least, one can not lose playing this MSE.

Ten computer science undergraduates were recruited. They were in average $22,7$ years old and 7 of them were male. They had to play the thread twice against another test person. Between the two threads, they played other games. In this way, 300 one-shot games or 600 single human decisions are gathered. Every person got €0.02 for a won one-shot game and €0.01 for a draw. The persons, who played against each other, sat in two separate rooms. One of the players used a cyber-glove and the other one a mouse as input for gestures. The graphical user interface showed the following information - own last and actual choice, opponents last choice, a timer and already gained money. According to statements of the persons, they had no problems to understand the game rules and to choose a gesture timely. All winners and 80% of losers attested that they had fun to play the game.

The second dataset is the recorded responder behavior from the CT experiment [22]. This dataset contains 371 single human decisions of 25 participating subjects. A positive decision of the responder updates the monetary payoff of both players, while a negative one does not change anything. The payoff update varied between $1.45 and $-1.35 for the responder. In 160 cases, responders update was zero. The equilibrium for the responder is to accept only proposals, which increase his payoff regardless of the proposer's payoff.

4 Methods

Statistical analysis of the datasets from the previous chapter exposed that the human behavior observed in the experiments can not be explained using only game theory [1]. The shape of equilibrium deviations confirms the one reported in relevant literature [9]. The goal is to find a model beyond game theory for the prediction of average deviations. In related work, the creation of a sophisticated model preceded the evaluation on the data. In this work, the evaluation on the data precedes the creation of a model. Of course, some people would not match into such a model like trained or somehow experienced individuals. Prediction of specific individuals is not addressed in this paper.

Machine done prediction without participation in game playing with human subjects should not be confused with prediction algorithms of artificial players. Quite the contrary, artificial players can manipulate the predictability of human subjects by own behavior. For instance, an artificial player, which always throws ”Paper” in RPS, would success at predicting a human opponent always throwing ”Scissors” in reaction. Otherwise, if an artificial player maximizes its payoff based on opponent modeling, it would face a change in human behavior and have to handle that. This case is more complex than a spectator prediction model for an ”only-humans” interaction. This paper restricts on a prediction model without participating.

Human behavior can be modeled as either deterministic or non-deterministic. Although human subjects fail at generating truly random sequences as demanded by MSE, non-deterministic models are especially used in case of artificial players in order to handle uncertainties. ”Specifically, people are poor at being random and poor at learning optimal move probabilities because they are instead trying to detect and exploit sequential dependencies. ... After all, even if people do not process game information in the manner suggested by the game theory player model, it may still be the case that across time and across individuals, human game playing can legitimately be viewed as (pseudo) randomly emitting moves according to certain probabilities.” [23] In the addressed case of spectator prediction models, non-deterministic view can be regarded as too shallow, because deterministic models allow much more exact predictions. Non-deterministic models are only useful in cases, where a proper clarification of uncertainties is either impossible or costly. To remind, deterministic models should not be considered to obligatory have a formal logic shape.

The deterministic function $\mathrm{HD}(\mathrm{Game},\mathrm{History}) \rightarrow \mathrm{Decision}$ denotes a human decision. $History$ denotes the previous turns in the game. *Game* and $History$ are the input and *Decision* – the output. Finding a hypothesis, which matches the regularity between input and output without a priori knowledge, is a typical problem called supervised learning [26]. There is already a big amount of algorithms for supervised learning. Each algorithm has its own hypothesis space (HS). For a Bayesian learner, e.g., the hypothesis space is the set of all possible Bayesian networks. There are many different types of hypothesis spaces - rules, decision trees, Bayesian models, functions and so on. Concrete hypothesis HD^I is a relationship between input and output described by using the formal means of the corresponding hypothesis space.

Which hypothesis space is most appropriate to contain valid hypotheses about human behavior? This is a machine learning version of the question about a formalism for human behavior. The most appropriate hypothesis space contains the most correct hypothesis for every concrete example of human behavior. A correct hypothesis does not only perform well on the given data (training set), but it performs also well on new data (test set). Further, it can be assumed that the algorithms which choose a hypothesis perform alike well for all hypothesis spaces. For instance, a decision tree algorithm creates a tree, a neuronal algorithm creates a neuronal network and the distance between the created tree to the best possible tree is the same as the distance between the created neuronal network and the best possible neuronal network. This assumption is a useful simplification of the problem for a preliminary demonstration. Using it, one can consider the algorithm with the best performance on the given data as the algorithm with the most appropriate hypothesis space. The standard method for measurement of performance of a machine learning algorithm or also a classifier is cross validation.

As it is already mentioned, a machine learning algorithm has to find hypothesis HD^{I} which matches best the real human behavior function HD. Human decision making depends mostly on a small part of the history due to bounded resources. This means that one needs a simplification function $\mathrm{S}(History) \rightarrow$ Pattern. Using function S the function $\mathrm{HD}(X,Y)$ is to be approximated through $\mathrm{HD}^{II}(X,S(Y))$. The problem for finding the most appropriate hypothesis can be formulated in equation 1. The function match in equation 1 is considered to be implemented through a cross validation run.

$$\arg\max_{\mathrm{HS}}(\max_{\mathrm{HD}^{II}\in\mathrm{HS}}(\mathrm{match}(HD(X,Y),\mathrm{HD}^{II}(X,S(Y)))))\qquad(1)$$

5 Empirical Results

The first dataset is transformed to a sets of tuples, each one consists of three own previous gestures, three opponent previous gestures and own next gesture. Therefore, every tuple has the length $3+3+1=7$. The simplification function is a window over three last turns. There are 2187 possible tuples for RPS. The decisions in the first three turns of game are not considered. Therefore, the size of the set results to 540 tuples. The second dataset is also transformed to a sets of 371 tuples, where every tuple includes the proposers payoff update, the responders payoff update and the responders boolean reply.

Implementations of classifiers provided by WEKA [27] are used for the cross validation on the both sets of tuples. For the first dataset, there are currently 45 classifiers available in the WEKA library, which can handle multi-valued nominal classes. Gestures in RPS are nominal, because there is no order between them. These classifiers belong to different groups - rule-based, decision trees, function approximators, baysian learners, instance-based and miscellaneous. A cross validation of all 45 classifiers on RPS dataset is performed. For the CT dataset, a cross-validation of 35 appropriate classifiers is performed. The number of subsets for cross-validation is 10. Both cross-validation runs are conducted with preserving order of the tuples.

Sequential minimal optimization (SMO) [28] showed 46.48% prediction correctness, which is about 1% higher than the sophisticated non-deterministic model for RPS of Warglen [29]. Unfortunately, decreasing and increasing the window size in the function S for the RPS dataset diminishes the performance. Using the single rule classifier (OneR), one can find out that 43.15% of the RPS dataset matches the rule: ”Choose paper after rock, scissors after paper and rock after scissors”. A number of classifiers including SMO achieve 95.42% correctness on the CT dataset in cross-validation. One of this algorithms is based on decission tables [30]. This algorithm finds out that 95.15% of the CT dataset conforms the rule: ”If an acceptance does neither change your payoff nor improve the proposers payoff, then refuse!” This result overperforms clearly the 72% reported from the non-deterministic approach of Pfeffer [22].

6 Conclusion

The strategic behavior consists out of the observable actions, whose origins are tried to be understood as generally as possible. Summarizing the results of this work, it can be said that SMO can find the most general deterministic hypothesis about regularities of human behavior in the investigated scenarios. The correctness of such a hypothesis overperforms the numbers reported in related work. The hypothesis space of SMO is one of complex functions and can be used for the design of a game behavior description formalism.

References

1. Tagiew, R.: Strategische Interaktion realer Agenten: Ganzheitliche Konzeptualisierung und Softwarekomponenten einer interdisziplinren Forschungsinfrastruktur. PhD thesis, TU Bergakademie Freiberg (2011)
2. Russel, S., Norvig, P.: Artificial Intelligence. Pearson Education (2003)
3. Osborne, M.J., Rubinstein, A.: A course in game theory. MIT Press (1994)
4. Morgenstern, O., von Neumann, J.: Theory of Games and Economic Behavior. Princeton University Press (1944)
5. Nash, J.: Non-cooperative games. Annals of Mathematics (54) (1951) 286 – 295
6. Li, D.H.: Kriegspiel: Chess Under Uncertainty. Premier (1994)
7. Genesereth, M.R., Love, N., Pell, B.: General game playing: Overview of the aaai competition. AI Magazine **26**(2) (2005) 62–72
8. Pool, R.: Putting game theory to the test. Science **267** (1995) 1591–1593
9. Camerer, C.F.: Behavioral Game Theory. Princeton University Press (2003)
10. Stevenson, L., Haberman, D.L.: Ten Theories of Human Nature. OUP USA (2004)
11. Bazerman, M.H., Malhotra, D.: Economics wins, psychology loses, and society pays. In De Cremer, D., Zeelenberg, M., Murnighan, J.K., eds.: Social Psychology and Economics. Lawrence Erlbaum Associates (2006) 263–280
12. Eysenck, M.W., Keane, M.T.: Cognitive Psychology: A Student's Handbook. Psychology Press (2005)
13. Chamberlin, E.H.: An experimental imperfect market. Journal of Political Economy **56** (1948) 95–108
14. Verbrugge, R., Mol, L.: Learning to apply theory of mind. Journal of Logic, Language and Information **17** (2008) 489–511
15. Wason, P.C.: Reasoning. In Foss, B.M., ed.: New horizons in psychology. Penguin Books (1966) 135–151
16. Oaksford, M., Chater, N.: The probabilistic approach to human reasoning. Trends in Cognitive Sciences **5** (2001) 349–357
17. Kahneman, D., Slovic, P., Tversky, A.: Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press (1982)
18. Kareev, Y.: Not that bad after all: Generation of random sequences. Journal of Experimental Psychology: Human Percetion and Performance **18** (1992) 1189–1194
19. Tagiew, R.: Hypotheses about typical general human strategic behavior in a concrete case. In: AI*IA, Springer (2009) 476–485
20. Gal, Y., Pfeffer, A.: A language for modeling agents' decision making processes in games. In: AAMAS, ACM Press (2003) 265–272
21. Kripke, S.: Semantical considerations on modal logic. Acta Philosophica Fennica **16** (1963) 83–94

22. Gal, Y., Pfeffer, A.: Modeling reciprocal behavior in human bilateral negotiation. In: AAAI, AAAI Press (2007) 815–820
23. Rutledge-Taylor, M.F., West, R.L.: Cognitive modeling versus game theory: Why cognition matters. In: ICCM. (2004) 255–260
24. Gluck, K.A., Pew, R.W., Young, M.J.: Background, structure, and preview of the model comparison. In Gluck, K.A., Pew, R.W., eds.: Modeling Human Behavior with Integrated Cognitive Architectures. Lawrence Erlbaum Associates (2005) 3–12
25. Taatgen, N., Lebiere, C., Anderson, J.: Modeling paradigms in act-r. 29–52
26. Mitchell, T.M.: Machine Learning. McGraw-Hill Higher Education (1997)
27. Witten, I.H., Frank, E.: Data Mining. Morgan Kaufmann (2005)
28. Platt, J.C.: Fast training of support vector machines using sequential minimal optimization. In: Advances in Kernel Methods - Support Vector Learning, MIT Press (1999) 185–208
29. Marchiori, D., Warglien, M.: Predicting human interactive learning by regret-driven neural networks. Science **319** (2008) 1111–1113
30. Kohavi, R.: The power of decision tables. In: ECML, Springer (1995) 174–189

Criteria Formation of Effective High-School Graduates Employment Based upon Data Mining Methods

I. Bolodurina, Y. Akhmayzyanova

Orenburg State University, Orenburg, Russia

yuliyanova@yandex.ru

Abstract. This work is devoted to the investigation of the main factors influencing high-school graduates' employment. This investigation is conducted using Data Mining methods. Two approaches to the identification of hidden patterns in data are employed. The first method is decision trees and the second method is production rules extraction. Both methods are combined and compared to each other.

Keywords: Data Mining, patterns identification, criteria, effective graduates employment.

Image Processing Using Dynamical NK-Networks Consisting of Binary Logical Elements

Daria M. Puchkova

Institute of cybernetics, informatics and communication,
Tyumen State Oil and Gas University,
Chervishevsky Trakt str. 13, Tyumen, 625008, Russia

dpuchkova@gmail.com

Abstract. In this paper a new method for image analysis is proposed, which uses a three dimensional neural network consisting of binary logical elements. The training process is divided into periods and a unique feature of this network is its self-organization capabilities which can be observed after the first period. Unique features of the initial image can be identified after a half-period. We studied these self-organizing properties of the NK-network consisting of binary logical elements. We also investigated the possibilities of applying such a system to image processing. Finally, peculiar features and advantages of the proposed method were discussed. The described method for image processing can be applied in the area of security informatics for signature authentication and detection of fraud even if the image is distorted by noise.

Keywords: NK-network, self-organization, image processing, image analysis, neural network, binary logical elements, signature authentication.

Author Index