Evolving membership functions in fuzzy linguistic summarization

Katarzyna Kaczmarek-Majer[1,*], Aleksandra Rutkowska[2] and Olgierd Hryniewicz[1]

[1]Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland

[2]Poznan University of Economics and Buisness, Niepodleglosci 10, 61-875 Poznań, Poland

Abstract

This paper introduces a time-dependent procedure for the construction of membership functions and applies it in linguistic summarization of time series. The primary goal is to dynamically reflect the changing interpretation of the linguistic terms by using a mathematical model and new data. In particular, the autoregressive and moving average models are applied. The proposed evolving linguistic summarization is illustrated for economic time series and supports the analysis of the expectations of customers towards inflation vs. the sentiment of communication of central banks. To reflect the changing nature of the economic variables, the definitions of linguistic terms are updated with time using statistical modeling. The preliminary results presented in this paper are promising. The proposed approach is illustrated with an example from economic time series, though it seems to have wider application potential.

Keywords

Linguistic summaries, Linguistic descriptions, Evolving fuzzy system, Inflation expectation

1. Introduction

Fuzzy linguistic summaries (also called linguistic descriptions) have been demonstrated successful to describe and summarize large datasets in various domains, see e.g., [1, 2, 3]. One of the main advantage of fuzzy linguistic summaries is their human-consistency [4]. At the same time, one of the main challenge when generating fuzzy linguistic descriptions with the use of type-1 fuzzy sets is the proper definition of membership functions that describe the linguistic terms related to the considered attributes such as e.g., *high inflation*. Furthermore, outcomes of the linguistic summarization highly depend on these definitions. For example, in [5], the authors show that relative linguistic variables constructed taking into account the historical data of an individual patient only in a healthy state enable to produce most informative results. In [6], the relative and personalized approach was guided by the semi-supervised online algorithm, and it resulted very promising in the context of sensor-based monitoring of bipolar disorder patients.

In this work, inspired by the general concept of evolving fuzzy systems, we propose to dynamically update the membership functions using statistical models. Similarly to the evolving fuzzy systems, components of the linguistic summarization can gradually change by learning from experience based on new data. In the literature, different techniques to automatically

OLUD 2022: First Workshop on Online Learning from Uncertain Data Streams, July 18, 2022, Padua, Italy
*Corresponding author.
k.kaczmarek@ibspan.waw.pl (K. Kaczmarek-Majer); aleksandra.rutkowska@ue.poznan.pl (A. Rutkowska)
0000-0003-0422-9366 (K. Kaczmarek-Majer); 0000-0002-2111-7764 (A. Rutkowska)

adapt fuzzy systems based on new data are described demonstrating their usefulness towards static approaches, see [7, 8]. For example, in [9], the structure of the control rule-base adapts to new data in an on-line mode with recursive, non-iterative learning. Although the evolving fuzzy systems have been developing dynamically, majority of the related work about generating fuzzy linguistic descriptions are static, see [10, 11], and it is assumed that the the fuzzy sets are constant in time. Evolving membership functions seem especially needed when the linguistic summaries are created for sequential data like time series or datastreams.

Section 2 describes the proposed method to construct evolving linguistic terms. The proposed procedure is illustrated with real-life economic data collected from Central Banks (CB) and Eurostat[1] in Section 3. The relations between sentiment of messages (also called tone) of CB and financial markets or inflation expectation were recently confirmed using econometric methods in some studies, such as vector autoregressive models with impulse response analysis or dynamic panels models, see [12, 13, 14]). In the experimental Section, we examine the relation between the sentiment of the CB communication and the expectations of customers towards inflation. In Section 4, main conclusions are discussed and future work is outlined.

2. Evolving membership functions in fuzzy linguistic summaries

Within this research, linguistic summaries based on extended protoforms in the sense of Yager and Kacprzyk [15, 4] are adapted. They describe with natural language the general facts about the evolution of numerical datasets. Although the resulting protoforms are quite simple, they enable to reveal complex relations between attributes [4]. It also needs to be noted that we use type-I fuzzy sets to describe the linguistic terms. Due to the lack of space we will not deal here with other protoforms or approaches to summarization such as type-2 fuzzy sets, linguistic summarization with Natural Language Generation, e.g., [16].

Formally, let $O = \{o_1, o_2 ..., o_C\}$ be a set of objects in a considered domain, e.g., economic situation in C countries. The properties of objects are measured by a set of attributes $A = \{r_1, r_2 ..., r_K\}$, such as e.g., economic indicators (inflation expectation, tone, etc.). Next, the set of linguistic terms set $l_k = \{l_k^1, ..., l_k^{j_k}\}$, $k \in \{1, 2..., K\}$ is established for each attribute from A such as e.g., *low, high.* Finally, type-I fuzzy sets are used to define these terms. Let fuzzy set A be represented by a membership function defined on a universe of discourse X as $\mu_A : X \to [0, 1]$, where A is the linguistic term (value) describing the variable $x \in X$. In the traditional static approach, fuzzy membership functions μ_A are created once at the beginning of the linguistic summarization process and are applied for all instances. In particular, one can apply e.g., the quartile-based approach to calculate the parameters of membership functions as depicted in Table 1.

The following three quartiles are computed from data: the first quartile (0.25) denoted by Q_1, the second quartile (0.5) denoted by Q_2 and the third quartile (0.75) denoted by Q_3. Next, *low* terms are expressed with z-shape fuzzy numbers and are characterized by the two parameters Q_1 and Q_2; *medium* terms are expressed with triangular fuzzy numbers that are characterized by

[1]https://ec.europa.eu/eurostat

Table 1
Quartile-based membership functions: static approach to the construction of membership functions of fuzzy numbers $[a_i, b_i, c_i]$ where Q_1 is the first quartile, Q_2 is median and Q_3 is the third.

Attribute	type	a_i	b_i	c_i
low	z-shape		Q_1	Q_2
medium	triangular	Q_1	Q_2	Q_3
high	s-shape	Q_2	Q_3	

the three parameters Q_1, Q_2, and Q_3; and the *high* terms are expressed s-shape fuzzy numbers that are characterized by the two parameters Q_2 and Q_3.

Next, we extend the quantile-based static approach in order to reflect the changing nature of the linguistic terms. We define a fuzzy set A_t represented by a evolving membership functions (dependent on time) defined on a universe of discourse X as follows: $\mu_{A_t} : X \to [0, 1]$, where A_t is the linguistic (value) term describing the variable $x \in X$ in time $t \in T$.

To construct these evolving memberships in an automatic manner, we select a time series that can be observed and is related to the interpretation of the linguistic term. Let us denote this time series as y. Next, we estimate a predictive model M describing y.

In particular, we consider that the observed discrete time series y comes from a stationary autoregressive and moving average (ARMA) process and limit the search for best model M to this class of models. ARMA has the following structure:

$$y_i = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i e_{t-i} + e_t \tag{1}$$

where $e_t \sim N(0, \sigma^2)$ are normally distributed independent random variables with the expected value equal to zero, and the finite standard deviation $\sigma^2 \in (0, 1)$ and $\phi_i \in$ (-1, 1) and $\theta_i \in$ (-1, 1) are paremeters of the model.

Finally, model M and new data of y are used to guide the construction of the evolving membership functions. In this work, for each moment t, we calculate the estimated value y_t and respective error e_t, and fuzzy numbers representing linguistic terms are created as depicted in Table 2. The *low* terms are expressed with z-shape fuzzy numbers and are characterized by the two parameters $y_t - e_t$ and y_t ; *medium* terms are expressed with triangular fuzzy numbers that are characterized by the three parameters $y_t - e_t$, y_t and $y_t + e_t$; and the *high* terms are expressed s-shape fuzzy numbers that are characterized by the two parameters y_t and $y_t + e_t$.

For example, in the considered application context of the economic use case, we select the observed time series of the inflation rate y to construct the membership functions for the linguistic terms related to the customer expectations of the inflation rate. Next, we identify best ARMA models to describe y, and these models are used in the linguistic summarization.

Having constructed the fuzzy numbers representing the attributes, we generate the linguistic summaries, also called fuzzy quantified sentences. In this work, we build the linguistic summaries based on the Yager's extended protoforms [15]. The linguistic summaries based on the extended protoforms enable to capture relations within and between the groups of attributes.

Table 2
Evolving membership functions (time-dependent): approach to the construction of fuzzy numbers that is based on one period forecast (y_t) and the error term (e_t).

Attribute	type	a_i	b_i	c_i
low	z-shape		$y_t - e_t$	y_t
medium	triangular	$y_t - e_t$	y_t	$y_t + e_t$
high	s-shape	y_t	$y_t + e_t$	

Following [5], we implement the summarization by the tree search algorithm with tree nodes corresponding to linguistic terms sets l_k. Linguistic summary based on the extended protoform LS_{re} takes the following form:

$$\textit{Among R objects, Q have P [DoT]} \tag{2}$$

where Q is the quantifier (the amount determination, e.g., *most*); R is a qualifier (attribute together with an imprecise label) about objects $o \in O$. P is the summarizer (attribute together with an imprecise label, e.g., low expected inflation rate; and $DoT \in [0, 1]$ measures the quality of the summary (level of confidence).

To compute the degree of truth DoT, we apply the Zadeh's degree of truth (validity) (DoT) of the summary defined in the following way:

$$DoT = \mu_Q\left(\frac{\sum_{i=1}^{n}\left(\mu_R(x_i) \wedge \mu_P(x_i)\right)}{\sum_{i=1}^{n} \mu_R(x_i)}\right) \tag{3}$$

where $\mu_R, \mu_P, \mu_Q : Re \rightarrow [0, 1]$ be the membership functions of fuzzy sets representing the qualifier R, summarizer P and quantifier Q, respectively, $\wedge$ is a t-norm.

The main purpose of experiments is to compare the linguistic summarization results for an exemplary application context that is economic use case. The comparative analysis assumes the static vs. the evolving approach to construction of membership functions that describe the linguistic terms.

3. Experimental results

3.1. About the application context: analysis of the expectations of customers towards inflation and the sentiment of communication of central banks

Communication entails both what is revealed to the public and how it is revealed. Some of the common and popular communication channel are minutes or press releases with the policy decision. The rationale in it remains the most important way of conveying monetary policy information to the public. That is why it is so important to measure the impact of information sent by CBs on the private forecasts of consumers. After each monetary council meeting the CB publishes minutes with information from the meetings, but these are formal texts without clear overtones and can be difficult for consumers to understand. This is why online information and the output of e.g. Twitter messages in the analysis of inflation expectations are becoming more and more important (see [17]. The main goal of experiments is to construct linguistic summaries describing the relationship between the tone (sentiment) of CB messages and customer inflation expectation in the dynamic economy situation and inflation perception. Next, we compare the static and dynamic approach to calculate the membership values that describe the linguistic terms related to inflation expectations.

3.2. About datasets

In this study, we focus on the following two economic variables:

1. **Inflation expectation** (*inf_exp*) - the consumer inflation expectations from Business and Consumers Surveys (BCS) estimated using the canonical probabilistic method of [18] in a subjectified version adjusted to a polychotomous survey [19].
2. **Tone** (*sentiment*) of communication of Central Banks (CBs) measures with the minutes of meeting, protocols or press release depends on Central Bank (further in short we use notation minutes) - it is sentiment of CB public messages - we classified words occurring in monetary policy related paragraphs according to the dictionary used by [20]. The tone variable is derived from corpora after words have been identified as negative or positive according to lexicons. The algorithm counts the words and returns the simple index of communication sentiment calculated as the difference between all positive (hawkish) words and negative (dovish) words divided by their sum. The tone is a continuous variable for each minute or press release, the value of which varies from -1 (all words are dovish) to 1 (all words are hawkish).

Six EU economies between 2001 and mid-2019, that implemented inflation targeting (IT), are considered in this study, namely: the Czech Republic, Hungary, Poland, Romania, Sweden, and the UK. These countries are members in the European Union[2] but not accompanied by the membership in the Euro area. The time series of inflation expectation and tone of CB minutes for one exemplary country (UK) are presented in Figure 1.

3.3. Evolving membership functions for fuzzy linguistic summaries

For the construction of the customer perception on inflation, the so called expected inflation, evolving fuzzy linguistic descriptions are constructed. To determine what is the customer perception of expected *low, high* inflation, we study the time series of actual inflation rate. Next, we identify an ARMA process that best describes it.

In experiments, we consider the following tree ARMA models:

inf_exp0 MA(6) model reflecting the 6 month rolling averages and the standard deviation of it;

inf_exp1 ARMA(1,1) model estimated on data from last 12 month and the estimated standard errors;

[2]the withdrawal of the United Kingdom (UK) from the European Union (EU) was on 31 January 2020

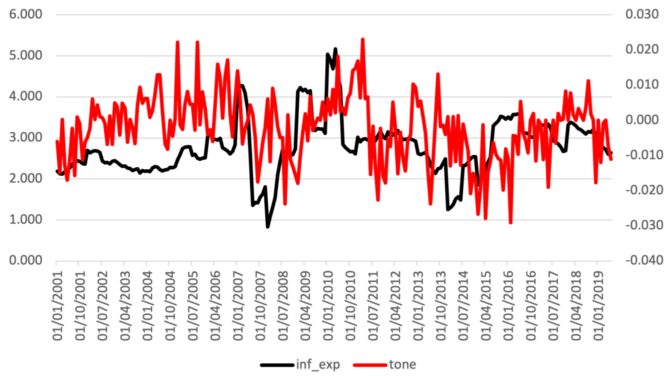

Figure 1: The inflation expectation (left axis) and tone (right axis) for one of the considered countries (UK) from Jan-2001 to June-2019.

inf_exp2 ARMA(1,1) model estimated on all available data in particular moment t, so from i=0 to $i = t - 1$ and the estimated standard errors.

Let us first look closer into the quartile-based membership values for an exemplary record of Jan-2001 in the UK. We know that the actual inflation rate in UK in Jan-2001 was 1.52. Figure 2 presents the memberships calculated with the quartile-based approach for the expected inflation in UK.

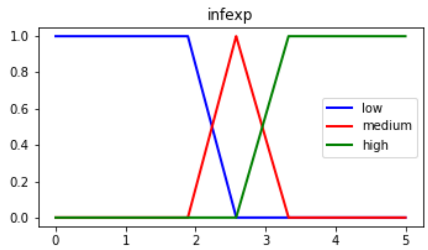

Figure 2: Quartile-based membership functions describing linguistic terms for the expected inflation.

As observed in Figure 2, according to the static approach (inf_exp) based on quartiles from expected inflation data, we calculate the that the expected inflation of 2.7 is *high* to a very small degree *high*(2.7)=0.1.

Let us now look closer into the dynamic approaches to membership values for this exemplary record. Figure 3. depicts evolving fuzzy memberships for the expected inflation in UK in January, 2001. We use the time series of the actual inflation rate and estimate the considered three ARMA models (MA(6) and ARMA(1,1)). Next, we calculate forecast y_{Jan-01} from infexp0 (MA), and it amounts to 1.18. Learning from infexp1 and infexp2 (ARMA) models, we calculate forecast y_{Jan-01} from it and amounts to 1.175. Those values are used as centers for fuzzy numbers describing the *medium* term, in line with Table 2. Finally, we construct the corresponding fuzzy numbers describing *low, medium, high* terms.

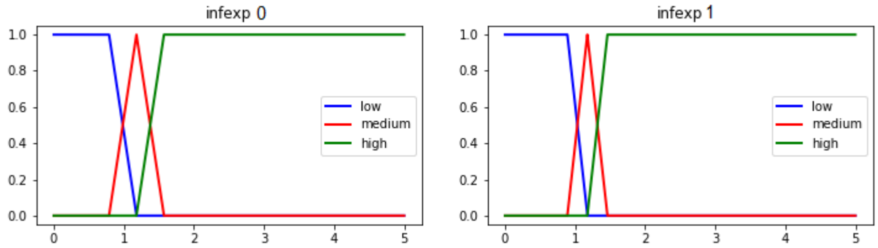

Figure 3: Evolving membership functions for Jan'2001 in UK from left to right: a) based on the MA(6) model (inf_exp0) b) based on the ARMA(1,1) model (inf_exp1). Fuzzy numbers based on ARMA(1,1) model (inf_exp2) are the same for this record as (inf_exp1).

As we know, for this particular record of January, 2001, the expected inflation was 2.7. As observed in Figure 3, the expected inflation of 2.7 is *high* ($high_{Jan-01}(2.7)$=1) according to all evolving variants.

Next, linguistic summaries are constructed using static and evolving approaches. Linguistic summaries for records among different tone and expected inflation are presented in Table 3, and in Figure 4. It is observed that in case of static approach to inflation expectation (*inf_exp*), the summary: *Among all low tone records, most inflation expectation are low* has the degree of truth 0.244 and the summary *Among all low tone records, most inflation expectation are high* is true to the degree of 0.148. Simultaneously, the summaries *Among all high tone records, most inflation expectation are high / low* is true to the degree of 0.197 / 0.169, respectively. We conclude that these degrees of truth are low and do not differ much, thus, one may suspect no relationship between the groups.

On the contrary, when we introduce evolving membership functions (*inf_exp0 - inf_exp2*), the summary: *Among all low tone records, most inflation expectation are low* has still the degree of truth on similar level 0.177, 0.233 (depending on the method adopted), but the summary *Among all low tone records, most inflation expectation are high* is true to the degree of between 0.519-0.600 (depending on the method adopted). The differences in degrees can be traced on the Figure 4 (left panel). Also, interestingly, in case of summaries for *all high tone records*, results using static fuzzy number for low and high are even closer, while using evolving membership functions the differences in the degrees of truth for the low and high are even greater (cf. Figure 4 (right panel), Table 3).

Looking at the results from the point of view of economic theories, one can notice that after introducing the evolving membership functions, relation between tone and inflation expectation seem to be more visible. Both, the linguistic summary *Among all high tone records, most inflation expectation are high* and the summary *Among all low tone records, most inflation expectation are high* are informative. The first one is more direct and it suggests CB low credibility. High (positive) tone transforms into higher expectations as as sign of CB lack of ability to constrain inflationary pressure. The second suggestion reveals that dovish tone transforms into accommodative monetary policy and high expectations. Also, the summaries type *Among all high(low) expectations records, most are with tone high(low)* have a low level of truth regardless of the case, which also makes economic sense due to the direction of the

Table 3
Linguistic summaries described tone of CBs' minutes and customers inflation expectation relationship with different approach to determined linguistic variable inflation expectation.

Among all low tone records, most inflation expectation are		
	...low	...high
inf_exp	0.244	0.148
inf_exp0	0.233	0.519
inf_exp1	0.233	0.589
inf_exp2	0.177	0.600
Among all high tone records, most inflation expectation are		
	...low	...high
inf_exp	0.197	0.169
inf_exp0	0.001	0.593
inf_exp1	0.130	0.593
inf_exp2	0.141	0.437
Among all low inflation expectation records, most are with tone		
	...low	...high
inf_exp	0.311	0.188
inf_exp0	0.399	0.084
inf_exp1	0.315	0.166
inf_exp2	0.283	0.202
Among all high inflation expectation records, most are with tone		
	...low	...high
inf_exp	0.216	0.198
inf_exp0	0.170	0.184
inf_exp1	0.205	0.169
inf_exp2	0.260	0.102

relationship between tone and expectations.

4. Conclusions

This study presents linguistic summaries using evolving membership functions that are updated gradually when new data arrive according to mathematical model. The need for such dynamic update is motivated with an illustrative example in economic time series. For this application context, the same absolute numbers can be perceived differently over time depending on the economic context. In such a case, keeping the membership functions constant seems too general and may lead to erroneous conclusions or missed relationships.

Preliminary results indicate that the tool of linguistic summaries under the condition of introducing time-varying membership functions is more adequate than static ones. It also needs to be noted that further experiments are planned to draw economic conclusions on larger sample including other countries with different transparency and different level of citizens' trust in the activities of central banks.

In terms of methodology, we plan to further investigate the selection of best statistical models reflecting the perception on linguistic terms. Fuzzy numbers considered in this work

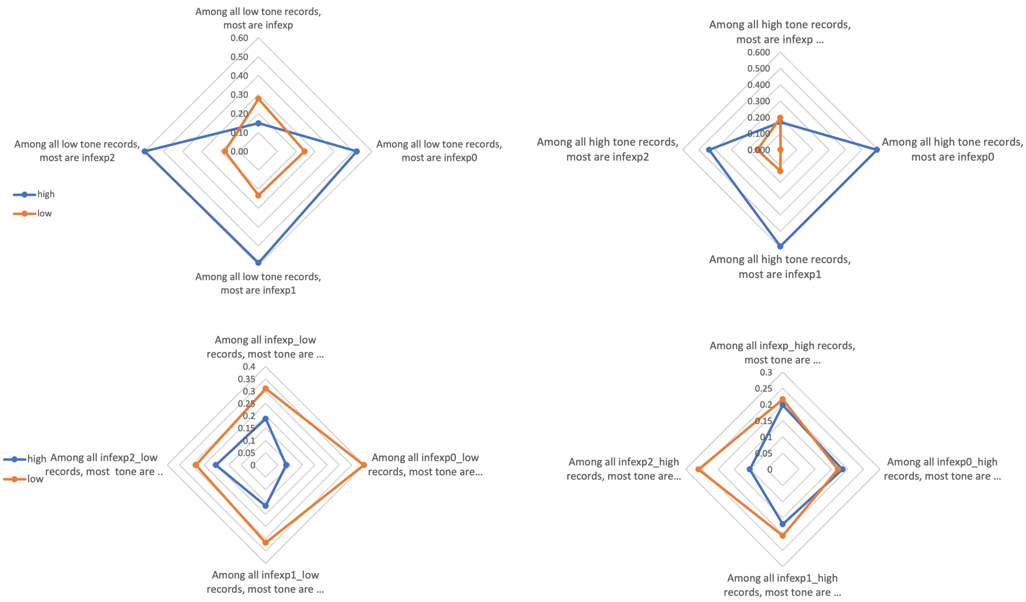

Figure 4: From top to bottom: the truth degree of linguistic summaries "Among all low \high tone records, most inflation expectation are high \low ", and the degree of truth "Among all low \high inflation expectation records, most tone are high \low " for the considered variants of infexp-related fuzzy numbers determination.

are a special case of the evaluative linguistic expressions in the sense of Novák [21]. Further research assumes in-depth study of other evaluative linguistic expressions and more recent fuzzy quantifiers. Finally, the proposed approach is illustrated with an example from economic time series, though it seems to have wider application potential.

Acknowledgments

Aleksandra Rutkowska is supported by the National Science Centre, Poland grant No.2020/37/B/HS4/02611

References

[1] J. Moreno-Garcia, J. Abián-Vicén, L. Jimenez-Linares, L. Rodriguez-Benitez, Description of multivariate time series by means of trends characterization in the fuzzy domain, Fuzzy Sets and Systems 285 (2016) 118–139. doi:https://doi.org/10.1016/j.fss.2015.05.011.

[2] K. Kaczmarek-Majer, O. Hryniewicz, K. R. Opara, W. Radziszewska, A. Olwert, J. W. Owsiński, S. Zadrożny, Control charts designed using model averaging approach for phase change detection in bipolar disorder, in: S. Destercke, et al. (Eds.), Uncertainty Modelling in Data Science, volume 832 of *Advances in Intell. Systems and Computing*, Springer International, 2019, pp. 115–123. doi:10.1007/978-3-319-97547-4_16.

[3] J. Moreno-Garcia, L. Rodriguez-Benitez, Jimenez-Linares, L., G. Triviño, A linguistic extension of petri nets for the description of systems: An application to time series, IEEE Transactions on Fuzzy Systems 27 (2019) 1818–1832. doi:10.1109/TFUZZ.2019.2892340.

[4] J. Kacprzyk, R. R. Yager, J. M. Merigo, Towards human-centric aggregation via ordered weighted aggregation operators and linguistic data summaries: A new perspective on zadeh's inspirations, IEEE Computational Intelligence Magazine 14 (2019) 16–30. doi:10.1109/MCI.2018.2881641.

[5] K. Kaczmarek-Majer, O. Hryniewicz, Application of linguistic summarization methods in time series forecasting, Information Sciences 478 (2019) 580–594. doi:https://doi.org/10.1016/j.ins.2018.11.036.

[6] K. Kaczmarek-Majer, G. Casalino, G. Castellano, O. Hryniewicz, M. Dominiak, Explaining smartphone-based acoustic data in bipolar disorder: Semi-supervised fuzzy clustering and relative linguistic summaries, Information Sciences 588 (2022) 174–195. URL: https://www.sciencedirect.com/science/article/pii/S0020025521012706. doi:https://doi.org/10.1016/j.ins.2021.12.049.

[7] R. R. Yager, Learning Methods for Evolving Intelligent Systems, John Wiley & Sons, Ltd, 2010, pp. 1–19. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470569962.ch1. doi:https://doi.org/10.1002/9780470569962.ch1. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470569962.ch1.

[8] P. Angelov, D. P. Filev, N. Kasabov, Evolving TakagiSugeno Fuzzy Systems from Streaming Data (eTS+), 2010, pp. 21–50. doi:10.1002/9780470569962.ch2.

[9] P. Angelov, A fuzzy controller with evolving structure, Information Sciences 161 (2004) 21–35. URL: https://www.sciencedirect.com/science/article/pii/S0020025503003165. doi:https://doi.org/10.1016/j.ins.2003.03.006, bio-Inspired Systems (BIS).

[10] A. Ramos-Soto, P. Martin-Rodilla, Enriching linguistic descriptions of data: A framework for composite protoforms, Fuzzy Sets and Systems 407 (2019) 1–26. doi:https://doi.org/10.1016/j.fss.2019.11.013.

[11] F. E. Boran, D. Akay, R. R. Yager, An overview of methods for linguistic summarization with fuzzy sets, Expert Systems with Applications 61 (2016) 356–377. doi:https://doi.org/10.1016/j.eswa.2016.05.044.

[12] G. C. Montes, L. V. Oliveira, A. Curi, R. T. F. Nicolay, Effects of transparency, monetary policy signalling and clarity of central bank communication on disagreement about inflation expectations, Applied Economics 48 (2016) 590–607. doi:10.1080/00036846.2015.1083091.

[13] P. Baranowski, W. Doryń, T. Łyziak, E. Stanisławska, Words and deeds in managing expectations: empirical evidence on an inflation targeting economy, Economic Modelling 95 (2021) 49–67. doi:https://doi.org/10.1016/j.econmod.2020.12.003.

[14] M. Szyszko, A. Rutkowska, A. Kliber, M. Próchniak, Central bank communication and expectations. evidence for inflation targeting economies, 2021. Submitted.

[15] R. R. Yager, A new approach to the summarization of data, Information Sciences 28 (1982) 69–86. doi:https://doi.org/10.1016/0020-0255(82)90033-0.

[16] M. Dragoni, I. Donadello, C. Eccher, Explainable ai meets persuasiveness: Translating reasoning results into behavioral change advice, Artificial Intelligence in Medicine 105 (2020) 101840. URL: https://www.sciencedirect.com/science/article/pii/S0933365719310140. doi:https://doi.org/10.1016/j.artmed.2020.101840.

[17] I. Korhonen, E. Newby, Mastering central bank communication challenges via twitter, Technical Report, BoF Economics Review, 2019. URL: https://helda.helsinki.fi/bof/handle/123456789/16534.

[18] J. A. Carlson, M. Parkin, Inflation expectations, Economica 42 (1975) 123–138. doi:https://doi.org/10.2307/2553588.

[19] R. A. Batchelor, A. B. Orr, Inflation expectations revisited, Economica (1988) 317–331. doi:https://doi.org/10.2307/2554010.

[20] H. Bennani, M. Neuenkirch, The (home) bias of European central bankers: new evidence based on speeches, Applied Economics 49 (2017) 1114–1131. doi:10.1080/00036846.2016.1210782.

[21] V. Novák, Evaluative linguistic expressions vs. fuzzy categories, Fuzzy Sets and Systems 281 (2015) 73–87. URL: https://www.sciencedirect.com/science/article/pii/S0165011415004042. doi:https://doi.org/10.1016/j.fss.2015.08.022, special Issue Celebrating the 50th Anniversary of Fuzzy Sets.