Mining and Linguistically Interpreting Data from Questionnaires
Influence of Financial Literacy to Behaviour

Miroslav Hudec[1] and Zuzana Brokešová[2]
[1]*Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1, Bratislava, Slovakia*
[2]*Faculty of National Economy, University of Economics in Bratislava, Dolnozemská cesta 1, Bratislava, Slovakia*

Keywords: Financial Literacy, Questionnaire, Data Mining, Quantified Sentence of Natural Language, Flexible Data Summarization, Fuzzy Logic.

Abstract: This paper is focused on mining and interpreting information about effect of financial literacy on individuals' behavior from the collected data by soft computing approach. Fuzzy sets and fuzzy logic allows us to formalize linguistic terms such as *most of*, *high literacy* and the like and interpret mined knowledge by short quantified sentences of natural language. This way is capable to cover semantic uncertainty in data and concepts. The preliminary results in this position paper have shown that for majority of people of low financial literacy angst and other treats represent serious issues, whereas about half of people with high literacy do not consider these treats as significant. Finally, influence of literacy to anchoring questions is mined and interpreted. Eventually, the paper emphasises needs for further data analysis and comparison.

1 INTRODUCTION

Adequate level of financial knowledge was identified as important for sound financial decisions (Lusardi, 2008). In addition, previous researches indicate that individuals' financial decisions are affected by different psychological heuristics and biases (Tversky and Kahneman, 1975). They include emotional aspects in decision-making, loss aversion, anchoring, framing and many others (see Gilovich et al., 2002). According to our knowledge, the effect of financial literacy to these heuristics and biases has not been widely examined in the literature. Hence, in the paper, we test the effect of high financial literacy to (i) feeling like angst, nervousness, loss of control and fear regarding possible catastrophic scenarios; (ii) how these people are influenced by anchor questions; (iii) how they decide about risky investments.

The second task is mining and interpreting this effect from the data collected by surveys in an easily understandable and interpretable way. First, summaries from the data are better understandable if they are not as terse as numbers (Yager et al., 1990). Secondly, there are often uncertainties in answers, which should not be neglected (Hudec, 2015; Viertl, 2011). Thirdly, answers to respective questions might be numbers, categorical data and short texts. Hence, mining and interpreting survey data by computational intelligence is beneficial. Fuzzy sets and fuzzy logic allows us to mathemiatically formalize linguistic terms such as *most of*, *low literacy*, *high angst* and the like. They make possible partial membership degrees of elements near the borderline cases to these concepts.

In order to meet both aforementioned challenges, we have conducted survey in Slovakia and adjusted Linguistic Summaries (LSs) in order to mine and interpret relational knowledge between financial literacy and respective attributes.

2 LINGUISTIC SUMMARIES AND THEIR QUALITY

LSs have been initially introduced by Yager (1982) in order to express summarized information from the data by linguistic terms instead of numbers. Overview of recent development can be found in (Boran et al., 2016). LSs of the structure *Q R entities in a data set are (have) S* developed by Rasmussen and Yager (1997) are relevant for our research. One example of such summary is *most of low financially literate people feel high level of angst*. The validity of such a LS is computed as

Hudec, M. and Brokešová, Z.
Mining and Linguistically Interpreting Data from Questionnaires - Influence of Financial Literacy to Behaviour.
DOI: 10.5220/0006462402490254
In *Proceedings of the 6th International Conference on Data Science, Technology and Applications (DATA 2017)*, pages 249-254
ISBN: 978-989-758-255-4

$$v(Qx(Px)) = \mu_Q\left(\frac{\sum_{i=1}^{n} t(\mu_S(x_i), \mu_R(x_i))}{\sum_{i=1}^{n} \mu_R(x_i)}\right) \qquad (1)$$

where n is the cardinality of elements in a data set (in our case number of respondents), $\frac{\sum_{i=1}^{n} t(\mu_S(x_i), \mu_R(x_i))}{\sum_{i=1}^{n} \mu_R(x_i)}$ is the proportion of the elements that belong to restriction R and satisfy summarizer S (fully or partially), t is a t-norm, μ_S, μ_R and μ_Q are membership functions explaining summarizer S, restriction R and relative quantifier Q, respectively.

The truth value (validity) v gets value from the unit interval. If v is closer to the value of 1, then the relation between R and S explained by Q is more significant. The goal of LS is to reveal such relations. LSs are graphically illustrated in Figure 1, where grey areas between sets *low*, *medium* and *high* emphasizes the uncertain area, i.e. area where unambiguous belonging to a particular set cannot be arranged.

Flexible summarizers, restrictions and quantifiers are mathematically formalized by fuzzy sets. For instance, fuzzy set *around anchor value of m* can be expressed as triangular fuzzy set (Figure 2).

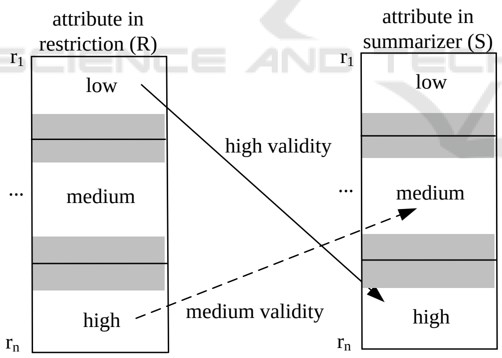

Figure 1: Graphical illustration of LSs with restriction.

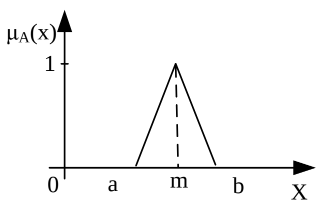

Figure 2: Triangular fuzzy set *around m*.

$$\mu(x) = \begin{cases} 1 & x = m \\ \frac{x-a}{m-a} & x \in (a,m) \\ \frac{b-x}{b-m} & x \in (m,b) \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

where belonging to the set decreases when distance to the anchor m increase.

The benefit against interval is in the intensity of belonging to a set. The closer is element to boundaries, the lower matching degree it has. Thus, defining boundaries is not as sensitive as for classical intervals. Similarly, fuzzy sets expressing concepts *small* and *high* are plotted in Figure 3. For categorical data matching degree is directly assigned to each element, i.e.

$$FL = \{(x_1, \mu(x_1)), \ldots, (x_n, \mu(x_n))\} \qquad (3)$$

where FL is a fuzzy set expressing concept in summarizer or restriction.

If LS with restriction has high validity v (1), it does not straightforwardly mean that such LS is relevant. In order to solve this problem several quality measures were suggested in (Hirota and Pedrycz, 1990). Due to their complexity and partial overlapping, simplified quality measure merging validity and coverage is suggested in (Hudec, 2017)

$$Q_c = \begin{cases} t(v, C) & C \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

where C is data coverage and t is a non-idempotent t-norm, e.g. product one. Coverage is calculated form the index of coverage.

$$i_C = \frac{\sum_{i=1}^{n} t(\mu_S(x_i), \mu_R(x_i))}{n} \qquad (5)$$

where parameters have the same meaning as in (1) and n is cardinality of data set by the transformation (Wu et al, 2010)

$$C = f(i_C) = \begin{cases} 0, & i_c \leq r_1 \\ 2\left(\frac{i_c - r_1}{r_2 - r_1}\right)^2, & r_1 \leq i_c < \frac{r_1 + r_2}{2} \\ 1 - 2\left(\frac{r_2 - i_c}{r_2 - r_1}\right)^2, & \frac{r_1 + r_2}{2} \leq i_c < r_2 \\ 1, & i_c \geq r_2 \end{cases} \qquad (6)$$

where $r_1 = 0.02$ and $r_2 = 0.15$, because in LSs with restriction only small subset of data is included in both sets R and S (Figure 1).

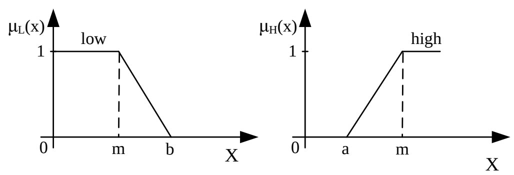

Figure 3: Fuzzy set expressing concepts *small* and *high*.

3 MINING RELATIONAL KNOWLEDGE AMONG ATTRIBUTES

This section explains procedure, results, discussion and challenges for further work. The data set consists of 644 answers with no missing data. Examples of questions, possible answers and fuzzy sets are provided in successive sections.

3.1 Experiments

The first step consists of construction of membership functions for attributes and their respective linguistic labels appearing in summaries.

Question regarding financial literacy is used in all experiments. The level of financial literacy was not included in the questionnaire. It was aggregated from answers to several questions. Financial literacy is expressed on the [0, 5] scale, where 0 is the lowest and 5 the highest level. Other questions were directly focused on feelings to potential kinds of threats, willingness to participate in financial games and influences by anchor questions.

The level of financial literacy attribute is fuzzified into three fuzzy sets: low literacy (LL), medium literacy (ML) and high literacy (HL) using notation (3) in the following way

HL = {(5, 1), (4, 0.75), (3, 0.45)}
ML = {(1, 0.35), (2, 1), (3, 1), (4, 0.35)}
LL = {(0, 1), (2, 0.75), (3, 0.45)}

It means that level 5 is without doubt high, level 4 significantly high but not fully and level 3 is partially high. Other two levels do not belong to this set. The recognized drawback in fuzzification is in set *ML*. It contains two elements with matching degree equal to 1. Hence, higher numbers of elements belong to this set. The ideal option is using scale of odd number of elements. Anyway, sets *HL* and *ML* do not cope with this issue.

3.1.1 Relation between Financial Literacy and Attributes of Fear, Angst, Nervousness and Loss of Control

Emotional heuristics and biases including fear, angst, nervousness or loss of control could significantly shape financial decision (e.g. Lee and Andrade, 2015). For example, in the case of insurance purchase, Zaleskiewicz et al. (2002) identified increase in the flood insurance demand after the recent flood. The role of financial literacy in this relation is questionable.

Our respondents were asked about emotions (angst, fear, nervousness and loss of control) evoked by imagination of possible negative events that could affect their life. These imagination was visualized in the pictures of unexpected natural disaster, death of relatives, war and deadly epidemic. Possible answers are: not at all, weakly, moderately, intensely and very intensely. These categorical attributes are fuzzified into sets *low* and *high* by (3):

L = {(not at all, 1), (weakly, 0.75), (moderately, 0.45)}
H = {(very intensely, 1), (intensely, 0.75), (moderately, 0.45)}

In order to get summaries, covered by sufficient number of respondents, we have strengthened Eq. (4) by condition $C \geq 0.75$, instead of 0.5.

Mined results are shown in tables 1 and 2, where terms *low* and *high* corresponds with terms illustrated in Figure 1. Financial literacy is restriction (*R*) and summarizers (*S*) are considered attributes. First row in Table 1 means that LS *about half respondents with high financial literacy have low angst* has significantly higher validity than sentence explained by quantifier *most of*.

Mined relational knowledge (Table 2) has shown that people with low financial literacy have high level of fear, angst, loss of control and nervousness. The most problematic attribute is angst, where significant proportion of respondents with low

Table 1: Result of summaries *Q respondents of high financial literacy have low values of respective attributes*.

high financial literacy		
low	**validity for quantifier Eq. (1)**	**coverage Eq. (6)**
angst	about half - 0.7129 most of - 0.1411	0.7888
loss of control	about half – 1	1.0000
fear	about half – 1	0.9988
nervousness	about half – 1	0.9346

Table 2: Result of summaries *Q respondents of low financial literacy have high values of respective attributes*.

low financial literacy		
high	**validity for quantifier (1)**	**coverage (6)**
angst	most of - 0.9996 about half – 0.0004	1
loss of control	most of - 0.5863 about half – 0.4137	1
fear	most of - 0.5818 about half – 0.4182	1
nervousness	most of - 0.7674 about half – 0.2326	1

financial literacy consider angst as a treat. These people might think that potential dangerous situation, if appear, will significantly devalue their lives and properties. On the other side, majority of people with high financial literacy does not have straightforwardly low intensities of these treats (Table 1). But at least about half of them have low values of respective threats.

3.1.2 Relation between Financial Literacy and Answer to Anchor Question

Anchoring represents a subconscious tendency of individuals to rely on onward information in the decision-making. This information does not have to be necessarily important for the decision. The role of financial literacy in the process of anchoring has not been previously studied.

In this experiment, we worked with numerical attribute: number of inhabitants in Iowa. Anchor value was set to 1 500 000 inhabitants. In order to reveal behaviour of people who have answered around this anchor value, we had relaxed crisp value 1 500 000 to the fuzzy number *around 1 500 000*, which is a convex fuzzy set with limited support and for value of 1 500 000 the matching degree is equal to 1. Fuzzy set around 1 500 000 is represented by (2) with the following values of parameters: a = 1200000, m = 1500000, b = 1800000 (Figure 2). Benefits against classical interval are explained in Section 2.

In this experiment, significant relations between literacy and answers to anchor questions have not been recorded by LSs. From the opposite view (restriction is anchor question and summarizer is financial literacy), we recorded high validity for summary: *most of respondents with answer around anchor have medium literacy*. These relations are summarized in Table 3.

Table 3: Result of summaries *Q respondents with answer around anchor value have (low, medium, high) values of financial literacy*.

number of inhabitants around the anchor value		
	validity for quantifier (1)	**coverage (6)**
low financial literacy	few – 0.5686 about half – 0.4314	0,1249
medium finan. literacy	most of – 1	0,9456
high financial literacy	few - 0.3018 about half – 0.6981	0,1918

We could conclude that individuals with high level of financial literacy do not reflect the anchor in their decision-making. However, also respondents with low level of financial literacy do not react to anchoring. Anyway, this experiment requires further analysis. We consider trying different terms regarding answers, for instance, *far from 1 500 000*, *significantly lower than 1 500 000*, *significantly higher than 1 500 000* and the like. These terms can be formalized by fuzzy sets plotted in Figure 3.

Another option is approach based on fuzzy functional dependencies, which might be helpful due to comparing conformance of each two respondents answer on both attributes.

3.1.3 Relation between Financial Literacy and Risk Taking

Risk aversion is an important parameter in financial decisions as majority of these decisions include risk. However, the effect of level of financial literacy is not clear. In economic literature, risk taking is often measured by the decisions in the lottery, where individuals could choose between sure and risky alternative (see e.g. Holt and Laury, 2002).

In our experiments, participation in lottery is numerical attribute. Respondents were asked how much they are willing to pay for the participation in the lottery, where one out of 10 players win 1000 €. Based on the expected value theory, those who are willing pay less than 100 € could be considered as risk averse, those who would like to pay 100 € are assumed as risk neutral and those who will pay more than 100 € are recognized as risk loving. The lowest recorded value was 0 and the highest 333, whereas mean value was 21.30 €. Data distribution is not uniform, which means that we cannot create three granules: *low*, *medium* and *high* by uniformly covering domain. The options are statistical mean based method (Tudorie, 2008) or logarithmic transformation (Hudec and Sudzina, 2012). Applying the former, fuzzy set *low participation* is expressed by parameters m = 25 and b = 37.5, whereas set *high participation* is expressed by parameters a = 62.5 and m = 75 (Figure 3). Results from this experiment are summarized in Table 4.

We have recorded that respondents belonging to all three categories of financial literacy prefer low participation in lottery. That is an expected result as the majority of the population is risk averse and prefer sure alternative over the risky one. We have not identified effect of financial literacy on risk taking. Other relations have not been recorded, due to low data coverage.

3.2 Discussion

The preliminary results are promising, but further research is highly advisable. Regarding questions angst, loss of control, fear and nervousness, we reveal that improving financial literacy might be beneficial for the population (tables 1 and 2). Although, majority of people with high financial literacy do not have straightforwardly low angst and other threats, at least about half of them do not consider the treats significant. For respondents with low financial literacy the treats represent problems.

On the other hand, reaction to anchoring is low in the group with high financial literacy and low financial literacy as well. Several studies supported the role of anchor in decision-making (Furnham and Boo, 2011). The absence of this effect on the two extremes of financial literacy (high, low) brings an interesting finding to the literature. One could speculate that those with high level of financial literacy notice the anchor but isolate its effect in decision-making. Those with low levels of financial literacy presumably do not notice anchor. This finding requires further research and robustness checks.

Our results suggest that risk aversion do not vary with financial literacy. Risk aversion is therefore superior characteristic and cannot be changed by the increase of the level of financial literacy.

Table 4: Result of summaries between financial literacy and participation in lottery using quantifiers *few*, *about half* and *most of*.

high financial literacy		
	validity for *Q*	**coverage**
low participation	most of – 1	1
high participation	few -1	0.0081<0.75
low financial literacy		
	validity for *Q*	**coverage**
low participation	most of – 1	1
high participation	few -1	0
medium financial literacy		
	validity for *Q*	**coverage**
low participation	most of – 1	1
high participation	few -1	0.2898<0.75

LSs are suitable for mining relational knowledge from variety of data: numbers (either crisp or fuzzy), (weighted) categorical data, short texts. The benefit is in low computational effort when optimization techniques are applied (Liu, 2011), covering non-linear dependencies and fast estimation. In order to reveal relational knowledge, domains are divided into several flexible granules: *low*, *medium* and *high*. This ensures that we can use all data types in (1), (4-6) considering their respective membership degrees.

3.3 Further Tasks and Opportunities

This is, according your best knowledge, first attempt to analyse impact of financial literacy to people perceptions and decisions as well as to analyse these data with linguistic summaries. The framework and mathematical formulation have been developed. In next parts of our project, we are going to analyse other possible combinations of attributes and their respective granules, such as answer far from anchor, significantly lower answers, etc.

Answers to majority of questions are linguistic terms. Such terms do not have clear definition in terms of set theory. Furthermore, concepts such as *around anchor value* are understandable, even though vague. Hence, we are considering building model based on Fuzzy Functional Dependencies (FFD) to compare with our results. Overview of FFD can be found in, e.g. (Vučetić and Vujošević, 2012). In this way, we can cover mining relational knowledge form questionnaires by two valuable approaches capable to manage semantic uncertainty.

Hence, the preliminary results are going to be confronted by approach based on FFD. Option based on FFD is more demanding in terms of computational effort, but is more powerful. It compares each two values in order to reveal whether similar values of one attribute causes similar values of another attribute.

Our results reveal influence of financial literacy to main threats and other attributes for Slovak respondents. Promising aspect for research is comparison with other countries as well as control the effect of the socio-economic characteristics of individuals in this relation. To reveal such answers an international survey and research is advisable.

4 CONCLUSIONS

Financial literacy, its improvement and it drivers is a relevant topic, which should be analysed. In this paper, we have shown that mining relational knowledge from well-designed questionnaires by fuzzy logic is valuable contribution to this field. Preliminary data mining results have shown that low literacy causes that people have lower quality of life, due to high level of angst, nervousness, fear and loss of control. On the other side, high literacy does not straightforwardly mean big improvement.

Further, we have recorded that respondents belonging to all three categories of financial literacy prefer low participation in lottery. It is an expected result as the majority of the population is risk averse and prefer sure alternative to the risky one.

Finally, result that reaction to anchoring is low in the group of respondents with high financial literacy and in the group with low financial literacy brings an interesting finding, which is contribution to further economical and data mining research.

These conclusions are preliminary. Regarding the data mining, in the next stage of our research, we are going to apply more powerful, but also computationally demanding approach for revelling flexible functional dependencies among attributes to confront already reached results. In addition, advisable are similar researches in other countries with different economic situation.

ACKNOWLEDGEMENTS

This paper is part of a research grant VEGA No. 1/0849/15 entitled ‘Economic and social aspects of the information asymmetry in the insurance market’ supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

REFERENCES

Boran, F. E., Akay, D., and Yager, R. R. 2016. An overview of methods for linguistic summarization with fuzzy sets. *Expert Systems with Applications*, 61:356–377.

Furnham, A., Boo, H. C. 2011. A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1):35-42.

Gilovich, T., Griffin, D., Kahneman, D. 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge University Press. Cambridge.

Hirota, K., Pedrycz, W. 1990. Fuzzy computing for data mining. Proceedings of IEEE, 87:1575-1600.

Holt, C. A., Laury, S. K. 2002. Risk aversion and incentive effects. *American economic review*, *92*(5):1644-1655.

Hudec, M. 2017. Merging validity and coverage for measuring quality of data summaries. In *Information Technology and Computational Physics*. Springer International Publishing, Switzerland (in press).

Hudec, M. 2015 Storing and analysing fuzzy data from surveys by relational databases and fuzzy logic approaches. In *XXV-th IEEE International Conference on Information, Communication and Automation Technologies (ICAT 2015)*. IEEE.

Hudec, M., Sudzina, F. 2012. Construction of fuzzy sets and applying aggregation operators for fuzzy queries. In *14th International Conference on Enterprise Information Systems (ICEIS 2012)* SCITEPRESS.

Lee, C. J., Andrade, E. B. 2015. Fear, excitement, and financial risk-taking. *Cognition and Emotion*, 29(1): 178-187.

Liu, B. 2011. Uncertain logic for modeling human language. *Journal of Uncertain Systems*, 5:3-20.

Lusardi, A. 2008. Financial literacy: an essential tool for informed consumer choice? (No. w14084). National Bureau of Economic Research.

Rasmussen, D., Yager, R.R. 1997. Summary SQL - A fuzzy tool for data mining. *Intelligent Data Analysis*, 1:49-58.

Tudorie, C. 2008. Qualifying objects in classical relational database querying. In *Handbook of Research on Fuzzy Information Processing in Databases*, pages. 218-245. Information Science Reference, Hershey.

Tversky, A., Kahneman, D. 1975. Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making*, pages. 141-162. Springer, Netherlands.

Viertl, R. 2011. Fuzzy data and information systems. In *15th WSEAS International Conference on Systems*. WSEAS Press.

Vučetić, M., Vujošević, M. 2012. A literature overview of functional dependencies in fuzzy relational database models. *Techniques and Technologies for Education and Management*, 7(4):1593-1604.

Wu, D., Mendel, J.M., Joo, J. 2010. Linguistic summarization using if-then rules. In *2010 IEEE International Conference on Fuzzy Systems*. IEEE.

Yager, R. R. 1982. A new approach to the summarization of data. *Information Sciences*, 28:69–86.

Yager, R.R., Ford, M., Cañas, A.J. 1990. An approach to the linguistic summarization of data. In *3rd International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU ’90)*. EUSFLAT.

Zaleskiewicz, T., Piskorz, Z., Borkowska, A. 2002. Fear or money? Decisions on insuring oneself against flood. *Risk, Decision and Policy*, 7(3):221-233.