

# AN STUDY OF THE INFLUENCE OF FUZZY VARIABLES ON THE ECONOMIC ANALYSIS

Javier Bonilla  
Facultad de Ciencias  
Matemáticas  
Universidad Complutense de  
Madrid  
Madrid, Spain  
[javierlb@ucm.es](mailto:javierlb@ucm.es)  
Started October 2015

Javier Montero  
Facultad de Ciencias  
Matemáticas  
Universidad Complutense de  
Madrid  
Madrid, Spain  
[monty@mat.ucm.es](mailto:monty@mat.ucm.es)

J. Tinguaro Rodríguez  
Facultad de Ciencias  
Matemáticas  
Universidad Complutense de  
Madrid  
Madrid, Spain  
[jtrodrig@mat.ucm.es](mailto:jtrodrig@mat.ucm.es)

**Abstract**— Predicting Gross Domestic Product is very important because until the final data is available, either temporarily or permanently, several weeks get spend after the quarter ends. So, we propose a new estimation based on text mining in economics documents based on a new proposed fuzzy clustering method with a new objective function.

**Keywords**—fuzzy logic; fuzzy clustering; intuitionistics fuzzy sets; fuzzy entropy; intuitionistic fuzzy generator; paired fuzzy sets; text mining; Gross Domestic Product estimation.

## I. Resume

### A. Description

Fuzzy logic enables to assign a membership degree between zero and one to each object to indicate the degree up to which that object belongs to a cluster, as introduced in [30]. After several decades, the work in [1] introduced an extension of fuzzy sets called intuitionistic fuzzy sets, which allow to also measure the non-membership degree and hesitation degree of an object with respect to a cluster, in order to try to find a better classification of reality. More information about different types of fuzzy sets can be find at [2] or [15].

Fuzzy models based on Belnap's logic further extends the intuitionistic model, so that the hesitation degree is divided in two fuzzy set, one is a degree of ignorance or lack of information and the second fuzzy set is contradiction, ambiguity or excess of information [11,17]. Moreover, in [16] and [20] the hesitation degree was divided in three fuzzy sets, indeterminacy, ambivalence and conflict, under the paired fuzzy sets approach. Other example of fuzzy model is the picture fuzzy sets [25] where a new fuzzy set appears, the refusal

degree, which is different from the neutral degree which is in turn more or less similar to the hesitation degree.

Text mining started in the sixties of last century with the work in [23] aimed to create the computer program General Inquirer<sup>1</sup>, which had the ability to extract information from data text using contextual analysis or programming natural language, with the use of dictionaries to classification of works. Other relevant papers about this topic are: [4,19] and [28].

Predicting Gross Domestic Product (GDP) is very important because until the final data is available, either temporarily or permanently, several weeks get spend after the quarter ends. So, different statistical methods as time series, confidence indicators or other economic variables can be used to predict the evolution of GDP. In order to apply the previous developments, we analyzed a data set consisting of the speeches of the president of the US Federal Reserve to build a new confidence index for predicting the quarterly evolution of US GDP. Examples for others authors with others statistical methods such as time series can be found in [7,10,12,13,21,22].

### B. General objectives

The general objective of this study is: To develop a new objective function and a new method of fuzzy clustering based on paired fuzzy sets [16] or intuitionistics fuzzy sets [1], with the application of the interactive descendent gradient method as the Fuzzy C-means Method [3,9]. The aim is enabling the classification of economic texts in order to create an advanced estimator of GDP.

---

<sup>1</sup> <http://www.wjh.harvard.edu/~inquier/>

### C. Specific objectives

The specific objectives we pursue are the following:

The development of the new mathematical methods for fuzzy clustering in order to classify data sets in clusters providing a better resume of information [5].

The estimation of some parameters of the models studied in [5], with the idea of not having to search a value for each of them with a grid or others methods such us genetic algorithms or particle swarm optimization.

The skills to extract information from text data sets [19], with the idea of splitting the text data sets on subjects or topics [28].

The comparison of the new methods developed with existing methods in the literature [3,8,14] and [25].

An application of the new methods is the classification of economics texts in order to create an advanced estimation of GDP.

### II. Methodology in the research plan

The methodology to be applied is based on the development of new fuzzy methods through the modification of the objective function of Fuzzy C-Means to allow the use of Tsallis entropy [27], and introducing a new parameter to link the entropy and tolerance of the model, as well as the application of intuitionistic fuzzy sets as part of the models or as a refinement of the solution.

We develop the estimation of the parameters of some negation fuzzy complement of Yager [29] and Sugeno [24], in order to generate fuzzy intuitionistic sets [6,18,26]. For example, the mode is used as part of the new estimators.

We compare the new methods with Fuzzy C-Means [3,9] and K-means [14] methods, with statistically significant differences between the new methods versus old methods. So we obtained better results on classification with the proposed methods.

These three first ideas appeared in the authors' work [5], containing proofs of the properties of both the new methods and the estimators of parameters.

To extract the information from text we change the way how to do it, in order to lemmatizer [4] versus stemmer the words [28], with the idea to get better results. Other important points [28] on this process are the tokenification of text, elimination

of stopwords and the use of term frequency and the inverse document frequency (tf-idf) metric. In [5] we used eight text dataset from [28].

About the application on GDP estimation, we use the speeches of the President of the US Federal Reserve, specifically Ben S. Bernanke and Alan Greenspan. The first has been president since 2006 to 2014 and the second from 1987 to 2006. Thus, we create the dataset speeches from the speeches of the President of US Federal Reserve (discarding the testimony or statements). To generate this dataset we use the General Inquirer program, as shown in [5].

### III. Relevancy

The main relevancy of this research is linked to the development of new methods of unsupervised classification, with the main idea of obtaining better partitions than those the previous methods offered. From a scientific point of view, we develop advances on the understanding of the behavior of objective functions and parameter settings.

The proposed methods can be applied to several fields, for example in economic analysis, social networks analysis, engineering, biostatistics or statistics.

A specific application of this research is in the estimation of GDP, a quite relevant issue in most economic processes, and thus in their associated decision making.

### References

- [1] K. T. Anatasov, "Intuitionistic Fuzzy Sets", In *Fuzzy Sets and Systems* 20, Elsevier, 1986, pp. 87-96.
- [2] B. Bede, "Mathematics of Fuzzy Sets and Fuzzy Logic", Berlin: Springer, 2013.
- [3] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", New York: Plenum Press, 1981.
- [4] S. Bird, E. Klein and E. Loper, "Natural language processing with Python", O'Reilly Media Inc., 2009.
- [5] J. Bonilla, J. Montero, J. T. Rodriguez and D. Velez, "New fuzzy non-suervised methods with applications to text mining", unpublished.
- [6] H. Bustince, J. Kacprzyk and V. Mohedano, "Intuitionistic fuzzy generators: Application to intuitionistic fuzzy complementation", *Fuzzy Sets and Systems*, num. 114, Elsevier, 2000, pp. 485-504.
- [7] J. L. Cendajas Bueno, S. de Lucas Santos and M. J. Delgado Rodríguez, "Evaluación del cambio estructural en los indicadores cíclicos", *Estadística Española*, Vol. 56, Num. 183, 2014 , pp. 39-59.

- [8] T. Chaira, "A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images", *Applied Soft Computing*, num. 11, Elsevier, 2011, pp. 1711–1717.
- [9] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters". *Cybernetics and Systems*, num. 3, vol. 3, IEEE, 1973, pp. 32-57.
- [10] M. Forni, M. Hallin, M. Lippi and L. Reichlin, "Coincident and leading indicators for the euro area", *The Economic Journal*, Vol. 111, Num. 471, Conference Papers, pp. 62-85, May 2001.
- [11] C. Franco, J. Montero and J. T. Rodríguez, "Relational structures for measures of ignorance", *EUSFLAT-LFA 2011*, Atlantis Press, July 2011.
- [12] M. Gallardo and M. Pedersen, "Resumen de metodologías de referencia para construir un indicador regional en América latina", *Estudios Estadísticos y Prospectivos*, 2007, pp. 1-59.
- [13] G. Gamba-Méndez, G. Kapetanios and R. J. Smith, "An Automatic Leading Indicator of Economic Activity: Forecasting GDP growth for European Countries", 1999, pp. 1-30.
- [14] J. B. MacQueen, "Some Methods of Classifications and Analysis of Multivariate Observations", *Proc. of 5th Berkeley Symposium on Math, Stat. and Prob.*, 1967, pp. 281-297.
- [15] J. Montero, D. Gómez and H. Bustince, "On the relevance of some families of fuzzy sets", *Fuzzy Sets and Systems*, Elsevier, Num. 158, 2007, pp. 2429-2442.
- [16] J. Montero, H. Bustince, C. Franco, J. T. Rodríguez, D. Gómez, M. Pagola, J. Fernández and E. Barrenechea, "Paried structures in knowledge representation", *Knowledge-Based Systems*, Elsevier, 2016.
- [17] M. Öztürk and A. Tsoukiás, "Modelling uncertain positive and negative reasons in decision aiding", *Decision Support Systems*, Num. 43, 2007, pp. 1512-1526.
- [18] N. R. Pal, H. Bustince, M. Pagola, U. K. Mukherjee, D. P. Goswami and G. Beliakov, "Uncertainties with Atanassov's Intuitionistic Fuzzy Sets: Fuzziness and Lack of Knowledge". *Information Sciences*, num. 228, Elsevier, 2013, pp. 61-74.
- [19] J. Pueyo and J. Quiles-Follana, "Trends in Natural Language Processing and Text Mining", *The European Journal for informatics Professional Upgrade*, Vol. XI, issue No. 3, pp. 33-39, June 2010.
- [20] J. T. Rodríguez, C. Franco, D. Gómez and J. Montero, "Paried structures and other opposites-based models", *16<sup>th</sup> IFSA and 9<sup>th</sup> EUSFLAT*, Atlantis Press, 2015.
- [21] D. Romer, "Advanced Macroeconomics", 3th edition, McGraw-Hill Companies, Inc. 2006.
- [22] J. H. Stock and M. W. Watson, "A probability model of the coincident economic indicators", *Nber working paper series*, num. 2772, 1988, pp. 1-41.
- [23] P. J. Stone, D. C. Dumphy, M. S. Smith and D. Ogilvie, "The General Inquirer: a computer approach to content analysis". Cambridge, MA: MIT Press, 1966.
- [24] M. Sugeno, "Fuzzy measures and fuzzy integrals: a survey", *Fuzzy Automata and Decision Process*, M. Gupta, G. N. Saridis, and B. R. Gaines, Eds., North Holland, 1977, pp. 82-102.
- [25] P. H. Thong and L. H. Son, "Picture fuzzy clustering: a new computational intelligence method", *Soft Comput Methodologies and Application*, num. 20, vol. 9, Springer, 2016, pp. 3549–3562.
- [26] E. Trillas, "On negation functions in fuzzy set theory", S. Barro, A. Bugarin, and A. Sobrino, Eds., "Advances in Fuzzy Logic", Universidade de Santiago de Compostela, 1998, pp. 31-43.
- [27] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics", *Journal of Statistical Physics*, num. 52, vol. 1-2, 1988, pp. 479-478.
- [28] D. Velez, J. Sueiras, A. Ortega, and J. F. Velez, "A method for K-means seeds generation applied to text mining", *Statistical Methods and Applications*, num. 25, vol. 3, Springer, 2016, pp. 477-499.
- [29] R. R. Yager, "On the measures of fuzziness and negation. Part II: lattices", *Information and Control*, num. 44, 1980, pp. 236-260.
- [30] L. A. Zadeh, "Fuzzy Sets", *Inf. Control*, Num. 8, Vol. 3, 1965, pp. 338-353.