Sentiment analysis via fractal dimension

Symeonidis Symeon
Department of Electrical and Computer Engineering
Democritus University of Thrace, Xanthi 67100, Greece
ssymeoni@ee.duth.gr

This paper presents a very early-stage idea about sentiment analysis with usage of theory of fractal dimension. It presents a short literature review about fractal dimension on datasets and an approach to sentiment analysis with fractal dimension.

Keywords: Sentiment analysis, Fractal Dimension, Classification

1. INTRODUCTION

Everyday more and more people use the Internet in order to communicate,work, search information or fro personal. The existing knowledge in conjunction with the vast amount of information collected by the use of Internet and the new ideas generated, imposed a shift in the scientists' to a more customize science which is called opinion mining and sentiment analysis.

Many applications for opinion mining and sentiment analysis have been deployed to analyze opinions, feelings and attitudes (Maks and Vossen 2012). In their previous studies, various authors categorized sentiment in three polars (da Silva, Hruschka, and Hruschka 2014) or in six "universal" emotions (Lemeignan, Guitart, and Bloch 1991). However is it possible for the dimensions of emotions to be predicted and classified? Furthermore, can a feeling be positive or negative and to what extent?

Fractal dimension is defined as "an estimate of the degrees of freedom of a data set" (Kumaraswamy 2003). Fractal analysis is employed by various sciences such as education and medicine, with main purpose to better quantify and describe, the deviation and complexity of images using a computable value (Park, Wang, and Zheng 2009). Despite the several limitations in providing accurate measures, fractal dimension and Euclidean geometry have provided crucial solutions to complicated everyday problems (Florindo and Bruno 2014).

An interesting approach, is this of information retrieval and specific of opinion retrieval - mining and sentiment analysis of users with the use of fractal dimension. The literature gap in this research area is a challenge in order to better explore and approach this issue.

2. RESEARCH DESCRIPTION

The last couple of years a high research interest is observed in the field of multimedia database management systems and more precisely in this of multimedia processing within computing systems (Nappi, Polese, and Tortora 1998). Application of fractal dimension cannot numerate sharply any objects and for such these objects is computed by estimation or approximation (Sadikin & Ito 2013).

Many data mining tasks such as dimensionality reduction, classification, clustering, learning patterns are used in sentiment analysis, mainly as patterns and an indicator of the way the data points are spread in the data space(Kumaraswamy 2003). In addition, the relationship between the spread of the data and the amount of information that can obtained, can enhance the performance of a given data mining method which is evaluated on the basis of the information captured. Nevertheless, all the above mentioned are expensive and require a large computation time during their implementation (Kumaraswamy 2003).

2.1 Data Fractal Dimension

As a dataset has fractal character, opinion and sentiment data can have the same capacity, if their properties such as the structure and the statistic distribution are the same in partial distribution(Ni, Ni, and Gao 2011; Bao et al. 2004).

Yan et al. (2006) measured the fractal dimension Dq of dataset by:

$$D_q = \begin{cases} \lim\limits_{r \to 0} \dfrac{\sum_i p_i \log p_i^q}{\log r} & q = 1 \\[2ex] \lim\limits_{r \to 0} \dfrac{1}{q-1} \dfrac{\log \sum_i p_i^q}{\log r} & q \neq 1 \end{cases} \quad r \in [r_1, r_2]$$

Figure 1: *Fractal dimension Dq of dataset*

where r is the grid size, pi is the probability of the data points in the ith grid, q is an integer (Ni, Ni, and Gao 2011).

2.2 Fractal Information Retrieval (FIR)

According to (Traina et al. 2010) "fractal dimension of an Euclidean object corresponds to its Euclidean dimension and it is always an integer number". Moreover Zhang et. al (2002) explain a technique which can discover and select the number of significant attributes to describe a dataset using fractal dimension. The study of combining Information Retrieval and fractal dimension can importantly enhance the results of retrieval techniques.

2.3 Fractal Sentiment Analysis (FSA)

Different techniques of sentiment analysis were published in the last years. Some of these approaches are based on natural language processing, lexicons and machine learning.

The problem in analyzing sentiment is how to convert every word to fractal. Is it necessary for every word to be converted according to the part of speech that it is (articles, nouns, pronouns, adjectives, verbs, adverbs, conjunctions, prepositions, and interjections)? Which technique of sentiment analysis is more appropriate when analyzing sentiment via fractal dimension?

My first approach is based on the ground that every word is related to the fractal dimension of a sentence and every word is correlated to the previous or following one. Every correlated word increases the complexity but the dimensional reduction method can avoid redundant dimensions and recover the original variables while preserving the topological properties of the sentence (Kumaraswamy 2003). Furthermore, in order to measure the spread of words and the intrinsic dimension of the sentence, fractal dimension is used which calculates the correlation integral and classification errors.

Another approach is this of computing every word with sentiment fractal dimension according to the part of speech that this is and by using priority to parts with "specific and essential" sentiment. Following the previous procedures, can be classified and different fractal levels for every category will be computing. This approach must be based on a sentiment analysis technique with strong sentiment evaluation for every word.

3. CONCLUSIONS

A very early-stage idea about fractal sentiment analysis was presented above. Further research in the existing literature combined with personal ideas and assumption s, will probably lead us to results.

4. ACKNOWLEDGEMENT

5. REFERENCES.

Bao, Yubin, Ge Yu, Huanliang Sun, and Daling Wang. 2004. *Advances in Web-Age Information Management*. Edited by Qing Li, Guoren Wang, and Ling Feng. *Advances in Web-Age Information Management: 5th International Conference, WAIM 2004, Dalian, China, July 15-17, 2004*. Vol. 3129. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/b98703.

Da Silva, Nádia F F, Eduardo R. Hruschka, and Estevam R. Hruschka. 2014. "Tweet Sentiment Analysis with Classifier Ensembles." *Decision Support Systems* 66. Elsevier B.V.: 170–79. doi:10.1016/j.dss.2014.07.003.

Florindo, João Batista, and Odemir Martinez Bruno. 2014. "Fractal Descriptors Based on the Probability Dimension: A Texture Analysis and Classification Approach." *Pattern Recognition Letters* 42 (1): 107–14. doi:10.1016/j.patrec.2014.01.009.

Kumaraswamy, Sree Krishna. 2003. "Fractal Dimension for Data Mining." *Ml.Cmu.Edu*. http://www.ml.cmu.edu/research/dap-papers/skkumar_kdd_project.pdf.

Lemeignan, M., P.L. Guitart, and S. Bloch. 1991. "Autonomic Differentiation of Six Basic Emotions." *International Journal of Psychophysiology* 11 (1): 52–53. doi:10.1016/0167-8760(91)90222-J.

Maks, Isa, and Piek Vossen. 2012. "A Lexicon Model for Deep Sentiment Analysis and Opinion Mining Applications." *Decision Support Systems* 53 (4). Elsevier B.V.: 680–88. doi:10.1016/j.dss.2012.05.025.

Nappi, M., G. Polese, and G. Tortora. 1998. "FIRST: Fractal Indexing and Retrieval SysTem for Image Databases." *Image and Vision Computing* 16 (14): 1019–31. doi:10.1016/S0262-8856(98)00084-5.

Ni, Li-Ping, Zhi-Wei Ni, and Ya-Zhuo Gao. 2011. "Stock Trend Prediction Based on Fractal Feature Selection and Support Vector Machine." *Expert Systems with Applications* 38 (5). Elsevier Ltd: 5569–76. doi:10.1016/j.eswa.2010.10.079.

Park, Sang Cheol, Xiao-Hui Wang, and Bin Zheng. 2009. "Assessment of Performance Improvement in Content-Based Medical Image Retrieval Schemes Using Fractal Dimension." *Academic Radiology* 16 (10): 1171–78. doi:10.1016/j.acra.2009.04.009.

Sadikin, Mujiono, and Wasito Ito. n.d. "FRACTAL DIMENSION AS A DATA DIMENSIONALITY REDUCTION." *The 7th International Conference on Information & Communication Technology and Systems (ICTS) 2013, 15 – 16 May 2013, Surabaya*, 105–10.

Traina, C, Agma Traina, Leejay Wu, and Christos Faloutsos. 2010. "Fast Feature Selection Using Fractal Dimension." *Journal of Information and Data Management* 1 (1): 3–16. http://repository.cmu.edu/compsci/580/.

Yan, Guanghui, Zhanhuai Li, and Liu Yuan. 2006. *MICAI 2006: Advances in Artificial Intelligence*. Edited by Alexander Gelbukh and Carlos Alberto Reyes-Garcia. *MICAI 2006: Advances in Artificial Intelligence, 5th Mexican International Conference on Artificial Intelligence, Apizaco, Mexico, November 13-17, 2006, Proceedings*. Vol. 4293. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/11925231.

Zhang, Haiqin, Chang-Shing Perng, and Qingsheng Cai. 2002. "An Improved Algorithm for Feature Selection Using Fractal Dimension." *Proceedings of the Second International Workshop on Databases, Documents, and Information Fusion*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.4407&rep=rep1&type=pdf.