Park et al., Int J Cancer Immunol Immun 2015, 1: 104
http://dx.doi.org/10.15344/ijcii/2015/104

International Journal of
Cancer Immunology & Immunotherapy

Commentary Open Access

# An Overview of Fuzzy Set Theory Based Approaches to Analyze Cancer Gene Expression Data

**Hye-Young Jung and Taesung Park***

*Department of statistics, Seoul national university, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea*

**Publication History:**

Received: October 14, 2015
Accepted: November 12, 2015
Published: November 14, 2015

**Keywords:**

Gene expression, Fuzzy set theory, DNA microarray

The development of DNA microarray technology has enabled researchers to monitor the expression profiles of thousands of genes simultaneously, which has proven to be a useful tool for researchers wishing to gain insight into cancer diagnosis, predicting clinical outcome of cancers, and new targets for cancer therapies. Since the massive amount of the gene expression data produced by this technology belongs to the class of high dimensional low sample size data with noisy, it has offered opportunities and challenges for statistical and machine learning methods including regression, clustering, classification, dimensionality reduction, and neural network.

Clustering is a useful exploratory technique for gene expression data as it groups similar objects together and allows the biologist to identify potentially meaningful relationships between genes [1]. Generally, classical clustering methods based on classical set theory allow a gene to be assigned into only one cluster.

In the classical set theory, an element either belongs or does not belong to the set. That is, a classical set is defined by a crisp boundary, which is usually called a crisp set. Most of classical methods based on classical set theory are described by crisp decision boundaries. By contrast, fuzzy set theory, which was firstly introduced by Zadeh[2] in 1965 as a generalization of classical set theory, allows a partial degree of membership between 0 and 1. In other words, an element may partially belong to the set and a fuzzy set is defined by a fuzzy boundary. Thus, a fuzzy set theory offers a tool for representing classes with soft boundaries.

For example, one of the most common analyses in microarray data analysis is to identify differentially expressed genes (DEG) between normal and tumor tissues. The DEGs are candidate cancer biomarkers which are up-regulated in tumor tissues but down-regulated in normal tissues and vise versa.

In this case, membership criteria of classes to be separated may not be precisely defined and may indeed be improperly defined based on an arbitrary threshold of expression required for classical approaches. Then, we can transform gene expression levels into three classes using fuzzy sets, down regulated, constitutive, and up regulated, despite gene expression (the x-axis) being a continuous variable (Figure 1). Point A in the graph represents the condition of 'not-likely to be up regulated', 'slightly constitutive' and 'mostly down regulated'. Similarly, Point B on the graph represents the condition of 'not-likely to be down regulated', 'slightly constitutive' and 'mostly up regulated [3].

Fuzzy sets are a natural way of characterizing non-probabilistic uncertainties as well as inherent imprecision's in the biological data.

In gene expression data, the boundary of a cluster can be regarded as fuzzy boundary for following reasons; One is that the gene expression data often contain noisy and missing values. Moreover, the similarity between expression levels of genes is continuous and there is no crisp cutoff value for group membership. In addition, a gene might express similarly to more than one gene under different sets of samples. Thus,

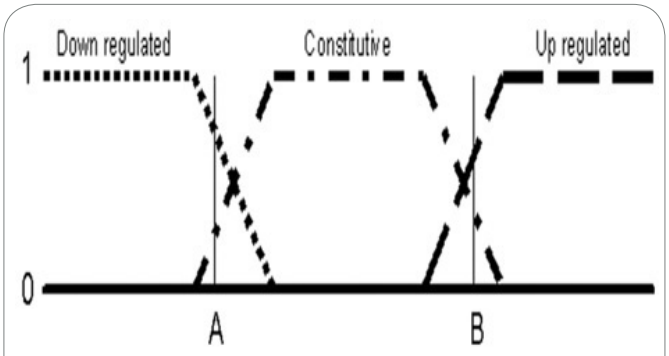


Figure 1: Example of fuzzy gene expression levels. (Gary B. Fogel Brief Bioinform 2008;9:307-316)

genes can belong to multiple clusters partially and simultaneously with different membership degrees. Therefore, fuzzy clustering which allows genes to simultaneously belong to multiple clusters is more applicable to gene expression data.

Among fuzzy approaches built by fuzzy set theory, fuzzy machine learning methods and fuzzy inference systems based on fuzzy if-then rules are well known and useful tools which have been used in gene expression analysis. Recently, fuzzy machine learning methods such as fuzzy clustering, fuzzy support vector machine, and fuzzy neural network have been successfully applied to cancer gene expression data for extracting cancer related genes and cancer classification.

The fuzzy inference systems can provide the qualitative aspects of human knowledge and reasoning processes using fuzzy rules. Fuzzy rules are a collection of linguistic statements that describe how the fuzzy inference system should make a decision regarding classifying an input or controlling an output. The process of fuzzy inference system is presented in Figure 2. The fuzzy inference system consists of

***Corresponding Author:** Dr. Taesung Park, Department of statistics, Seoul national university, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea, E-mail: taesungp@gmail.com

**Citation:** Jung HY, Park T (2015) An Overview of Fuzzy Set Theory Based Approaches to Analyze Cancer Gene Expression Data. Int J Cancer Immunol Immun 1: 104. doi: http://dx.doi.org/10.15344/ijcii/2015/104

three main stages: fuzzification, inference, and defuzzification. For example, input variable is a size of tumor and output variable is a risk of cancer. As well known, a size of tumor and a risk of cancer are not precisely defined. Since the crisp sets with crisp boundaries can cause loss of information in data, the fuzzy sets are appropriate to represent these types of data. Thus, fuzzy inference system is used for finding a risk of cancer from a size of cancer using fuzzy if-then rules. First, we define three fuzzy sets of size of tumor: small, average, and large along with three fuzzy sets of risk of cancer: low, moderate, and high. Second, we make up the all possible fuzzy rules from the given data.

Rule 1: If size of tumor is small, then risk of cancer is low.
Rule 2: If size of tumor is average, then risk of cancer is low.
Rule 3: If size of tumor is average, then risk of cancer is high.
⋮
Rule k: If size of tumor is large, then risk of cancer is high.

Then, we choose rules from a fuzzy rule set to infer the predicted fuzzy risk of cancer. Finally, rules applied individually are aggregated by using fuzzy union operation with max function and then we defuzzify predicted fuzzy output to get crisp output.

Since it may be natural to postulate that everything in biological problems is fuzzy, the fuzzy inference systems play an important role in cancer study.The fuzzy approaches based on the fuzzy inference system, as qualitative computational approaches, provide formalized tools to deal with the imprecision intrinsic to many biological problems. In addition, the hybrid approach combining of classical approach and fuzzy approach can exploit the advantages of both approaches. This can not only provide simple classification results, but also easily be explained and interpreted by understandable fuzzy rules. There are many fuzzy or hybrid approaches to identifying cancer related information regarding overlapping clusters and overlapping

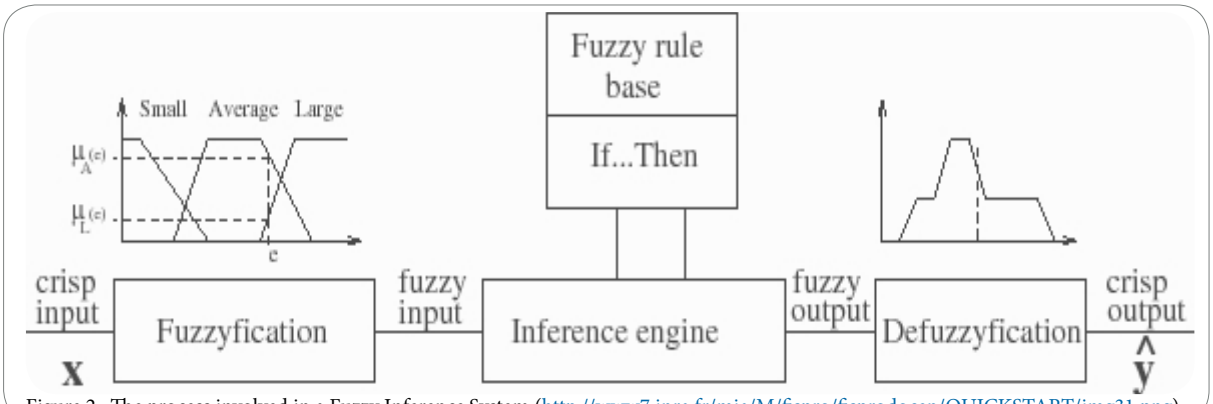


Figure 2 : The process involved in a Fuzzy Inference System.(http://www7.inra.fr/mia/M/fispro/fisprodocen/QUICKSTART/img31.png)

cellular pathways such as fuzzy neural network, fuzzy c-means clustering, fuzzy gene regulatory network and fuzzy rule classifier. The following are some examples.

1. Clustering analysis of expression profile using fuzzy adaptive resonance theory [4].
2. The application of a local search heuristic, Fuzzy J-Means, embedded into variable neighborhood search metaheuristic for the clustering of microarray gene expression data[5].
3. To attribute cluster membership values to genes applying a fuzzy partitioning method, fuzzy C-means[6].
4. To analyze gene expression data using fuzzy logic [7].
1. To classify cancer microarray data using a fuzzy standard additive model for cancer classification based on gene selected by analytic hierarchy process [8].
2. To select multiple highly informative gene subsets for cancer classification and diagnosis using a new Fuzzy Granular Support Vector Machine-Recursive Feature Elimination algorithm[9].
3. To produce interpretable classifier with knowledge expressed in terms of if-then rules and membership function using the fuzzy expert system[10].
4. To develop some multi-objective evolutionary algorithms based Interpretable fuzzy methods for analyzing high dimensional microarray gene expression cancer data sets[11].
5. To survey the fuzzy methods used in gene regulatory networks (GRNs) inference[12].
6. To show how fuzzy logic can be applied using information from genes previously shown to be important[13].
7. To classify cancer tissue using evolving fuzzy neural networks [14].

As Zadeh said, "Fuzzy sets can play a very important role in human thinking, particularly in the domains of pattern recognition, communication of information, and abstraction"[2]. Since cancer studies include these types of domains, fuzzy approaches are expected to play an important role in cancer studies.

To date, many fuzzy approaches have been successfully applied to microarray data to identify cancer related genes. However, most analyses based on classical methods have ignored the fuzzy nature of microarray data.We expect the hybrid approaches combining fuzzy approaches with other classical methods such as artificial neural networks and/or genetic algorithms as well as fuzzy approaches would be quite effective to identify cancer related genes and/or proteins.

## Conflict of Interest

The authors declare that they have no competing interests.

## Author Contributions

Both the authors substantially contributed to the drafting of the manuscript.

## Funding

This work was supported by the Industrial Strategic Technology Development Program (#10045352), funded by the Ministry of Knowledge Economy (MKE, Korea).

## References

1. Ghosh S, Dattagua R, Mitra S (2014) Fuzzy clustering with biological knowledge for gene selection. Applied Soft Computing 16: 102-111.
2. Zadeh LA (1965) Fuzzy Sets. Information and Control 8: 338-353.
3. Fogel GB (2008) Computational intelligence approaches for pattern discovery in biological systems. Brief Bioinform 9: 307-316.
4. Tomida S, Hanai T, Honda H, Kobayashi T (2002) Analysis of expression profile using fuzzy adaptive resonance theory. Bioinformatics 18: 1073-1083.
5. Belacel N, CuperloviÄ‡-Culf M, Laflamme M, Ouellette R (2004) Fuzzy J-Means and VNS methods for clustering genes from microarray data. Bioinformatics 20: 1690-1701.
6. Dembélé D, Kastner P (2003) Fuzzy C-means method for clustering microarray data. Bioinformatics 19: 973-980.
7. Woolf PJ, Wang Y (2000) A fuzzy logic approach to analyzing gene expression data. Physiol Genomics 3: 9-15.
8. Nguyen T, Khosravi A, Creighton D, Nahavandi S (2015) Hierarchical gene selection and genetic fuzzy system for cancer microarray data classification. PLoS One 10: e0120364.
9. Tang Y, Zhang YQ, Huang Z, Hu X, Zhao Y (2008) Recursive fuzzy granulation for gene subsets extraction and cancer classification. IEEE Trans Inf Technol Biomed 12: 723-730.
10. Ganesh Kumar P, Aruldoss Albert Victoire T, Renukadevi P, Devaraj D (2012) Design of fuzzy expert system for microarray data classification using a novel genetic swarm algorithm. Expert Systems with Applications 39: 1811-1821.
11. Wang Z, Palade V (2011) Building interpretable fuzzy models for high dimensional data analysis in cancer diagnosis. BMC Genomics 12 Suppl 2: S5.
12. Al Qazlan TA, Hamdi-Cherif A, Kara-Mohamed C (2015) State of the art of fuzzy methods for gene regulatory networks inference. ScientificWorldJournal 2015: 148010.
13. Ohno-Machado L, Vinterbo SA, Weber G (2002) Classification of gene expression data using fuzzy logic. Journal of Intelligent and Fuzzy Systems 12: 19-24.
14. Futschik ME, Reeve A, Kasabov N (2003) Evolving connectionist systems for knowledge discovery from gene expression data of cancer tissue. Artif Intell Med 28: 165-189.