International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-9, Issue-6, November 2021
https://doi.org/10.55524/ijircst.2021.9.6.15
Article ID IRPS12025, Pages 65-68
www.ijircst.org

# A Review Paper on Fuzzy Logic Algorithms

**Mrinal Paliwal[1], and Pankaj Saraswat[2]**

[1,2] SOEIT, Sanskriti University, Mathura, Uttar Pradesh, India

Correspondence should be addressed to Mrinal Paliwal; mrinalpaliwal.cse@sanskriti.edu.in



**ABSTRACT**-Logic for gene expression analysis in a flurry. We developed a new method for analyzing gene expression data. To convert expression values into Quality descriptors, this method uses a fluid logic that can be evaluated using heuristic rules. We developed a model for identifying three different activators, repressors, and objectives in a data set for yeast gene expression in our experiments. The test predictions generated by an algorithm match the experimental data in the literature very well. Algorithms can identify a much larger number of transcription factors that could be identified at random in defining the function of unspecified proteins. Using only expression data in the form of clustering, this method allows the user to construct a linked network of genes. The interpretation of gene expression categorization models is typically difficult, however, it is an essential component of the analysis procedure. In five databases ranging in size, experimental origin, and physiological field, we investigate the effectiveness of micro rules fuzzy systems. The classifiers resulted in regulations that are simple to understand for biomedical researchers. The classifiers resulted in regulations that are simple to understand for biomedical researchers.

**KEYWORDS**-Algorithm, Data, Fuzzy logic,Gene, Models.

## I. INTRODUCTION

Common data extraction approaches such as kernel supporting linear computers, machine learning, and logistically extrapolation were used to categorise biological information. They create models that are difficult for biological and clinical scientists to comprehend due to the large number of elements and factors. It will be much simpler to understand systems if straightforward but exact principles could be generated from a little amount of learning information. People can grasp basic rules like 'when A is elevated and B is downregulated,' which are easier to comprehend than formulas with several variables, interactions circumstances, and variables; some writers recommended adopting regulations approaches to interpret macroscopic information[1].The development of more sophisticated fuzzy models is hampered by computational time. Pre-processing data might be a solution to the problem. If three genes can be found which are unlikely to match the model before the algorithm is operated, the improbable triplets may be disregarded if the method is executed, thereby reducing the number of triplets not evaluated in future stages to about a quantity of time.

Various algorithms rely on a restricted number of categories to discretize the data. Discretization is also beneficial since it simplifies the interpretation of data. The accuracy of the values acquired by microarrays is excellent.However, who should be addressed in the communication. It is disputed the sense of this apparent accuracy. The scanners measuring the grade of fluorescence in the different coloured channels read the values. The process is susceptible to saturation and other errors;therefore, biologist is typically dependent on experiments to gain a notion of the quantity of mRNA, rather than absolute gene expression data. Moreover, this reduction might be used to construct simple rules which can be readily comprehended by people rather than actual numbers multiplied by particular coefficients in order to generate a classification device. In order to induce these rules from the data, 2 types of algorithms are required: algorithms for categorising continuous values and algorithms for rule discovery and filtering resulting in compact, readable rules. Discretization ofCrispdoes not take into consideration that boundary values across value categories might be extremely similar (in opposition to fussy discretization)[2][3].For example, it is feasible to identify a sample at the top end of the "low" category by combining membership values in multiple categories. Please note that fluid memberships are not likely and their operators differ from those in that context[4].

### *A. Clustering to Improve Run Time*

The grouping of genes in relation to expression is a first step towards quick and thorough data interpretation. The clustering of genes with comparable expression patterns is commonly employed. We can try to use gene clusters to provide gene dataset metadata. It is improbable that any genes with comparable expression profiling is will fit the model well if a specific mix of clusters does not match the model well.

Since we assume that, the data may be clustered so that most of the expression profiles of genes are comparatively close to cluster centres, the cluster centres, and their respective gene profiles are comparable, and the MSE difference is low.Thus, if the model does not fit in well with a group of cluster centres, the genes around these cluster centres will not fit well into the model[5].

Therefore, we can save enormous calculation time by not examining these combinations in the foggy model.
Although clustering does not let us to identify direct connections between genes, it enables us to successfully limit the dataset to a few time series that reflect the data in general generally.If the standard difference between genes and their associated cluster centre is minimal, we can say that the expression profile of the cluster centre is roughly identical to the cluster genes[6].

### *B. Improved Algorithm*

In this part, we discuss procedures to evaluate gene expression data in the modified Woolf and Wang method. First, the method is executed utilising the triplet cluster centres with no error or variance limitations. In reference to the model, each triplet of cluster centres is classified according to its inaccuracy. Then the algorithm runs on the genes. A file with all triplets of the cluster and their errors with respect to the model is read before gene analysis. The target separates the triplets of the cluster. There are two approaches of evaluating the centres of the clusters they belong to:

- Method of percentage ranking. For the target gene cluster, the matching cluster triplet needs to be above a specific percentile. The model error for cluster triplets determines the ranking, with a smaller error meaning a better rank.
- Method of error threshold a mistake score must be below a previously defined threshold in the respective cluster combination. If the associated triplet cluster of the triplet gene does not meet the threshold given, triplet.is not examined and the algorithm continues to the following triplet.

### *C. Developing a Generalized Model*

The generalised model is founded on the concept that reactants should be limited. To create the complex, all proteins in a complex must be present. If one or more of the genes is not strongly expressed, the proteins encoded by them will not be highly expressed, which will lead to poor complex expression since certain component proteins are lacking. If not all activators or repressors needed for activation or repression of the complex target gene are not highly expressed, complexes are unlikely to have a meaningful impact on target gene expression.
Let X be a Gene expression matrix that comprises gene expression profiles (in the simple model, X consists of two vectors $x_a$ and$x_r$representing two expression profiles, i.e. activator and repressor, respectively), and let y be a model output, let us define y=f(X). The y output is the optimal expression profile of the objective gene. Let z be a vector that displays the actual target gene expression profile. We can extend $x_a$, $x_r$, to $X_a$ and $X_r$, which are matrices containing an arbitrary number of activator or repressor profile vectors.
Suppose there exist vectors $x_{ma}$ and $x_{mr}$,
Where$x_{mai}$ = min ($x_{a2i}$, $x_{a2i}$, … … …$x_{aji}$)
$x_{mri}$= min ($x_{r2i}$, $x_{r2i}$, … … …$x_{rji}$)
And
Wherej is the activator number in Xij k is the repressors number in Xij and i=I, 2.
A' where N is the expression profile number. Now, xij and x contain the minimal level of expression for each of Xi and Xj's genes, at every point of expression. We may suppose that, because of the limiting reactants, xij and x are an expression profile of the activator/pressure complex.
2.2 Fuzzy rules
If U is a collection of samples of tissues, let G = $\{g_j\}_j$ be a set of symbols of a gene, let C be a set of class labels and allow c: U as C be a partial function, which applies classes to U as a result of U as a whole. Let g(x) indicate the value of the tissue sample gene g expression x. According to the aforementioned assumptions, genes are controlled at a qualitative level such as up (u), neutral (n) or downwards (d). Let L be the collection of such standards. We combine a fluid set and allow μ(l, G, x) to form part of X with the gene g and level l for each gene or qualitative level.
May the aboveground descriptors be D = G = L. In a proposal context, the descriptor d=(g, l) is often seen in U {x alternate with U|g(x) = l{}. This allows d to be used for a set of elements in U for which g(x) = l may be seen as a function. This view is expanded to a member function, allowing g(x) in l to include the descriptor d. In other words, with the descriptor d = (g, l) we have this d(x) = μ (l, g, x). This allows the typical crook-free combination and disconnection, as less (and max) of 2 feature functions, to be extended straight to the standard flush case.
The R = 2D = C rule is defined as an element. We designate the antecedent D by ant(r) for a rule r = (D, c), and thus c of r by cons (r).
The application r(x) of a rule r = (D, c) in respect of the element x = U shall be defined as r(x) = min d in respect of the element D(d(x) in question. (1) We consider that r(x) is the x membership of class c in accordance with r. Our idea of membership is extended to a collection of R rules and is classified accordingly. The c-type of a c-type is μR(c, x) = max ({0}).
By selecting one with a maximum membership, we may now assign a class to x. In some cases, if you cannot be sufficiently sure that a case is a class, it is occasionally helpful to be able to disregard a categorization. This is how we implement this concept. If x > 00 else, leave s(x) = x.
Also add a threshold tc to every class label c = C. The tc threshold is the threshold for which class c is rejected. What we mean by categorization may now be defined properly. Given the corresponding tc criteria, the categorization of x to be determined by R is defined. Functions for length three sequence classes with R maximum membership. Formally, classR(x) = arg max c∈C (s (μR(c, x) − tc)).
If classR(x) contains more than one element, we decide to reject the classification. This happens if either all classes have been rejected or many classes have the same maximum membership for x.

D. ***Learning Membership Functions and Rejection Thresholds***

In order to apply a set of rules, we must know in the rules the membership functions that are consistent with the descriptors and the refusal levels for each class label.

We present a simple system, which allows us to learn both the membership features and the refuse thresholds given a set U = U.

The triangular functions result in a sequence of length three. Our set of labels L are now ordered to have a quality of the meaning of the li mark lower than the lj mark for 0 < i<j to n − 1. An instance is that of L = {l0, l1}, in which l0 = "down" and l1 = "up." Increased sequence of real numbers sg n = |L| is associated to each gene g. We are now allowing μ(li,g,x) = μi(x), where μi for sequence sg is described above. We propose to utilise quantiles in the gene g sequence U over observed expression values, assuming expert knowledge dictates nothing else.

We used n = 2, v0 = min (g (U)) and v1 = max (g (U)) in our experiments.

Then the tc threshold is the minimum non-zero assignment across U. We suppose that it is useful to carefully pick the training set U, preferably in cases where partial function c is specified[7,8].

They applied each conceivable gene combination of activators and repressors using standardised Saccharomyces cerevisiaedata[9,10].

SOM is a neural network that maps the data from a multidimensional one or two-dimensional space into a discrete one. It is a strong, scalable, adaptable and rather quick approach. In addition, comparable profiles are provided in neighbouring clusters and progressively varied clusters in distant clusters.

We have grouped four datasets in our experiments: Yeast, CNS and yeast elutriation, as well as cdcl5 data using SOMs with varied node counts. We run Woolf and Wang's algorithms at the cluster centres, and classified the cluster three times, according to how well they match the fluid model.If only gene triplets whose respective cluster centre trials exceed a given cut-off were evaluated, it would be improbable that the three triplets will match the activator/repressor model if the gene triplets of the corresponding cluster centres. The test was carried out on four datasets of distinct clusters in the SOM and varied cut-off limits.

## II. DISCUSSION

Overall, the algorithm results correspond well with the literature's experimental data. This should not come as a surprise as the algorithm seeks connections that are in keeping with our knowledge of the interaction between an activator, a repressor and a target. Thus, the fuzzy logic algorithm approximates the mental process an expert would use in analysing this data, based on essentially the same criteria that an investigator may use to define the regulatory function of a protein. However, the algorithm is automated, impartial and broad in contrast to an expert. It can be difficult to evaluate expression data, and it can easily be misinterpreted if not carefully analysed.We have created a "lens" through which the results may be sorted without problems, quickly and efficiently, using a calculation method for data processing.

Although the algorithm has been employed in this work to search for just three times the activator, repressor and target genes, other method variants are also conceivable. An easy way of showing that this technology can provide physiologically relevant outcomes is to use the activator repressor model.The technology, however, is generic and may be extended to more complex interactions and systems.Examples of these include additional connection types, such as coactivators or intricate systems involving genes whose production by a number of transcription factors has been controlled in a complex manner. The technique may also potentially be expanded to represent full generic networks of gene connections based just on data from expression.However, there are several limits to the use of fuzzy logic to analyse expression data. The interaction between many proteins is largely linear in the initial analysis and hence the programme looked for linear behaviour. However, this linear approximation is not correct in the situation of many redundant promoters.

It might remedy this problem by taking a more advanced "fuzzification" to incorporate nonlinear effects; but, this additional complexity only remedy for a few missing connections while eliminating many of the more typical close-linear connections.The aim of this method is also to draw broad patterns that relate to numerous genes rather than to provide quantitative previsions. Thus, inclusion of certain nonlinear effects would not assist to establish many links, but would increase an already tough task to a substantial computing load.In the functions of activators and repressors the flush logic method identified a disproportionately large number of transcription factors, however not all activators and repressors were found transcript factors. Two probable causes of this variation include 1) low-level transcription factors, which are difficult to detect, and 2), which are indirectly regulatory for transcription in other gene products such as enzymes.Transcription factors normally only occur at extremely low concentrations; hence variations for expression of the transcription factor cannot be recognised by existing approaches of profiling.Probably the fluid-logic algorithm might identify many more differences from transcription factors in the activator and repressive functions, if expression-profiling technology became more sensitive. In many situations, however, the amount of expression of a specific protein does not depend on the transcription factor expression, but on the concentration of certain intracellular compositions.

In these situations, the variations in the expression level of the enzyme will be recognised by the algorithm as activating or repressing impact and are "transcription-factor-like."These enzymes are maybe more important in drug design than real transcript factors since the activity of the cytosol in the cytosol is usually easier to modify with a substance instead of blocking a true transcript factor in the nucleus. In addition, the collection of data used in this study was based on a single experiment where the major study procedure was cell cyclic control.Transcription factors not engaged in pathway associations with this cellular function did not demonstrate substantial changes in their expression and

so the Fuzzy Logic Algorithm could not be assessed.We analyse a data set that covers genetic expression patterns of both wild and different mutant yeast cells in order to conduct a more thorough assessment of transcription factors. Many additional transcription factors may be assessed since they have disrupted the mobile processes they govern.An additional benefit to the fuzzy logical algorithm is that information may originate from any source inside an organism (tissue, cell type, therapy or physiological condition). This improvement has been due to the necessity for the algorithm to monitor changes in protein expression level relative to changes in other levels of expression.The varied degrees of expression of each new data set may be examined to see if they are suitable for the proposed regulatory model. In our research, several data sets were only discarded because the combinations of expressive levels (too high a sigma value) were not properly explored and their forecasts could not be believed.

The major use of this method is to validate or identify pharmacological targets independently. Traditional drug target identification approaches require an in depth understanding of the biology of the illness, which may be slow and hard to achieve.By contrast, expression profiling is a fast, high-performance method that provides a lot of cell information in a form that can be handled simply on a computer. The mechanism of a known objective may be confirmed by means of a fuzzy logic technique for analysing expression profile data. In addition, as the fumbling logic method requires no biological gene knowledge, genes with unknown functions may be equally simply added as genes with established activities.In the development of medication targets, this capacity to detect functional indications of uncharacterized genes is a major benefit as prospective drug targets may be traced to the detailed biology.

## III. CONCLUSION

In this work, we show that clustering can save substantial time as a pre-processing step in the build-up of a fluid, logic-based model. In contrast to a modelling strategy employed in past years that examines all gene combinations, our method uses self-organized mapping cluster centres to find genes that can interact. This method accelerates the process of modelling gene interaction and enables the development of sophisticated models, including co-activators and corepressors. This will increase the attraction of fuzzy techniques for reversing the transcriptomic networks. There is presently no precise method for determining the numbers of nodes to keep and the fraction of clustered permutations to keep. We observed that the frequency of clustering in a collection from one of the four datastore searches used should be at least 50 percent the quantity of available points. The increase beyond that number of clusters leads in several clusters with identical characteristics, which offer little information to the clusters. The author is looking at using a double self-organizing map (DSOM) and adaptable frequency theories to find the right amount of groups. When it comes to picking the proportion of clusters permutations to be kept, 67 percent appears to be adequate; there are significant increases around 50 and 67 percent, but very little, if any, increase among 67 and 75 percent. The methods proposed in this paper can open the path for the construction of a generic gene regulatory model that can be applied to any group of proteins.The enhancements will enhance the practicality of using fuzzy logic to analyse genes' relations using existing micro array methods.

## REFERENCES


[1]. Sahoo A, Dash JK. A Novel Method for Optimal Solution of Fuzzy Chance Constraint Single-Period Inventory Model. Adv Fuzzy Syst. 2016;

[2]. Butterworth R, Simovici DA, Santos GS, Ohno-Machado L. A greedy algorithm for supervised discretization. J Biomed Inform. 2004;

[3]. Gbolahan K. Williams. Architecting Tacit Information in Conceptual Data Models for Requirements Process Improvement. Kings College London. 2013.

[4]. Ohno-Machado L, Vinterbo S, Weber G. Classification of gene expression data using fuzzy logic. J Intell Fuzzy Syst. 2002;

[5]. Das D, Chakravorti T, Dash PK. Hilbert huang transform with fuzzy rules for feature selection and classification of power quality disturbances. In: 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics, UPCON 2017. 2017.

[6]. Acharya S, Ranarahu N, Dash JK, Acharya MM. Solving multi-objective fuzzy probabilistic programming problem. J Intell Fuzzy Syst. 2014;

[7]. D'haeseleer P, Wen X, Fuhrman S, Somogyi R. Linear modeling of mRNA expression levels during CNS development and injury. Pac Symp Biocomput. 1999;

[8]. Woolf PJ, Wang Y. A fuzzy logic approach to analyzing gene expression data. Physiol Genomics. 2000;

[9]. Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, et al. A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell. 1998;

[10]. Soinov LA, Krestyaninova MA, Brazma A. Towards reconstruction of gene networks from expression data by supervised learning. Genome Biol. 2003;