fuzzy reasoning,
fuzzy truth value,
fuzzy systems

Przemysław KUDŁACIK[1]

AN ANALYSIS OF USING TRIANGULAR TRUTH FUNCTION IN FUZZY REASONING BASED ON A FUZZY TRUTH VALUE

Fuzzy systems are widely used in research and applications considering complex information like gene recognition and classification. Because of the character of genetic data, the extensive knowledge bases of such systems contain complex rules described by even thousands of atomic premises. This paper presents an analysis of fuzzy reasoning based on a fuzzy truth value, presented by Baldwin. The solution is an interesting, alternative approach to fuzzy inference. Considering the Zadeh's compositional rule of inference, the idea of Baldwin has an advantage of resolving the whole inference process within the truth space. The approach is especially convenient for systems with large number of premises in rules, like mentioned gene classification systems. Although the solution of Baldwin is characterized by significantly lower computational complexity than the full implementation of the compositional rule of inference, it is not applied in contemporary systems. Over the years different researchers proposed simplified approaches, which are easier to implement and faster. The analysis presented in this paper considers possible simplifications that could be applied to the approach of Baldwin, where facts and fuzzy truth values are described by triangular membership functions. Such assumptions open the possibility of implementation of fast Baldwin's inference and applications even for complex genetic data. Nevertheless, the solution would preserve one of the biggest advantage, which is the fuzzy relation, in form of the truth function, between a fact and a premise, throughout the whole inference process. Other fast approaches reduce this relation to only one value.

1. INTRODUCTION

Fuzzy systems are widely used in many areas of engineering for a couple of decades [7], [18], [21]. Their popularity is based on simple and clear approach to uncertainty, which is inevitable in measuring and examining processes of surrounding world [34], [35], [36].

Throughout the years researchers developed different approaches to the problem of fuzzy inference. The theoretical basics were introduced by Zadeh [34] and since then other solutions were proposed. The most popular in applications are approaches presented by Mamdani and Assilan [21], Larsen [18], Takagi, Sugeno and Kang [28], [27] as well as solution of Tsukamoto [29].

In the yearly years of the first fuzzy systems an interesting idea derived directly from classical logic was presented by Baldwin [1], [2]. Unfortunately, probably because of it's complexity in comparison with other approaches, the solution based particularly on a fuzzy truth value was forgotten. Although the idea of a fuzzy truth value used in such form for a fuzzy inference was considered by Belman and Zadeh [3] as well as Dubois and Prade [8], [9], the approach was not widely applied.

Nevertheless, the approximate reasoning introduced by Baldwin in 1979 has advantages, especially in case of systems with large knowledge bases [14]. Such problems are very often encountered in different

[1]University of Silesia, Institute of Computer Science, ul. Będzińska 39, 41-200 Sosnowiec.

research and applications considering complex information. Gene recognition and classification are typical examples, where rules in the knowledge base consist of even thousands of atomic premises. The problem of gene classification is extensively studied using fuzzy systems by many researchers throughout the recent years. The following papers show that the problem is very important and valid [4], [5], [6], [10], [11], [12], [13], [19], [20], [22], [23], [24], [25], [26], [30], [31], [32], [33].

The aim of this paper is to propose a faster approach of Baldwin's fuzzy reasoning using triangular membership functions of premises and triangular truth functions. The analysis presented in the following sections is a direct continuation of [16], which considers the computational complexity of Baldwin's approach. The conclusion of the previous work directed future research to simplification of obtaining truth function of a premise and compound truth function. These operations invoked repeatedly determine the output computation time in case of mentioned rules with multiple atomic premises.

A general structure of subsequent sections is organized in the following order. First, the approach of Baldwin is shortly presented. Next section describe the problem of obtaining a piece-wise linear function on the basis of algorithms implemented within the Fuzzlib library [17], [15]. Consequences of using triangular membership functions in Baldwin's reasoning and simplified junctions are analyzed afterwards. The final sections contain the analysis of output computational complexity for simplified approach, summarized with the conclusion.

2. THE FUZZY INFERENCE OF BALDWIN

The process of inference based on the fuzzy truth value can be broken down into four stages. The first involves obtaining a truth function of a premise, which directly reflects conformity of a fact and a premise. The second phase is responsible for obtaining one compound truth function in case the rule has more atomic premises.

Assuming the following rule *"if humidity is high and temperature is medium, then fan speed is medium"* the first phase would be responsible for calculating two truth functions for two atomic premises: *"humidity is high"* and *"temperature is medium"*, based on the relevant facts (the real state of humidity and temperature). In case of K atomic premises K truth functions of a premise have to be obtained. The second stage in this case would join two obtained functions into one fuzzy truth function describing overall conformity of two facts with consideration of the junction type (*and*, *or*). For K truth functions $K-1$ junction operations have to be performed.

Fuzzy truth functions (τ) are defined in $[0,1]$ ranges (considering the domain and counter-domain)

$$\tau : [0,1] \to [0,1]. \tag{1}$$

Therefore, the inference process preserves the fuzzy relation between a fact and a premise. In other simplified approaches that are widely applied, this relation is reduced into only one value from $[0,1]$ range. Obtaining a compound result in this case for multiple atomic premises within a rule involves simple operations on single values, which is faster.

After obtaining the compound truth function of the compound premise, the inference moves into the third stage: calculating the truth function of conclusion. This function is further used to modify the fuzzy conclusion and by that obtaining the fuzzy result of inference at the fourth and the last phase.

In case of the sample rule presented above (the conclusion *"fan speed is medium"*) the output truth function would transform the membership function of fuzzy expression *"medium"* into the appropriate form prepared for aggregation with results from other rules.

The last two phases are not significant for large number of atomic premises [16] and will not be further considered in this paper.

2.1. COMPUTATIONAL COMPLEXITY

The analysis presented in [16] assumed the piece-wise linear description of membership functions, which allowed to generally describe the computational complexity of the presented process by the following equation [16]

$$KC_{minDY}\log_2 N + (K-1)2C_{minDY}^2 + C_{minDY}^2 + N\log_2(C_{minDY}). \tag{2}$$

The parameter N stands for the number of fuzzy sets' description nodes (piece-wise linear membership functions). K determines the number of premises and C_{minDY} denotes the complexity of truth function description according to additional $minDY$ parameter (the smaller $minDY$ the more description nodes are needed for obtained truth function) [16].

The first two elements of (2) are involved with obtaining K truth functions of atomic premises and $K-1$ operations of obtaining compound truth function (subsequent junction of given K truth functions). It can be observed that the two stages strongly depend on value of C_{minDY}, which is a number of truth function nodes. Therefore, the analysis presented in this paper considers decreasing the number of nodes to a very low and almost constant level, needed only to properly describe all possible forms of triangular truth functions.

3. OBTAINING TRUTH FUNCTION: THE GENERAL APPROACH

The Fuzzlib library [17], [15] implements a divide and conquer algorithm in order to obtain description nodes of truth functions. Such solution was chosen because generally a truth function can take many different forms, which depend on chosen junction operator (a variety of T-norms and S-norms [7]) and type of used implication. The algorithm is relatively fast because of logarithmic computational complexity of the approach.

Subsequent steps of the algorithm are shown in Fig. 1 for three examples. The solution aims to obtain nodes of piece-wise linear functions to match desirable forms of the functions designated by dashed lines. Each column of charts show three following steps of function creation for three characteristic examples.

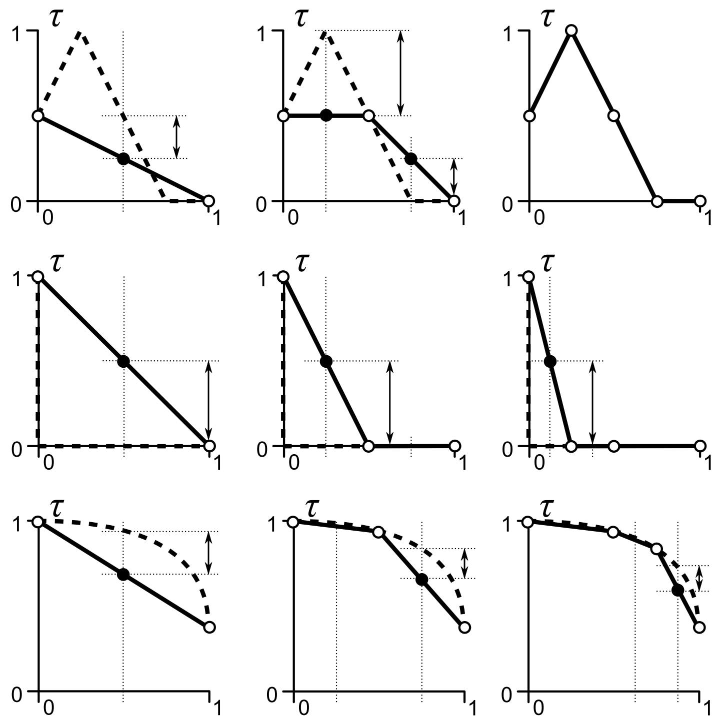

Fig. 1. Visualization of the algorithm obtaining a piece-wise linear truth function for three sample cases presented in subsequent rows.

Procedures creating each type of truth function differ only in calculations of node values. Therefore, further analysis focuses on procedures obtaining a truth function of a premise and compound truth function.

3.1. TRUTH FUNCTION OF A PREMISE

The truth function of a premise denoted by τ_P can be described by the following equation [1]

$$\underset{\eta\in[0,1]}{\forall} \quad \tau_P(\eta) = \sup_{\substack{x\in X \\ \eta=\mu_A(x)}} \Big[\mu_{A'}(x)\Big], \tag{3}$$

where μ_A and $\mu_{A'}$ represent respectively membership functions of a premise and a fact, described in the domain X. In other words, τ_P represents the largest values of $\mu_{A'}$ in these ranges of X, where μ_A takes the same values.

Therefore, obtaining one node of piece-wise linear τ_P for specified value of η needs searching through the X domain (description nodes of μ_A) and looking for points where $\tau_P(x) = \eta$. Further, for these points $\mu_{A'}$ is analyzed and it's maximum value is chosen. The complexity in this case is obviously linear and depends on number of nodes describing μ_A membership function.

3.2. COMPOUND TRUTH FUNCTION

The compound truth function τ_C obtained from two other truth functions τ_1 and τ_2 can be described by the following equation [1]

$$\underset{z\in[0,1]}{\forall} \quad \tau_C(z) = \sup_{\substack{x,y\in[0,1] \\ x\star_{T_1}y=z}} \Big[\tau_1(x) \star_{T_2} \tau_2(y)\Big], \tag{4}$$

where $\star_{T_1}, \star_{T_2}$ represent any T-norms, from which $\star_{T_1}$ models the and junction. In case of or junction the equation presented above takes the following form [1]

$$\underset{z\in[0,1]}{\forall} \quad \tau_C(z) = \sup_{\substack{x,y\in[0,1] \\ x\star_S y=z}} \Big[\tau_1(x) \star_T \tau_2(y)\Big], \tag{5}$$

where similarly, $\star_T$ represents any T-norm, but $\star_S$ is any S-norm modeling junction or. Therefore, basing on the last equation, obtaining one node of compound truth function (given value of z) involves analysis of range where $x \star_S y = z$ and finding the maximum value of $\tau_1(x) \star_T \tau_2(y)$ results. Implementation of this approach would also be characterized by a linear time complexity depending on the number of nodes in truth functions (an analysis of whole description of τ_1 and τ_2 in the worst case).

4. TRIANGULAR MEMBERSHIP FUNCTIONS

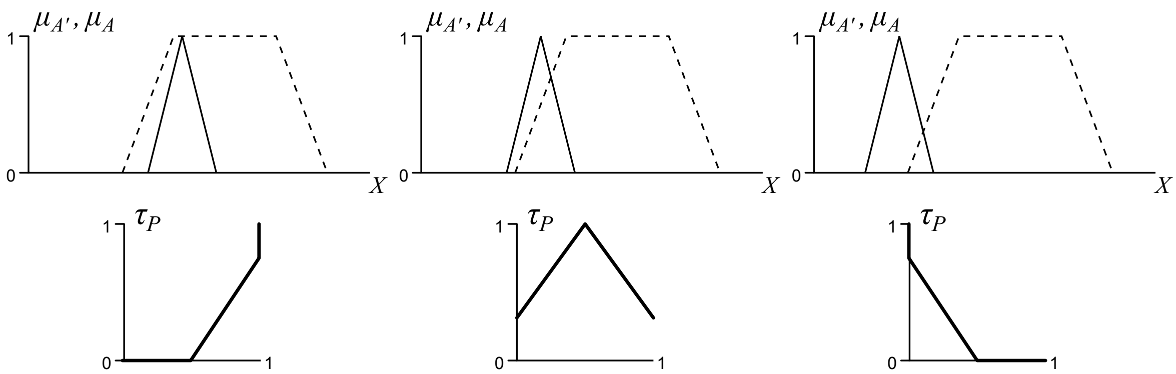

Fig. 2. Simplified description of truth functions in case of triangular membership function of a fact and trapezoidal membership function of a premise.

In many cases the uncertainty of a fact is modeled by using a triangular membership function. It is particularly convenient in piece-wise linear description of membership functions, because the number of nodes is the smallest. Considering a trapezoidal description of premises, which is also very popular, the description of truth functions is also simplified. The problem is depicted in Fig.2, where three truth functions were presented for three sample relations between facts and premises.

The aim is to eliminate the process of finding description nodes to meet the certain level of precision and provide the solution based on only several nodes. However, in this case their position can be obtained directly from characteristic nodes of analyzed fact and premise. Such approach would be much more efficient and in case of using triangular facts and trapezoidal premises even equal to the full approach.

4.1. SIMPLIFIED TRUTH FUNCTION OF A PREMISE

The Fig.3 presents sample truth functions of a premise. Depicted situations cover all characteristic types of τ_P assuming described environment, which is triangular membership function of a fact and trapezoidal membership function of a premise.

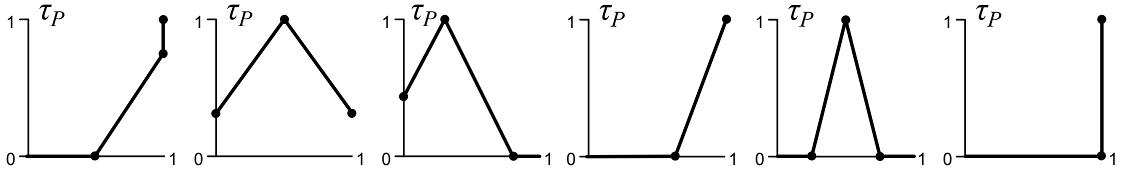

Fig. 3. Sample triangular truth functions and nodes needed to describe them.

It can be noticed, that each type of situation can be described using just three nodes and sometimes even two. The characteristic nodes of fact and premise responsible for the shape of produced truth function are marked in Fig.4 for three cases. Subsequent situations show transition from low to high conformity between fact and premise. In each presented case the most important are values marked as a, b and c.
Therefore, analyzing the cases, it can be observed that for computation purposes the nodes describing the fact should be confronted with nodes representing slopes of membership function of the premise. This involves locating the position of the fact triangle according to the appropriate slope of the premise and then obtaining correct values. Assuming the larger width of the slope than the width of the triangle, only 4 nodes in both functions need to be processed.

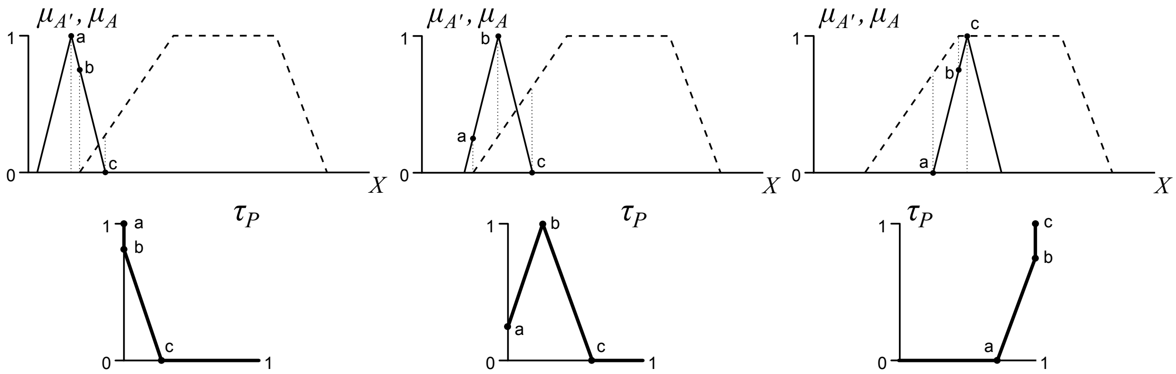

Fig. 4. Characteristic description nodes and values of fact and premise during calculations of truth function.

It must be mentioned, that such approach allows to use also other types of membership functions, like i.e. Gaussian. The solution would analyze only three points of each function (assumed beginning point, the peak and the end point), just as in the triangle case. However, the specifics of Gaussian slopes would be appropriately reflected in the first and the last produced nodes of the truth function and will be different than simple trapezoidal and triangular case. Although, the result in this situation would not be precise the general idea of fuzzy truth function describing a relation between facts and premises would be preserved.

4.2. SIMPLIFIED COMPOUND TRUTH FUNCTION

The definition of compound truth function provided by (4) and (5) does not clearly indicate the behavior of junction operations. Equations are directly obtained from Zadeh's extension principle, that extends junction of two precise values to junction of two fuzzy sets: truth functions. The flexible implementation of the process is also quite problematic because of the possibility of using any T-norms and S-norms. Due to the big difference in analysis for different norms the Fuzzlib library contains only several available functions (only continuous): minimum, maximum, product, probabilistic sum, Einstein's t-norm and s-norm and t-norm of Hamacher.

The simplified version of truth functions described mostly by three nodes allows to consider a simplified approach to composition. The general idea is depicted in Fig.5, where results of junction with "*or*" and "*and*" operators are presented. All examples use minimum and maximum as junction operators. Each row of functions corresponds to different example. The first and the second column contain parameters of composition, while the third and the fourth contain the results of "*and*" and "*or*" operations respectively.

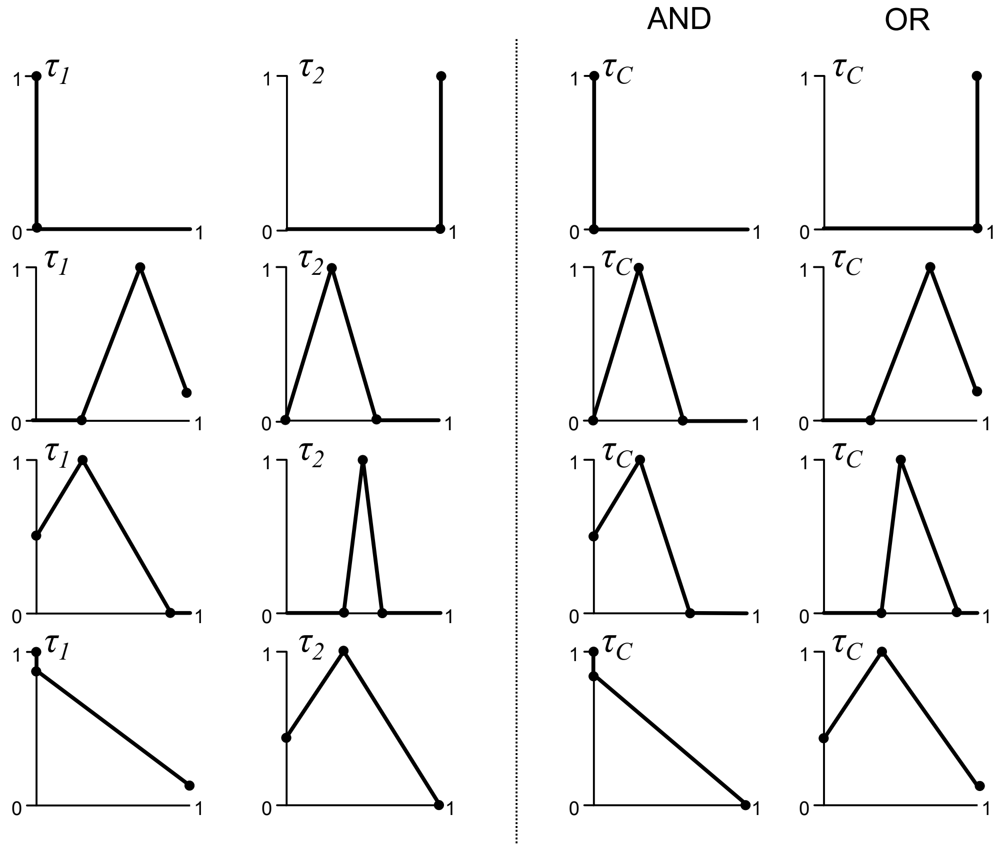

Fig. 5. Visualization of simplified composition of two truth functions.

First situation from the top show two extreme truth functions. $\tau_1 =$ "*absolutely false*" represent full inconsistency of some fact with a premise: triangular description of the fact is either on the left or on the right side of the trapezoidal description of the premise

$$\underset{\substack{x\in X \\ \mu_A(x)>0}}{\forall} \mu_{A'}(x) = 0. \tag{6}$$

On the other hand, $\tau_2 =$ "*absolutely true*" corresponds to full consistency of a fact and a premise: triangular description of the fact is included within the kernel of the premise

$$\underset{\substack{x\in X \\ \mu_A(x)<1}}{\forall} \mu_{A'}(x) = 0. \tag{7}$$

The results of composition are in this case obvious. "*Absolutely false*" as the result of "*and*" operation and "*absolutely true*" as the result of "*or*" junction.

Similar results are obtained for the second example, where the difference between truth functions is big enough that either one or the second function is copied.

The contribution of both truth functions in the results can be observed for the last two cases, where nodes of τ_1 and τ_2 are used in the output. Generally, the algorithm used to obtain the result can perform the appropriate junction operation (T-norm or S-norm) on domain values of relevant nodes in two truth functions. This produces the new domain coordinates of the output nodes. If the relevant nodes differ in values of counter-domain coordinates, an appropriate value is chosen respectively to the type of junction.

5. COMPLEXITY ANALYSIS

Considering simplified stages of obtaining two analyzed types of truth functions, the general time complexity of the approach can be described by the following equation

$$KC\log_2 N + (K-1)C + C^2 + NC, \tag{8}$$

where in comparison to (2) C_{minDY} was replaced by a constant C. As it was mentioned before, the computation does not depend on specified precision parameter anymore (C_{minDY}) but on a constant (C) involved with the analysis of only several important nodes. Therefore, for big numbers of K the analysis becomes comparable to other widely used simplified approaches, which was the aim of this work.

6. CONCLUSION

The paper presented the analysis of possible simplifications in the fuzzy reasoning of Baldwin. According to the previous work ([16]), the problem was focused on two important stages: obtaining a truth function of a premise and a compound truth function.

Suggested modifications described in previous sections strongly simplify the calculations needed for the full approach of Baldwin implemented within the Fuzzlib library. Now, obtaining truth functions of a premise and compound truth functions does not depend on precision parameter and involves processing of only several nodes.

Because of uncomplicated and efficient computations the solution allows to easily apply any T-norm and S-norm as the junction operation. Moreover, the advantage of extended fuzzy relation between facts and premises is preserved in the reasoning process. Described benefits makes the simplified reasoning based on a fuzzy truth value a very flexible tool in research and applications.

BIBLIOGRAPHY

[1] BALDWIN J. F., A new approach to approximate reasoning using a fuzzy logic. Fuzzy Sets and Systems, 1979, No. 2, pp. 309-325.

[2] BALDWIN J. F., GUILD N. C. F., Feasible algorithms for approximate reasoning using fuzzy logic. Fuzzy Sets and Systems, 1980, No. 2, pp. 225–251.

[3] BELLMAN R. E., ZADEH L. A., Modern Uses of Multiple-Valued Logic, chapter Local and Fuzzy Logics, 1977, pp. 103-165.

[4] BRINTHA S. J., BHUVANESWARI V., Clustering microarray gene expression data using type 2 fuzzy logic. 3rd National Conference on Emerging Trends and Applications in Computer Science (NCETACS), 2012, pp. 147-151.

[5] CHUANG C. L., CHEN C. M., SHIEH G. S., JIANG J. A., A neuro-fuzzy inference system to infer gene-gene interactions based on recognition of microarray gene expression patterns. IEEE Congress on Evolutionary Computation (CEC 2007), 2007, pp. 904-910.

[6] CHIU S. H., TSAI M. H., WU H. C., CHEN W. C., SHRESTHA U., KUO S., Application of Fuzzy c-Means and Self-organizing maps for genes clustering in mouse brain microarray data analysis. 9th Int.Conf. Fuzzy Systems and Knowledge Discovery (FSKD), 2012, pp. 1104-1108.

[7] CZOGAŁA E., ŁESKI J., Fuzzy and Neuro-Fuzzy Intelligent Systems. Physica-Verlag, Springer-Verlag Comp., Heidelberg, 2000.

[8] DUBOIS D., PRADE H., Fuzzy sets in approximate reasoning, part 1: Inference with possibility distribution. Fuzzy Sets and Systems, 1991, No. 40, pp. 143–202.

[9] DUBOIS D., LANG J., PRADE H., Fuzzy sets in approximate reasoning, part 2: Logical approaches. Fuzzy Sets and Systems, 1991, No. 40, pp. 203–244.

[10] FUTSCHIK M. E., KASABOV N. K., Fuzzy clustering of gene expression data. Proc. 2002 IEEE Int. Conf. Fuzzy Systems (FUZZ-IEEE'02), 2002, Vol. 1, pp. 414-419.

[11] HAVENS T. C., KELLER J. M., POPESCU M., BEZDEK J. C., MACNEAL REHRIG E., APPEL H. M., SCHULTZ J. C., Fuzzy cluster analysis of bioinformatics data composed of microarray expression data and gene ontology annotations. Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS 2008), 2008, pp. 1-6.

[12] HE Y., TANG Y., ZHANG Y. Q., SUNDERRAMAN R., Mining fuzzy association rules from microarray gene expression data for leukemia classification. IEEE Int. Conf. Granular Computing, 2006, pp. 461-464.

[13] HO S. Y., HSIEH C. H., CHEN H. M., HUANG H. L., Interpretable gene expression classifier with an accurate and compact fuzzy rule base for microarray data analysis. Biosystems, 2006, No. 85, pp. 165-176.

[14] KUDŁACIK P., Advantages of an Approximate Reasoning Based on a Fuzzy Truth Value. Journal of Medical Informatics & Technologies, 2010, Vol. 16, pp. 125-132.

[15] KUDŁACIK P., Operations on fuzzy sets with piece-wise linear membership function (in Polish), Studia Informatica, 2008, 29(3A):91–111.

[16] KUDŁACIK P., Performance Evaluation of Baldwin's Fuzzy Reasoning for Large Knowledge Bases. Journal of Medical Informatics & Technologies, 2012, Vol. 20, pp. 29-38.

[17] KUDŁACIK P., Structure of knowledge base in FUZZLIB library (in Polish). Proc. 4th Conf. Databases: Application and Systems'10. Ustron 25-28 May 2010, Studia Informatica, 31(2A):469–478.

[18] LARSEN P. M., Industrial Applications of fuzzy logic control. International Journal of Man-Machine Studies, 1980, Vol. 1, No. 12, pp. 3-10.

[19] MA P. C. H., CHAN K. C. C., Incremental Fuzzy Mining of Gene Expression Data for Gene Function Prediction. IEEE Trans. Biomedical Engineering, 2011, Vol. 58, Issue 5, pp. 1246-1252.

[20] MAJI P., PAUL S., Rough-Fuzzy Clustering for Grouping Functionally Similar Genes from Microarray Data. IEEE/ACM Trans on Computational Biology and Bioinformatics, 2012, Vol. PP , Issue 99, pp. 1.

[21] MAMDANI E. H., ASSILAN S., An experiment in linguistic synthesis with a fuzzy logic controller, International Journal of Man-Machine Studies, 1975, 20(2):1–13.

[22] MARGHNY M. H., EL-SEMMAN I. E., Extracting fuzzy classification rules with gene expression programming. Proc. Int. Conf. Artificial Intelligence and Machine Learning (AIML), 2005, Cairo, Egypt.

[23] MOHAMMADI A., SARAEE M. H., Estimating Missing Value in Microarray Data Using Fuzzy Clustering and Gene Ontology. IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM '08), 2008, pp. 382-385.

[24] PAUL A., SIL J., Sample selection of microarray data using rough-fuzzy based approach. World Congress on Nature & Biologically Inspired Computing (NaBIC 2009), 2009, pp. 379-384.

[25] PEREZ M., RUBIN D. M., SCOTT L. E., MARWALA T., STEVENS W., A Hybrid Fuzzy-SVM classifier, applied to gene expression profiling for automated leukaemia diagnosis. Proc. of IEEE 25th Convention of Electrical and Electronics Engineers, Israel, 2008.

[26] SCHAEFER G., NAKASHIMA T., YOKOTA Y., ISHIBUCHI H., Fuzzy Classification of Gene Expression Data. Proc. IEEE Int. Conf. Fuzzy Systems, 2007, pp. 1-6.

[27] SUGENO M., KANG G. T., Structure identification of fuzzy model, Fuzzy sets and systems, 1988, Vol. 28, pp. 15-33.

[28] TAKAGI H., SUGENO M., Fuzzy identification of systems and it's application to modeling and control, IEEE Trans. Systems, Man and Cybernetics, Vol. 15(1), 1985, pp. 16-132.

[29] TSUKAMOTO Y., Advances in Fuzzy Set Theory and Applications, chapter: An Approach to Fuzzy Reasoning Method, North-Holland, Amsterdam, 1979, pp. 137–149.

[30] VERMA N. K., GUPTA P., AGRAWAL P., YAN CUI, Fuzzy rule based unsupervised approach for salient gene extraction. IEEE Applied Imagery Pattern Recognition Workshop (AIPRW), 2009, pp. 1-5.

[31] WANG Y. P., GUNAMPALLY M., CHEN J, BITTEL D., BUTLER M. G., CAI W. W., Accurate Quantification of Gene Expression using Fuzzy Clustering Approaches. IEEE Int. Workshop on Genomic Signal Processing and Statistics (GENSIPS 2007), 2007, pp. 1-4.

[32] WANG Z., PALADE V., A Comprehensive Fuzzy-Based Framework for Cancer Microarray Data Gene Expression Analysis. Proc. 7th IEEE Int. Conf. Bioinformatics and Bioengineering (BIBE 2007), 2007, pp. 1003-1010.

[33] WU G. P. K., CHAN K. C. C., WONG A. K. C., BIN WU, Unsupervised discovery of fuzzy patterns in gene expression data. Proc. IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM), 2010, pp. 269-273.

[34] ZADEH L. A., Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man and Cybernetics, 1973, Vol. 1, No. 3, pp. 28-44.

[35] ZADEH L. A., The concept of a linguistic variable and it's application to approximate reasoning. parts 1-3. Information Science, 1975, Vol. 8-9, No. 8, pp. 43-80, pp. 199-249, 301-357.

[36] ZADEH L. A., Fuzzy logic and approximate reasoning. Syntheses, 1975, No. 30, pp. 407-428.