# A Soft Computing Based Approach For Anomaly Detection In IDS

Priyanka Dutta, Bijaya Kumar Panda, Manoranjan Pradhan

**Abstract**— Protecting computer networks from internal and external threats has become a major issue. Intrusion detection system (IDS) collects information from a computer or network of computers and attempts to detect intruders or system abuse. Soft computing techniques are used in IDS based on anomaly detection methods and also in rule-based expert systems where the knowledge is usually in the form of if-then rules. Network behaviors can be categorized into normal and abnormal. We need to extract the most important data that can be used to efficiently detect network attacks.This research work attempts to use soft computing techniques for Intrusion Detection System. The aim is to reduce the False Positive rate and False Negative rate.

**Index Terms**—Anomaly Based Detection, False Positive, False Negative, Fuzzy Decesion Module, Fuzzy Inference Engine, Fuzzy Logic, Fuzzy Rule Based System,Intrusion Detection System, KDD CUP'99 Dataset.

## 1 INTRODUCTION

NOWADAYS computer security is becoming a major issue due to the higher use of computers and tremendous growth of computer networks[1]. Intrusion is a successful attack or a successful unauthorized access in a computer or in a network. So in order to detect the intrusion from the system or the network different soft computing based methods are used to develop a tool called Intrusion Detection System (IDS). But nowadays many organizations are preferred to use Intrusion Detection and Prevention System (IDPS) because of its primary objective is to identify the problems with its security issues, documenting the existing threats. It typically collects the information about the observed events. Mainly the Intrusion Detection is classified into two types: Misuse Detection and Anomaly Detection [2]. Misuse/Signature Based Detection is mostly followed some fixed patterns. So it is very effective at detecting the known attacks but very poor in detecting unknown threats. Whereas Anomaly Detection, is very effective in detecting the previously unknown attacks. Anomaly detection methods are designed to face the problem when the intruders try to mask their illegal behavior to deactivate the detection system. The anomaly detection method always tries to find the normal behavior pattern with the assumption that an intrusion will generally include some deviation from this normal behavior [3]. Here our aim is to use both fuzzy logic and neural network to design, implement, and evaluate the Anomaly Based Intrusion Detection System. This thesis explains the method for integrating fuzzy logic with neural network to improve the flexibility of an Intrusion Detection System.

## 2 LITARATURE SURVEY

Several intrusion detection methods have been proposed for detecting the anomalies. Intrusion Detection Expert System (IDES) [4], one of earliest intrusion detection system which was developed at the Stanford Research Institute. The IDES always eyed on user behavior and detected the suspicious events to be occurred.

In [5], it is suggested that an intrusion detection method is used to detect the intrusion efficiently.

In [6], Denning considered that any changes or any differences in the normal behavior of user are treated as anomalous. For observing and detecting user's events, an Expert System of intrusion detection was developed by Stanford Research Centre. This centre also developed next generation mechanism which includes audit profiles of user's and can monitor the current status of the user, if any change occurs with user's activity in compared to audit profile of user then it will produce an alarm.

In [7], it is stated that intrusion detection systems have been generally built using expert system technology. But, Intrusion Detection System (IDS) researchers have been focused in building systems which are difficult to handle, inconvenient to use in real life and lack of insightful user interfaces. To find out attacks the proposed adaptive expert system has used fuzzy sets.

In [8], it has shown a method that detects real-time network anomaly attack for discovering suspicious activity against computer network by using Fuzzy-Bayesian. By combining fuzzy and Bayesian classifier, the overall performance of the intrusion detection system (IDS) based on Bayes has been improved.

In [9], it is briefly explained about an advanced fuzzy and data mining methods based on hybrid model to find out both misuse and anomaly attacks. Their primary objective was to decrease the quantity of data processing and also to improve the detection rate of the existing IDS using attribute selection process and data mining technique respectively. An improved fuzzy data mining algorithm is used for implementing fuzzy rules which enabled the generation of if-then rules that show common ways of expressing security attacks. They have achieved faster decision making using Mamdani inference mechanism with three variable inputs in the fuzzy inference engine which they have employed.

In [10], back propagation model for intrusion detection is briefly described. This method makes training pair with a combination of input and equivalent target were generated and implemented into the network. Performance success can be measured by false alarm and detection rate. Detection rate was proven to be less than 80% for U2R, R2L, DoS and Probe attacks. However, the major issue of the method was found to be much inefficient to detect hidden attackers present in the system.

## 3 DATASETS TO BE USED

With the wide use of computer networks, the number of attacks has grown tremendously, and different types of new hacking tools and intrusive methods have discovered. The main aim of our research is to develop an anomaly-based network intrusion detection system using soft computing techniques. We have decided to use KDD Cup 1999 Dataset that contains 41 features labeled as either normal or attack [11]see Table.1 in this research work.

In 1998, DARPA in concert with Lincoln Laboratory at MIT launched the DARPA 1998 dataset for evaluating IDS[12] .The DARPA 1998 dataset contains seven weeks of training and also two weeks of testing data. In total, there are 38 attacks in training data as well as in testing data. The tcpdump data provided by 1998 DARPA Intrusion Detection Evaluation network was further processed and used for the 1999 KDD Cup contest at the fifth International Conference on Knowledge Discovery and Data Mining[13]. The KDD Cup is an annual Knowledge Discovery and Data Mining competition organized by the ACM Special Interest Group on Knowledge Discovery and Data Mining.

The input KDD Cup 1999 dataset is divided into two subsets such as, training dataset and testing dataset. At first, the training dataset is classified into five subsets so that, four types of attacks (DoS (Denial of Service), R2L (Remote to Local), U2R (User to Root), Probe) and normal data are separated. Then the consequent part of fuzzy if-then rule is formed by matching the randomly generated fuzzy rules with each and every obtained association rule. Thus, we obtain a set of fuzzy if-then rules with consequent parts that represent whether it is a normal data or an abnormal data. In the testing phase, the test data is matched with fuzzy rules to detect whether the test data is an attack data or a normal data.

Various types of attacks incorporated in the dataset which are broadly classified into following four major categories:

**Denial of Service Attacks:** It is an attack where the attacker builds some memory resources which are unable to manage the legitimate requirements, or reject legitimate user's right to use a machine.

**User to Root Attacks:** It is an attack where the attacker initiates by accessing a normal user account on the system and take advantage of some susceptibility to achieve root access to the system.

**Remote to User Attacks:** This attack takes place when an attacker who has the capability to send packets to a machine over a network but does not have an account on that machine, makes use of some vulnerability to achieve local access as a user of that machine.

**Probes:** Probing is a type of attacks where an attacker tests a network to collect information or discover well-known vulnerabilities.

Different types of attacks in each category are summarised in table.2.

| Category 1 | Category 2 | Category 3 | Category 4 |
|---|---|---|---|
| F1:duration<br>F2:protocol-type<br>F3:service<br>F4:flag<br>F5:src-bytes<br>F6:dst-bytes<br>F7:land<br>F8:wrong-fragment<br>F9:urgent | F10:hot<br>F11:num-failed-logins<br>F12:logged-in<br>F13:num-compromised<br>F14:root-shell<br>F15:su-attempted<br>F16:num-root<br>F17:num-file-creations<br>F18:num-shells<br>F19:num-access-files<br>F20:num-outbound-cmds<br>F21:is-host-login<br>F22:is-guest-login | F23:count<br>F24:srv-count<br>F25:serror-rate<br>F26:srv-serror-rate<br>F27:rerror-rate<br>F28:srv-rerror-rate<br>F29:same-srv-rate<br>F30:diff-srv-rate<br>F31:srv-diff-host-rate | F32: dst-host-count<br>F33:dst-host-srv-count<br>F34:dst-host-same-srvrate<br>F35:dst-host-diff-srvrate<br>F36:dst-host-same-srcport-rate<br>F37:dst-host-srv-diffhost- rate<br>F38:dst-host-serror-rate<br>F39:dst-host-srv-serrorrate<br>F40:dst-host-rerror-rate<br>F41:dst-host-srv-rerrorrate |

Table1. List of features given in KDD cup 99 dataset

| Sl. No. | Category | Attack Types |
|---|---|---|
| 1 | Denial of Service Attacks | Back, land, neptune, pod, smurf, teardrop |
| 2 | User to Root Attacks | Buffer_overflow, loadmodule, perl, rootkit, |
| 3 | Remote to Local Attacks | Ftp_write, guess_passwd, imap, multihop, phf, spy, warezclient, warezmaster |
| 4 | Probes | Satan, ipsweep, nmap, portsweep |

Table 2. Different types of attacks described in four major categories:

## 4 PROPOSED MODEL

This research work proposed a soft computing based anomaly detection(SCBAD) Model. The different components of this model are shown in the following figure-1. Basically the proposed system for Anomaly Detection is described in 4 major categories.

a. Training data Classification
b. Generation of Fuzzy Rules
c. Fuzzy Decision Module
d. Classification for a test input

***a) Training data classification*:** The dataset we have taken for analyzing the intrusion detection behavior is KDD-Cup 1999 data. The KDD-Cup 1999 data contains four types of attacks and normal behavior data with 41 attributes that have both continuous and symbolic attributes. The proposed system is designed only for the continuous attributes because the major attributes in KDD-Cup 1999data are continuous in nature. Therefore, we have taken only the continuous attributes for instance, 34 attributes from the input dataset by removing discrete attributes. Then the dataset is divided into five subsets of classes based on the class label prescribed in the dataset. The class label describes several attacks, which comes under four major attacks (Denial of Service, Remote to Local, U2R and probe) along with normal data. The five subsets of data are then used for generating a better set of fuzzy rules automatically so that the fuzzy system can learn the rules effectively.

***b) Generation of fuzzy rules*:** Generally, the fuzzy rules given to the fuzzy system is done by manually or by experts, who are given the rules by analyzing intrusion behavior. But, in our case, it is very difficult to generate fuzzy rules manually due to large input data and having more attributes. So we use FIS editor for generation of fuzzy rules..

*Identification of suitable attributes for rule generation:* In this phase, we have chosen only the most suitable attributes for identifying the classification whether the record is normal or attack. The reason behind this step is that the input data contains 34 attributes, in which all the attributes are not so effective in detecting the intrusion detection. So to identify the suitable attribute, we have used deviation method. And by using this method we choose 14 attributes for experimentation which are mentioned in Table.5.

*Rule generation:* The effective attributes chosen from the previous step is utilized to generate rules that is derived from the {max, min} deviation. By comparing the deviation range of effective attributes in between the normal and attack data, the intersection points are identified for the effective attributes. By making use of these two intersection points, the definite and indefinite rules are generated.

***c) Fuzzy Decision module:*** Here we describe the designing of fuzzy inference system for finding the suitable test dataset. Fuzzy inference system is the process of formulating the mapping from a given input to an output using fuzzy logic. Zadeh in the late 1960s [14,15] introduced fuzzy logic and is known as the rediscovery of multi valued logic designed by Lukasiewicz. Here from thirty four input, we select 14 attributes and so these are used as input and single output of Fuzzy inference System (FIS) with area of defuzzification Strategy was used for this purpose.

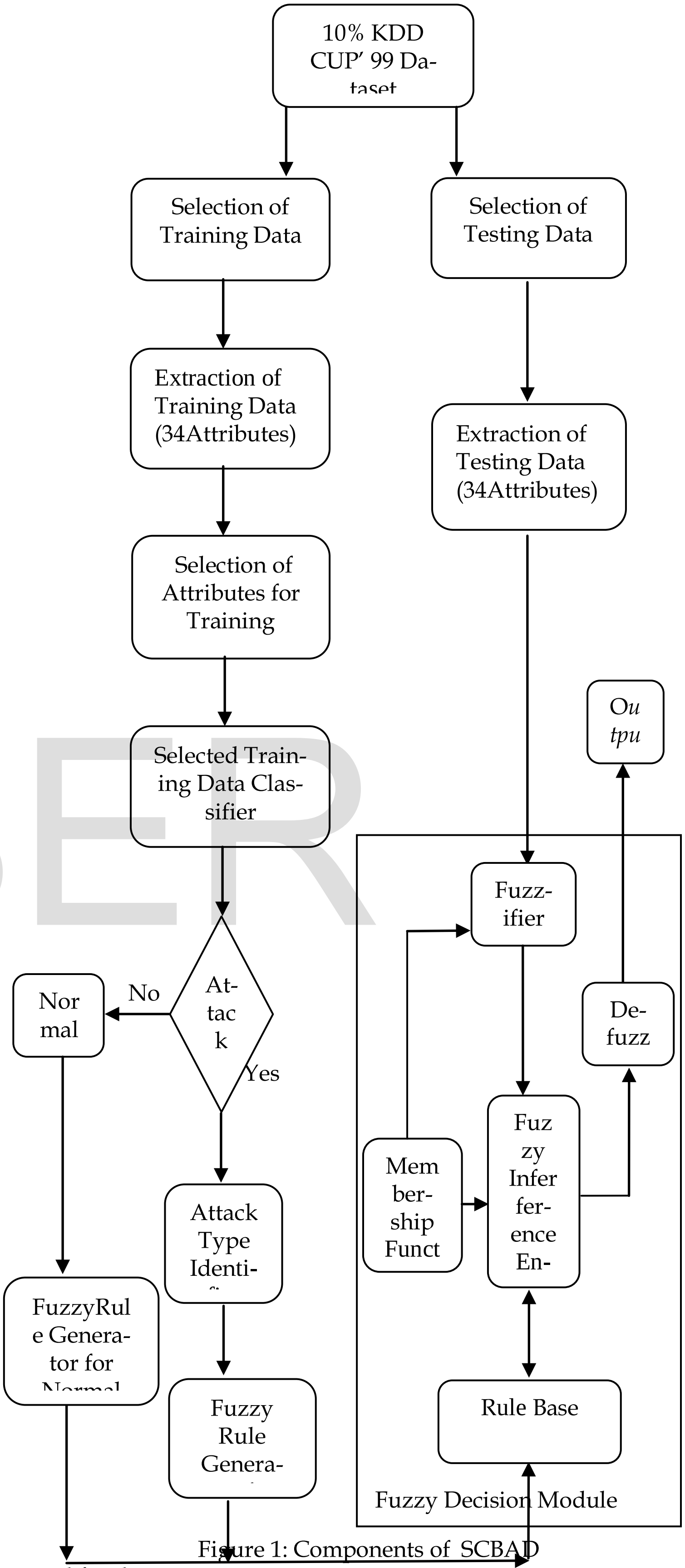
10% KDD CUP' 99 Dataset
Selection of Training Data
Selection of Testing Data
Extraction of Training Data (34Attributes)
Extraction of Testing Data (34Attributes)
Selection of Attributes for Training
Selected Training Data Classifier
Attack
No
Yes
Normal
Attack Type Identi-
FuzzyRule Generator for Normal
Fuzzy Rule Genera-
Outpu
Fuzzifier
Defuzz
Fuzzy Inference En-
Membership Funct
Rule Base
Fuzzy Decision Module

Figure 1: Components of SCBAD

***d) Classification for a Test Input:*** For testing phase, a test data from the KDD-cup 99 dataset is given to the designed fuzzy logic system for finding the fuzzy score. At first, the test input data containing 34 attributes is applied to fuzzifier, which converts 34 attributes (numerical variable) into linguistic variable using the triangular membership function. The output of the fuzzifier is fed to the fuzzy inference engine which in turn compares that particular input with the rule base. Rule base is a knowledge base which contains a set of rules obtained from the definite rules. The output of fuzzy inference engine is one of the linguistic values from the following set {Low and High} and then, it is converted by the defuzzifier as crisp values. The crisp value obtained from the fuzzy inference engine is varied in between 0 to 1, where '0' denotes that the data is completely normal and '1'specifies the completely attacked data.

## 5 EXPERIMENTATION AND RESULT ANALYSIS

In this section, the experimental results and performance evaluation of the proposed system are analyzed. We are using MATLAB(9.0) for implementation and the performance of the system is evaluated using Precision, Recall and F-measure. For experimental evaluation, we have taken KDD cup 99 dataset [13], which is mostly used for evaluating the performance of the intrusion detection system. For evaluating the performance, it is very difficult to execute the proposed system on the KDD cup 99 dataset since it is a large scale. Here, we have used 10% of KDD Cup 99 dataset for testing and training. The number of records taken for testing and training phase is given in table 3 and table 4.

| Training Dataset | |
|---|---|
| Normal | 24950 |
| DOS | 24950 |
| Probe | 4106 |
| R2L | 76 |
| U2R | 41 |

Table 3. Records taken forTraining Data

| Testing Dataset | |
|---|---|
| Normal | 25000 |
| DOS | 25000 |
| Probe | 4106 |
| R2L | 76 |
| U2R | 41 |

Table 4. Records taken for Testing Data

The training dataset contains normal data along with four types of attacks (DoS, U2R, R2L, Probe), which are given to the proposed system for identifying the suitable attributes. The selected attributes for rule generation process are given in table 5. After that using fuzzy rule learning strategy, the system generates definite and indefinite rules and finally, fuzzy rules are generated from the definite rules.

| Sl. No. | Attribute Index | Selected Attributes | Range |
|---|---|---|---|
| 1 | F1 | Duration | [0. 58329] |
| 2 | F5 | src_bytes | [0.1.3 one billion] |
| 3 | F6 | dst_bytes | [0.1.3 one billion] |
| 4 | F8 | wrong_fragment | [0.3] |
| 5 | F9 | Urgent | [0,14] |
| 6 | F10 | Hot | [0.101] |
| 7 | F11 | num_failed_logins | [0.5] |
| 8 | F13 | num_compromised | [0.9] |
| 9 | F16 | num_root | [0.7468] |
| 10 | F17 | num_file_creations | [0,100] |
| 11 | F18 | num_shells | [0,5] |
| 12 | F19 | num_access_files | [0.9] |
| 13 | F23 | Count | [0.511] |
| 14 | F24 | srv_count | [0.511] |

Table 5. Selected attributes for rule generation

In testing phase, the testing dataset is given to the proposed system, which classifies the input as a normal or attack. Then obtained result is used to compute overall accuracy of the proposed system. The overall accuracy of the proposed system is computed based on the definitions, namely Precision, Recall and F-measure which are normally used to estimate the rare class prediction. It is very effective to accomplish a high recall devoid of loss of precision. F-measure is a weighted harmonic mean which evaluates the trade-off between them.

$$\text{Precision} = T.P / (T.P+F.P)$$

$$\text{Recall} = TP / (T.P+F.N)$$

$$\text{F-measure} = [(\alpha^2+1)\ (\text{Precision. Recall})] \,/\, [(\alpha^2.\text{Precision}+\text{Recall})] \quad \text{Where } \alpha = 1$$

$$\text{Overall Accuracy} = (T.P+T.N) / (T.P+T.N+F.N+F.P)$$

Where, T.P: - True positive
T.N:-True negative
F.P: - False Positive
F.N: - False negative

These are computed using the confusion matrix in Table 6, and defined as follows:

| | **Pridicted Class** | | |
|---|---|---|---|
| **Actual Class** | | **Positive Class** | **Negative Class** |
| | **Positive Class** | True Positive(T.P) | False Negative(F.N) |
| | **Negative Class** | False Positive(F.P) | True Negative(T.N) |

Table 6. Confusion matrix

The evaluation metric is computed for both training and testing dataset in the testing phase and the obtained result for all attacks and normal data are given in table 7, which is the overall classification performance of the proposed system on KDD cup 99 dataset. By analyzing the result, the overall performance of the proposed system is improved significantly and it achieves more than 90% accuracy for all types of attacks.

| Metric | | Proposed System | |
|---|---|---|---|
| | | Training | Testing |
| PROBE | Precision | 0.912532 | 0.912532 |
| | Recall | 0.37085 | 0.37085 |
| | F-measure | 0.52735557 | 0.52735557 |
| | Accuracy | 0.906218 | 0.909333 |
| DOS | Precision | 0.993553 | 0.993818 |
| | Recall | 0.90145 | 0.904156 |
| | F-measure | 0.94526246 | 0.94687246 |
| | Accuracy | 0.9488 | 0.949268 |
| U2R | Precision | 0.051958 | 0.051958 |
| | Recall | 0.180476 | 0.180476 |
| | F-measure | 0.08263265 | 0.08263265 |
| | Accuracy | 0.992112 | 0.993188 |
| R2L | Precision | 0.075959 | 0.075959 |
| | Recall | 0.155744 | 0.155744 |
| | F-measure | 0.10212866 | 0.10212866 |
| | Accuracy | 0.992586 | 0.992909 |
| NORMAL | Precision | 0.828479 | 0.829328 |
| | Recall | 0.99426 | 0.994375 |
| | F-measure | 0.90376539 | 0.90438029 |
| | Accuracy | 0.910872 | 0.903119 |

Table 7. The classification performance of the proposed IDS

## 6 Conclusion

This research work successfully demonstrated the use of soft computing techniques in intrusion detection. In this paper, a soft computing based model was proposed for detecting the intrusion using anomaly detection method. A fuzzy decision module was designed to build the system more accurate for attack detection, using the fuzzy inference approach. An effective set of fuzzy rules were identified, which are more effective for detecting intrusion in a computer network. Firstly, the definite rules were generated. Then, fuzzy rules were identified by fuzzifying the definite rules and these rules were given to fuzzy system, which classify the testing data. We have used KDD cup 99 dataset for evaluating the performance of the proposed system and experimentation results showed that the proposed method is effective in detecting various intrusions in computer networks.

## References

[1] Yao, J. T., S.L. Zhao, and L.V. Saxton, "A Study On Fuzzy Intrusion Detection", In Proceedings of the Data Mining, Intrusion Detection, Information Assurance, And Data Networks Security, SPIE, Vol. 5812, pp. 23-30,Orlando,Florida, USA, 2005.

[2] S. Zhong, T. Khoshgoftaar, N. Seliya, "Clustering-based network intrusion detection", In Intl. Journal of Reliability, Quality and Safety, Vol. 14, pp. 169-187, No. 2, 2007

[3] Pakkurthi Srinivasu, P.S. Avadhani, Vishal Korimilli, Prudhvi Ravipati, "Approaches and Data Processing Techniques for Intrusion Detection Systems", Vol. 9, No. 12, pp. 181-186, 2009.

[4] T.F. Lunt, A. Tamaru, F. Gilham, R. Jagannathm, C. Jalali,P.G.Neumann, H.S. Javitz, A. Valdes, T.D. Garvey, "A Real-time Intrusion Detection Expert System (IDES)," Computer Science Laboratory, SRI International, Menlo Park, CA, USA, Final Technical Report, February1992.

[5] Shah, K., Dave, N., Chavan, S., Mukherjee, S., Abraham, A. and Sanyal S. Adaptive neuro-fuzzy intrusion detection system. IEEE International Conference on Information Technology: Coding and Computing (ITCC'04), vol. 1. USA: IEEE Computer Society;2004; 70–74.

[6] Anderson, D., Frivold, T. and Valdes, A. Next-generation intrusion detection expert system (NIDES): A summary Technical Report SRI–CSL–95–07,Computer Science Laboratory,SRI International, May 1995.

[7] Stephen F. Owens, Reuven R. Levary, "An adaptive expert system approach for intrusion detection", International Journal of Security and Networks, Vol: 1, No: 3/4, pp: 206-217, 2006.

[8] O. Adetunmbi Adebayo, Zhiwei Shi, Zhongzhi Shi, Olumide S. Adewale, "Network Anomalous Intrusion Detection using Fuzzy-Bayes", IFIP International Federation for Information Processing, Vol: 228, pp: 525-530, 2007.

[9] Bharanidharan Shanmugam, Norbik Bashah Idris, "Improved Intrusion Detection System Using Fuzzy Logic for Detecting Anamoly and Misuse Type of Attacks", in Proceedings of the International Conference of Soft Computing and Pattern Recognition, pp: 212-217, 2009.

[10] Jaiganesh, V., Sumathi, P. and Mangayarkarasi, S.,"An Analysis of Intrusion Detection System using back propagation neural network", IEEE Computer Society Publication;2013.

[11] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu and Ali A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set", in Proceedings of the Second IEEE international conference on Computational intelligence for security and defense applications, pp. 53-58, Ottawa, Ontario, Canada, 2009.

[12] http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/1998data.html\

[13] http://www.sigkdd.org/kddcup/index.php?section=1999&method=data

[14] www.cse.unr.edu/~bebis/CS365/Papers/FuzzyLogic.pdf Available: http://www.aptronix.com/fide/ ... .edu/~fishwick/paper/paper.html.

[15] Zadeh, L.A., "Fuzzy sets", Information and control, vol.8, pp. 338-353, 1965.