A Network Intrusion Detection Framework based on Bayesian Network using Wrapper Approach

Md Reazul Kabir
Ahsanullah University of Science and Technology
Dhaka
Bangladesh

Abdur Rahman Onik
Ahsanullah University of Science and Technology
Dhaka
Bangladesh

Tanvir Samad
Ahsanullah University of Science and Technology
Dhaka
Bangladesh

ABSTRACT
Increasing internet usage and connectivity demands a network intrusion detection system combating cynical network attacks. Data mining therefore is a popular technique used by intrusion detection system to prevent the network attacks and classify the network events as either normal or attack. Our research study presents a wrapper approach for intrusion detection. In this framework Feature selection technique eliminate the irrelevant features to reduce the time complexity and build a better model to predict the result with a greater accuracy and Bayesian network works as a base classifier to predict the types of attack. Our experiment shows that the proposed framework exhibits a superior overall performance in terms of accuracy which is 98.2653 , error rate of 1.73 and keeps the false positive rate at a lower rate of 0.007. Our model performed better than other leading state-of-the-arts models such as KNN, Boosted DT, Hidden NB and Markov chain. The NSL-KDD is used as benchmark data set with Weka library functions in the experimental setup.

General Terms
Pattern Recognition. Intrusion detection system, Data Mining

Keywords
Intrusion Detection System, Feature Selection, Genetic Search, Bayesian Network, Weka, NSL-KDD.

1. INTRODUCTION
Computer security became vulnerable because of the massive expansion of the computer networks and rapid emergence of the hacking tools and intrusion incidents. As technology is rolling out this attacks make the network security more vulnerable and therefore intrusion detection system is introduced to eliminate these threats. Intrusion detection system is assigned to shield the system from malicious attacks and network vulnerability. Intrusion detection system is categorized in two different form a) anomaly detection b) misuse detection [1]. In anomaly detection, the system builds a profile of that which can be considered as normal or expected usage patterns over a period of time and triggers alarms for anything that deviates from this behavior (Adel, Zeynep & Adnan, 2014) [1]. The misuse detection is used to identify attacks in a form of signature or pattern [10].Misuse detection came with the drawback, incapable of identifying the unknown attacks.

In data mining, feature selection is an approach to reduce the dimensionality and remove the irrelevant and inappropriate features from the dataset. Feature selection technique represents a subset of features. This technique is very much familiar in the field of classification, machine learning, data mining, pattern recognition, intrusion detection, image processing and so forth. Feature selection builds a model with higher accuracy by eliminating irrelevant data that reduce the time complexity. Researchers have been working in the field of feature selection from early 1970 (Adel, Zeynep & Adnan, 2014), Levent Koc, Thomas, Shahram, 2012 categorized feature selection model in three different parts. i) Filter method ii) Wrapper method ii) Embedded method [7] [1].Several proposed architecture has been built with the idea of a feature selection technique. Yin Shui, Li 2012 proposed an improved feature selection approach named GFR method, selecting 19 features NSL KDD dataset out of 41 features [19]. Adel, Zeynep & Adnan, 2014 proposed a new feature selection approach based on the cuttlefish optimization algorithm. The framework uses the cuttlefish algorithm (CFA) as a search strategy to ascertain the optimal subset of features on KDD Cup 99 dataset. [1]. Ming-Yang Su, 2011 proposed genetic algorithm combined with k nearest neighbor for feature selection and weighting [18]. Initially, 35 features in the training step were weighted and based on their weight the top ones were selected to implement testing phase [18]. For known attacks, 19 features were considered and provide an accuracy of 97.42% on the contrary, for unknown attacks 28 features were considered and the accuracy rate was 78%. [18].

Classification is the identification of the instances label which is described by a set of features. Learning classifier models learn from the given training data and infer the class labels for the instances of the new data [9]. Researchers have applied numerous classifier models to intrusion detection problem. Rule based detection (Lunt, 1989) , Neural networks (Cannady , 1998; Lippman & Cunningham,2000; Zhang 2001; M. Bahrololum, E.S, 2009), Fuzzy logic(Bridges & Vaughn,2000), Hidden Naïve Bayes (Levent Koc, 2012), Data Mining (Lee, Stolfo, & Mok, 1999) and Bayesian analysis (Bar-bara, Wu & Jajodia,2001) [9] . M.Bahrololum used the neural network model with Ranker search method to achieve an accuracy of 97.0 % [8]. Yung-Tsung Hou (2010) proposed the Boosted_DT model using the ranker method in intrusion detection system [20]. This system gains an accuracy of 96.14%. Levente koc (2012) applied the Hidden Naïve Bayes model in intrusion detection system that suffers from dimensionality and the accuracy of his proposed model was 93.72% [7]. In 2011, Shun-Sheng Wang proposed Adaptive Resonance Theory with ranker search based on SVM in intrusion detection [16]. The proposed model gives an accuracy of 95.13% [16]. Adel, 2014 proposed a novel feature selection approach based on cuttlefish optimization algorithm for intrusion detection system with an accuracy of 91.986% [1]. Seongjun Shin, 2013 proposed advance probabilistic approach for intrusion detection using Markov chain with an accuracy of 90.0% [13].

2. INTRODUCTION TO GENETIC ALGORITHM AND BAYESIAN NETWORK

2.1 Genetic Algorithm in Feature Selection:

Genetic algorithm is an optimization algorithm based on natural selection of evolutionary algorithm which is described from Darwin's theory. The basic idea of Genetic algorithm consist chromosome also knows as genetic information which encodes the candidate solution (i.e. individuals) to optimize the problem set [3]. Binary strings like 0's or 1's are used to represent genetic information and sets of bits to encode the solution of the problem set [3]. Crossover and mutation are the two types of operators which are applied on the individuals for the next generation. Crossover operator copies the selected bit strings from parents and then creates two offspring strings [3]. Mutation randomly changes the string bits value [3]. Fitness function assures that the survival probability of a single bit increased throughout the evolutionary process [3]. Genetic algorithm is more efficient with large search space and has a little probability of reaching local optimal solution than other algorithms. Genetic algorithms are stochastic optimization procedure which works efficiently to choose the small subset of features for classification with a lower computational requirement. Often the fitness function of the genetic algorithms employed for feature selection task is manipulated to achieve parsimonious solutions concerning the size of the feature selection subset [3, 5, and 9].

2.2 Wrapper approach

Feature selection can be categorized in three different approaches I) Filter approach II) Wrapper approach III) Embedded approach [7]. Filter approach assesses the relevance of the features from the dataset and the selection of the features is based on the statistics. Wrapper approach is more appropriate because in filter approach the performance of a particular classifier in the selection of the features are ignored. In wrapper, the classifiers performance is used in feature selection as a part of the search for the evaluation of the feature subset. Wrapper approach takes the classifier performance to evaluate the features resulting in better prediction accuracy

2.3 Bayesian netwrok

Bayesian network is one of the most common approaches in data mining and pattern recognition. Bayesian network for classification is a graphical representation of the set of random variables in probabilistic relation. A probabilistic graphical model is called Bayesian network when the graph connecting its variable is a directed acyclic graph [11]. Bayesian network consists of two major components I) Directed acyclic graph (DAG) II) Probability distribution [11]. Stochastic variables are represented by the nodes in the directed acyclic graph and arcs represent directed dependencies among variables that are quantified by conditional probability distributions [11]. Bayesian network is based on the probability theory and support for missing data during learning and classification [11]. Fig 1 shows the structure of a Bayesian network. The dotted lines denoted potential links and the blue box are used to indicate that additional nodes and links can be added to the model in-between input and output [21].

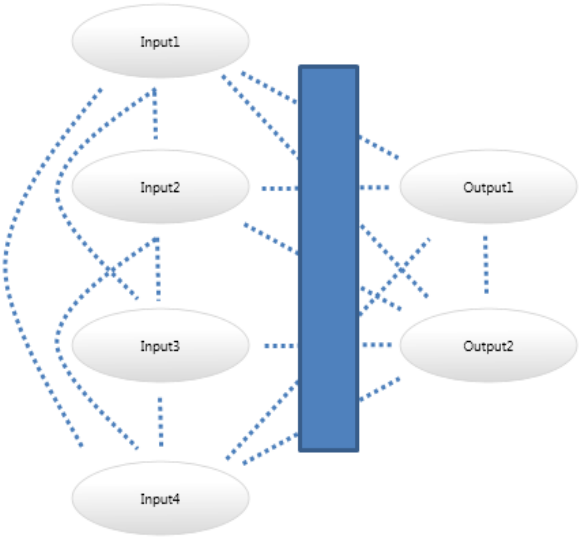

Fig 1: Structure of a Bayesian Network [21].

3. PROPOSED FRAMEWORK

The proposed framework operates in two different phases.

Phase I: Feature selection using a wrapper approach with Genetic Search algorithm

Phase II: Classification of Test instances using Bayesian Network.

Figure 2 gives a quick glance about the whole IDS system that has been proposed in this research paper in order to get better performance where the wrapper feature selection step belongs to phase I and just after that the classification model represents phase II. In phase I, where the redundant features are been cut off we have proposed a wrapper feature selection method in order to get better accuracy. In wrapper approach a search algorithm, genetic search is used to search through the space of possible features and evaluate each subset by running a Bayesian Network based model on the subset. Feature subset with better performance is been selected. In phase I, we have proposed to use a genetic search algorithm where the probability of crossover that means the probability that two Population members will exchange genetic material had been set to 0.6 as for the number of generations to evaluate had been set to 20. The probability of mutation occurring had been set to 0.0333 and as for the population size 20 had been taken.

In phase II, for building up a classification model a Bayesian Network has been developed with a simple estimator with the value of 0.5 for alpha where the estimator algorithm is been used to find the conditional probability tables of the Bayes Network. Again, K2 search algorithm is been used for developing the Bayesian Network based classifier for searching network structures. A test instance is been classified by the Bayesian Network based classification model.

Our proposed wrapper type Bayesian Network based attack detector algorithm is given below:

Algorithm of the proposed Wrapper type Bayesian Network Based Attack Detector (WBNAD)

Input: Training and Test Data set.

Output: Intrusion detection Model.

Step 1: Begin by randomly generating an initial population P;

Step 2: Calculate e(x) for each member $x \in$;

Step 3: Define a probability distribution p over the members of P where $(x) \propto (x)$;

Step 4: Select two population members x and y with respect to p;

Step 5: Apply crossover to x and y to produce new population members x' and y';

Step 6: Apply mutation to x' and y';

Step 7: Insert x' and y' into P' (the next generation);

Step 8: If |P'| < |P|, go to 4;

Step 9: Let, $P \leftarrow P'$;

Step 10: If there are more generations to process, goto 2;

Step 11: Return $x \in P$ for which e(x) is highest;

Step 12: Measure the performance of the selected feature subset using Bayesian Network;

Step 13: If, performance improved take the subset on account;

Step 14: If, Stopping Criterion reached, stop the feature extraction method;

Else, go to step **1**;

Step 15: Buildup a Bayesian Network based classifier using Training dataset and the final feature subset;

Step 16: Classify Test instances using Bayesian Network based classifier using final feature subset selected by the wrapper approach;

4. EXPERIMENTAL RESULT

4.1 Experimental result:

The experiment run on an Intel® Core™ i5-3210M CPU @ 2.59Ghz(4 CPUs),~2.5GHz with 4.09 G memory running on Windows 8.1. The experiment conducted with the help of JAVA programming language, WEKA 3.6 machine learning tool and Weka Library functions for feature selection techniques. In our experiment we used NSL-KDD benchmark dataset, which is the most popular data set developed by the MIT Lincoln Lab to compare the performance of different intrusion detection approach. NSL-KDD training dataset consists of approximately 4,900,000 single connection vectors each of which contains 41 features and is labeled as either normal or an attack, with exactly one specific attack type. The classes in NSL-KDD Dataset can be classified in five types such as normal and four types of attack names: DoS, Probing, U2R, and R2L.

4.2 Performance measurement:

True positive rate:

$$True\ positive\ rate = \frac{TP}{P}$$

False positive rate:

$$False\ Positive\ rate = \frac{FP}{N}$$

Accuracy and Precision:

$$Precision = \frac{TP}{TP + FP}$$

$$Accuracy = \frac{TP + TN}{P + N}$$

ROC Curve:
The curve is created by plotting the true positive rate against the false positive rate at various threshold settings. A classification model with ROC area of 95% or above is a good one.

4.3 Experimental evaluation:

The experiment conducted on a part of renowned benchmark dataset NSL-KDD where the whole dataset contained 25,192 instances with 41 features and beside the normal type of class label it had 4 more attack type of class known as- DoS, Probing, U2R, and R2L. All the experiments had been conducted by using 10-fold cross validation as k-fold cross validation is the well-known process for conducting any classification system as it eliminates the possibility of building up an over-fitted classification system. In table 1 showed the overall performance of our proposed IDS system is been showed in a glance where the proposed system showed a higher true positive rate of 98.3% along with a very low false positive rate 0.7%. Again our proposed system showed a very high ROC area of 99.9% which is clearly a sign of an outstanding performance.

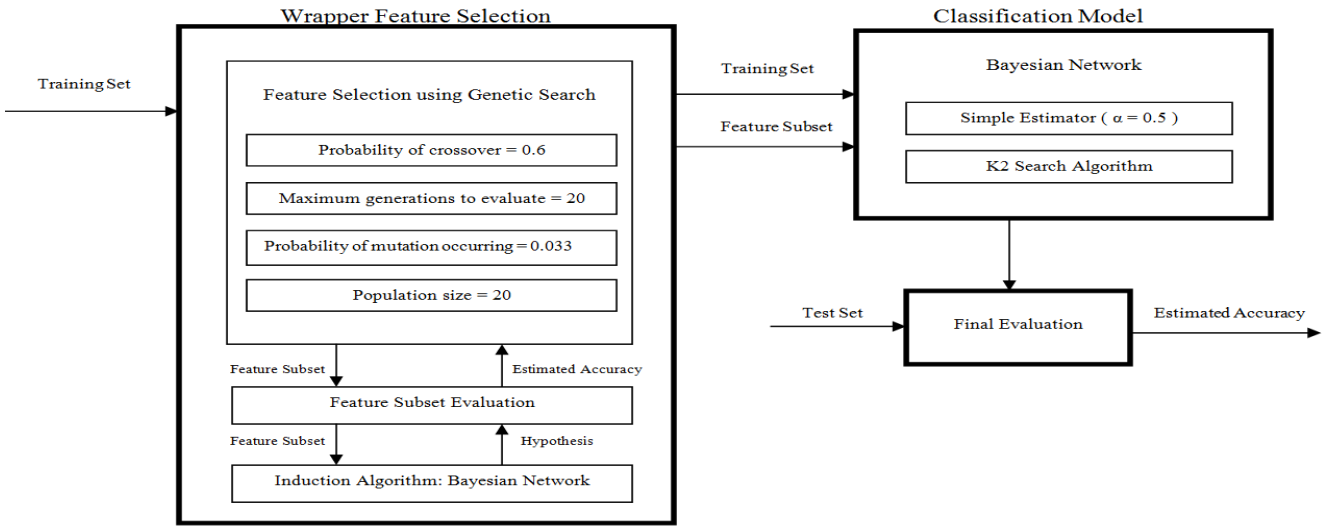

Fig 2: Proposed framework of Bayesian algorithm using wrapper approach

Class	**True Positive rate (TP)**	**False Positive rate (FP)**	**ROC area**
Normal	98.5	0.7	99.9
DoS	98.9	0.7	99.9
R2L	87.1	0.2	99.4
Probing	95.4	0.5	99.6
U2R	75.5	0.3	95.8
Weighted Average	98.3	0.7	99.9

Table 1: Overall Performance of the Proposed IDS

Table 2 showed the comparison of our proposed algorithm with some renowned algorithms where our proposed system again stands as a better performer with an accuracy rate of 98.26% where the other algorithms like Naive Bayes showed an accuracy rate of 84.86%, a decision tree type algorithm C4.5 showed 97.44% and another popular algorithm SMO showed an accuracy of 97.99%. Again the time taken by our proposed algorithm in training phase is very low with respect to other popular algorithms, such as the proposed system took only 0.23 seconds where SMO took 15.01 seconds in training phase of the classification model.

Algorithms	Naive Bayes	C4.5	SMO	Proposed Algorithm
Accuracy (%)	84.86	97.44	97.99	98.26
Precision (%)	92.3	97.8	98.0	98.9
Time taken in training phase (sec)	0.26	1.93	15.01	0.23

Table 2: Performance Comparison between Several Algorithms

Table 3 showed the comparison of the accuracy rate of our proposed system for attack detection with some recently conducted other existing research works where our proposed system stands as a better performer than all of them.

Table 4 showed the performance comparison of our proposed wrapper approach with some other popular feature selection techniques. Our proposed wrapper approach selected only 16 important features within 41 features and showed better performance than CFS, consistency type feature selection techniques where CFS type filter approach showed 95.58 % of accuracy rate and consistency type feature selection technique showed 94.93% of accuracy rate using rank search. Using genetic search CFS type filter showed an accuracy of 94.25% which is much lower than our proposed wrapper approach which also used genetic search for searching the feature space.

Research Paper	Accuracy (%)
Proposed wrapper type Bayesian Network based attack detector (WBNAD)	98.26
Su, M.-Y. (2011). Real-time anomaly detection systems for Denial-of-Service attacks by weighted k-nearest-neighbor classifiers. Expert Systems with Applications, ELSEVIER [18].	97.42
Yung-Tsung Hou, Y. C.-S.-M. (2010). Malicious web content detection by machine learning. Expert Systems with Applications, ELSEVIER [20].	96.14
Shun-Sheng Wang, K.-Q. Y.-C.-W. (2011). An Integrated Intrusion Detection System for Cluster-based Wireless Sensor Networks. Expert Systems with Applications, ELSEVIER [16].	95.13
Levent Koc, T. A. (2012). A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier. Expert Systems with Applications, ELSEVIER [7].	93.72
Adel Sabry Eesa, Z. O. (2014). A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems. Expert Systems with Applications, ELSEVIER [1].	91.986
Seongjun Shin, S. L. (2013). Advanced probabilistic approach for network intrusion forecasting and detection. Expert Systems with Applications, ELSEVIER [13].	90.0

Table 3: Performance comparison of the proposed model with some existing research work sorted by accuracy

Fig 3 depicts the graphical representation of accuracy of several feature selection algorithm derived in Table 4.

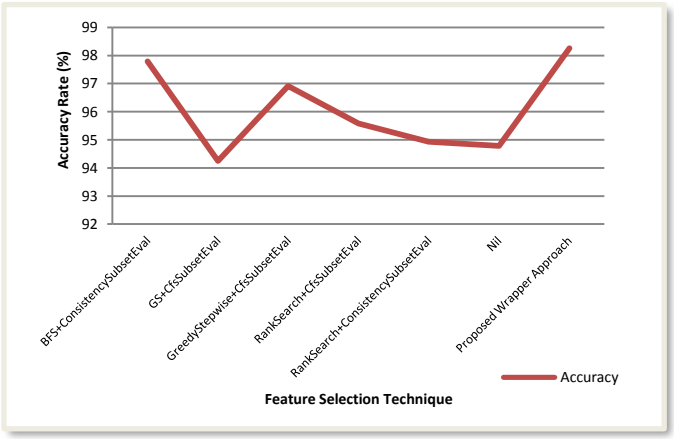

Figure 3: Graphical representation of the Accuracy using Several Feature Selection Techniques.

Algorithm	#Feature Selected	Accuracy (%)
BestFirst+ConsistencySubsetEval	9	97.79
GeneticSearch+CfsSubsetEval	23	94.25
GreedyStepwise+CfsSubsetEval	11	96.91
RankSearch+CfsSubsetEval	14	95.58
RankSearch+ConsistencySubsetEval	26	94.93
Nil	41	94.78
Proposed Wrapper Approach	16	98.26

Table 4: Performance Comparison between Several Feature Selection Techniques

5. CONCLUSION AND FUTURE WORK

The proposed framework detects attack using Bayesian Network classifier with wrapper approach for feature selection with the following procedures: construct a proper NSL-KDD train dataset, reduce the features into 16 from 41 by the wrapper approach; classification of test instances using Bayesian Network classifier. The false positive rate (FP) of the proposed model is 0.007 with a true positive rate (TP) of 98.26%.The result shows that the proposed framework is a reliable one and outruns other classifiers performance in efficiency and accuracy. The wrapper method consists reasonable features and shows an advantage in the performance in intrusion detection.

This experimental study showed us a way to observe the intrusion detection by using lesser features which led towards the time and complexity reduction in training and testing phase. Our future work can be conducted in the following perspectives:

- Developing simpler feature selection approach by experimenting with other feature selection techniques.
- Implement this proposed approach with real cloud data for analyzing real-time experience effects.
- Developing our proposed methodology for designing an adaptive intrusion detection framework

6. REFERENCES

[1] Adel Sabry Eesa , Zeynep and Brifcani (2015). A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems in Expert Systems with Applications. Volume 42, Issue 5, Pages 2670-2679.

[2] A.M. Chandrashekhar, K. (2013). Fortification of hybrid intrusion detection system using variants of neural networks & support vector machines. International Journal of Network Security & Its Applications (IJNSA) .

[3] Chih-Fong Tsai, William Eberle , Chi-Yuan Chu. (2013), Genetic algorithms in feature and instance selection. Expert Systems with Applications, ELSEVIER

[4] C.M.Bishop. (1995). Neural networks for pattern recognition. England: Oxford University.

[5] Carlos A. Catania, F. B. (2012). An autonomous labeling approach to support vector machines algorithms for network traffic anomaly detection. Expert Systems with Applications, ELSEVIER .

[6] Chengpo Mua, Y. L. (2010). An intrusion response decision making model based on hierarchical. Expert Systems with Applications, ELSEVIER .

[7] Chih-Fong Tsai, Y.-F. H.-Y.-Y. (2009). Intrusion detection by machine learning: A review. expert systems with applications, ELSEVIER .

[8] D. Sa´nchez, M. V. (2009). Association rules applied to credit card fraud detection. Expert Systems with Applications,ELSEVIER .

[9] Dahlia Asyiqin Ahmad Zainaddin, Z. M. (2013). HYBRID OF FUZZY CLUSTERING NEURAL NETWORK OVER NSL DATASET FOR INTRUSION DETECTION SYSTEM. Journal of Computer Science .

[10] Dewan Md. Farid, L. Z. (2013). An Adaptive Ensemble Classifier for Mining Concept-Drifting Data Streams. Expert systems with Applications,ELSEVIER .

[11] Dewan Md. Farid, M. Z. (2011). Adaptive Intrusion Detection based on Boosting and. International Journal of Computer Applications

[12] Feng Jiang, Y. S. (2009). Some issues about outlier detection in rough set theory. expert systems with application,ELSEVIER .

[13]] G. Davanzo, E. M. (2011). Anomaly detection techniques for a web defacement monitoring service. Expert Systems with Applications,ELSEVIER .

[14] Gisung Kim, S. L. (2013). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. Expert Systems with Applications,ELSEVIER

[15] Han-Ching Wu, S.-H. S. (2010). Neural networks-based detection of stepping-stone intrusion. Expert Systems with Applications,ELSEVIER .

[16] Haq NF, Onik AR, Shah FM. An ensemble framework of anomaly detection using hybridized feature selection approach (HFSA). InSAI Intelligent Systems Conference (IntelliSys), 2015 2015 Nov 10 (pp. 989-995). IEEE.

[17] Haq NF, Onik AR, Shah FM. "Application of Machine Learning Approaches in Intrusion Detection System: A Survey." (IJARAI) International Journal of Advanced Research in Artificial Intelligence.

[18] Onik AR, Haq NF, Alam L, Mamun TI. An Analytical Comparison on Filter Feature Extraction Method in Data Mining using J48 Classifier. International Journal of Computer Applications. 2015 Jan 1;124(13).

[19] Onik AR, Haq NF, Mustahin W. Cross-breed type Bayesian network based intrusion detection system (CBNIDS). InComputer and Information Technology (ICCIT), 2015 18th International Conference on 2015 Dec 21 (pp. 407-412). IEEE

[20] Yinhui Li, J. X. (2012). An efficient intrusion detection system based on support vector machines and gradually feature removal method. Expert Systems with Applications, ELSEVIER .

[21] Yung-Tsung Hou, Y. C.-S.-M. (2010). Malicious web content detection by machine learning. expert systems with applications, ELSEVIER .

[22] Yusuf Sahin, S. B. (2013). A cost-sensitive decision tree approach for fraud detection. Expert Systems with Applications, ELSEVIER.

[23] Zimmermann, H.-J. (2010). Fuzzy set theory. Advanced Reviw Zimmermann, H.-J. (2010). Fuzzy set theory.