Towards the Scalable Classifier with Rule Based Fuzzy Logic for Big Data Learning

Khin May Win
University of Computer Studies, Pathein
winn.km05@gmail.com

Abstract

There is no doubt that big data are now rapidly expanding in all science and engineering domains. While the potential of these massive data is undoubtedly significant, fully making sense of them requires new ways of thinking and novel learning techniques to address the various challenges. Big Data Fuzzy Supervised Learning has been the main focus of latest research efforts. Large-scale data sets are collected and studied in numerous domains, from engineering sciences to social networks, commerce, biomolecular research, and security. Particularly, digital data, generated from a variety of digital devices, some traditional learning methods suffer from a loss of information. First, this paper focuses on the analysis on the machine learning techniques and highlights some promising learning methods in recent studies and then discussions about the challenges and possible solutions. The proposed solution comes through integration of fuzzy rules and binary classification which provide problem solving for big data learning.

Keywords : big data, supervised learning, fuzzy, rule based, classification.

1. Introduction

Big Data Analytics has become feasible as well as recent powerful hardware, software, and algorithms developments; however, these algorithms still need to be fast and reliable [11]. The excessive amount of data have been being continually generated at unprecedented and ever increasing scales. To clarify what the big data refers to, several good surveys have been presented recently and each of them views the big data from different perspectives, including challenges and opportunities [2], background and research status, and analytics platforms [5].Among these surveys, comprehensive overview of the big data from three different angles, i.e., innovation, competition, and productivity. Over the past decade, machine learning techniques have been widely adopted in a number of massive and complex data-intensive fields such as medicine, astronomy, biology, and so on.

The enormous variety of data is the main dimension that makes big data both interesting and more challenging. This is resulted from the phenomenon that data generally come from various sources and are of different types. Structured, semi-structured, and even entirely unstructured data sources stimulate the generation of heterogeneous, high-dimensional, and nonlinear data with different representation forms. Computer science history in the field of machine learning has been shown significant development particularly in the area of Supervised Learning (SL) applications. Many supervised learning applications and semi-supervised learning algorithms were developed with Boolean logic rather than using Fuzzy logic. The machine learning algorithms were typically fed with relatively accurate data from well-known and quite limited sources, so the learning results tend to be unerring, too; As the time for big data is coming, the collection of data sets is so large and complex that it is difficult to deal with using traditional learning methods since the established process of learning from conventional datasets was not designed to and will not work well on high volumes of data. An effective Fuzzy Supervised Learning (FSL) approach which contributes a new algorithm with Fuzzy logic implemented a new fuzzy membership function for nonlinear classification. With the emergence of big data, much more needs to be done to address many challenges including learning for large scale of data, learning for different types of data, learning for high speed of streaming data, learning for uncertain and incomplete data, and learning for extracting valuable information from massive amounts of data.

2. Related Work

From the data analytical viewpoint, it is essential to secure sensitive data, to protect private information and to manage data quality, exists whether data sets are big or small. However, the specific properties of big data (volume, variety, velocity, veracity) create new types of risks that necessitate a comprehensive strategy to solve the big data considerations.

For research, many challenges are significant to develop theories and scalable techniques that can extract knowledge from large, dynamic, multi relational information sources and to close the semantic gap between structured data and human notions and concepts. For effective future prediction, data analysis using statistical modeling techniques may be applied enhances and supports the organization’s business strategy.

2.1. Analysis on Big data using Machine learning approaches

The massive amounts of data highlight the current research efforts and the challenges to big data, as well as the future trends. The other is to analyze the connections of machine learning with modern supervised learning techniques for big data processing from different perspectives. Machine leaning is a field of research that formally focuses on the theory, performance, and properties of learning systems and algorithms. Thus, it is a highly inter disciplinary field building upon ideas from many different kinds of fields such as artificial intelligence, information theory, statistics, and many other disciplines of science.

Particularly, digital data, generated from a variety of digital devices, are growing at astonishing rates. According to [2], digital information has grown nine times in volume in every 5 years. Therefore, the term “Big Data” was coined to capture the profound meaning of this data explosion trend. Besides describing the fundamental techniques and technologies of big data, a number of more recent studies have investigated big data under particular context. For example, [7, 8] gave a brief review of the features of big data from Internet of Things. Some researchers also analyzed the new characteristics of various big data processing models and algorithms from the data mining perspective. Over the past decade, most machine learning techniques have been used in a number of massive and complex data-intensive fields provide possible solutions to mine the information hidden in the data. For instance, most traditional machine learning algorithms are designed for data that would be completely loaded into memory [9], non separable data for which there does not exist a hyperplane that successfully separates the positive from the negative examples.

Because of its implementation in a wide range of applications, machine learning has covered almost every scientific domain, which has brought great impact on the science and society [10]. It has been used on a variety of problems, including recommendation engines, recognition systems, informatics and data mining, and autonomous control systems [12].Generally, the field of machine learning is divided into three sub domains: supervised learning, unsupervised learning, and reinforcement learning [9].

Briefly, supervised learning requires training with labeled data which has inputs and desired outputs. In contrast with the supervised learning, unsupervised learning does not require labeled training data and the environment only provides inputs without desired targets. Reinforcement learning enables feedback received through interactions with an external environment. Based on these three essential learning paradigms, a lot of theory mechanisms and application services have been proposed for dealing with data tasks [11]. Alternating direction method of multipliers (ADMM) [3,4] serving as a promising computing framework to develop distributed, scalable, online convex optimization algorithms is well suited to accomplish parallel and distributed large-scale data processing. The key merits of ADMM is its ability to split multiple variables in optimization problems, which enables one to find a solution to a large-scale global optimization problem by coordinating solutions to smaller sub-problems [4].

2.2. Learning for high dimensional data

Data mining is of an exploratory nature and can also be seen as exploratory data analysis with a special focus on large data collections. Some well-known analysis methods and tools that are used in data mining are, for example, statistics (regression analysis, pattern recognition analysis etc.), and Fuzzy classification analysis and association rules. On developing concepts, such methods are concerning data analysis and classification of fuzzy-valued data (fuzzy-set).

Real time data are high velocity(heterogeneous, high dimensional, non-linear) and high velocity(data streams with high speed).So data distribution is changing over time, which needs the learning algorithms need to solve the challenging issue of data are often non-stationary. To surmount these obstacles, rule-based fuzzy learning is significant different from traditional learning methods. Convex optimization is integrated into the learning progress by randomization and distributed computing. They mainly referred to the scalable, randomized and flexible fuzzy logic for big data analytics.

2.3. Possible remedies : Fuzzy Set

Learning from these massive data is expected to bring significant opportunities and transformative potential for various sectors. However, most traditional machine learning techniques are not inherently efficient or scalable enough to handle the data with the characteristics of large volume, different types, high speed, uncertainty and and low value density. In response, machine learning needs to reinvent itself for big data processing. So, the critical issue for Fuzzy learning tends to emphasize the importance of the managing the uncertainty and incompleteness on data quality.

Fuzzy sets are sets whose elements have degrees of membership. Fuzzy sets are an extension of the classical notion of set. Important points for the interpretability of a fuzzy system are that there are only few fuzzy rules in the rule base. Each rule should use only a few variables and the variables should be partitioned by few meaningful fuzzy sets. It is also important that no linguistic label is represented by more than one fuzzy set.

There are several ways to induce the structure of a fuzzy system. Cluster-oriented and hyper box-oriented approaches to fuzzy rule learning create rules and fuzzy sets at the same time. Structure-oriented approaches need initial fuzzy partitions to create a rule base. The main problem of both approaches is that each generated fuzzy rule uses individual membership functions and thus the rule base is hard to interpret. Cluster-oriented approaches additionally suffer from a loss of information and can only determine an appropriate number of rules.

Structure-oriented approaches avoid all these drawbacks, because they do not search for hyper-rectangular clusters in the data space. By providing (initial) fuzzy sets before fuzzy rules are created the data space is structured by a multidimensional fuzzy grid. A rule base is created by selecting those grid cells that contain data. This can be done in a single pass through the training data. Structure-oriented methods allow the user to provide appropriate fuzzy partitions in advance such that all rules share the same fuzzy sets. Thus the induced rule base can be interpreted well.

3. Structured Rule Based Fuzzification

The structured rule based approach operates in an iterative manner to eliminate rules weighted by weak weights specified in a k-norm on classifier model.

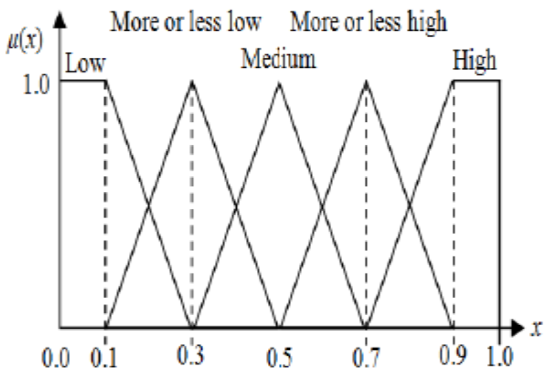

Figure 1. Member function with rule based

(1) Generate rules as **Rule Rj:** IF x_1 IS A^1j AND …AND x_n IS A^nj THEN Class = Cj with RWj"

(2) Builds the fuzzy partition using equally triangular membership function

(3) Builds the RB creating a fuzzy rule associated to each example

(4) Output rules may be created:

- Same consequent → Delete duplicated rules
- Different consequent → Preserve highest weight rule

The same antecedent rule can be generated proximity between two rules to remove more than one per-iteration. For every membership function,

$$\mu_A : X \rightarrow [0,1] \qquad (1)$$

$$\mu_i(x) = \frac{\sum_{j=1}^{k} \mu_{ij} \left({}^{1}\!/_{||x-j||^{2/(m-1)}} \right)}{\sum_{j=1}^{k} \left({}^{1}\!/_{||x-j||^{2/(m-1)}} \right)} \qquad (2)$$

X – space; $x \in X$, **A** - linguistic variable, concept, fuzzy set.

$\boldsymbol{\mu_A}$ – a Membership Function (MF), determining the degree, to which *x* belongs to *A*.

The evaluation of mean squared error based on K-Nearest Neighbor method. Thus the size of the rule base can be determined automatically by adding rules ordered by their proximity until all training data is covered.

Problem solving rules were selected by the model according to measures of usefulness, applicability, and simplicity. Rules were further discriminated by their use of symptomatic information for pattern recognition or information seeking. A production rule consists of two parts: condition (antecedent) part and conclusion (action, consequent) part:

IF (conditions) THEN (actions)

IF x_1 is small and x_2 is small THEN Class 2
IF x_1 is small and x_2 is medium THEN Class 2
IF x_1 is small and x_2 is large THEN Class 1

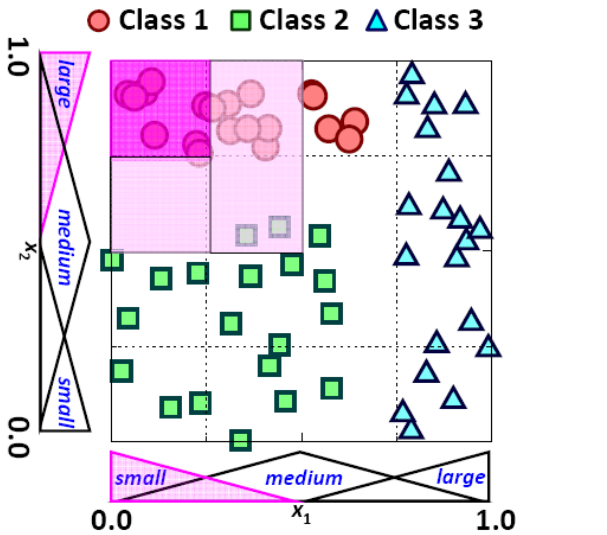

Figure 2. High interpretability of rule evaluation

Data can be classified and categorized into fuzzy sets (with membership value), developed a rule-base structure. The number of fuzzy rules can also be restricted by including only the best rules in the rule base. It is also possible to use pruning methods to choose the number of rules and the number of variables used by the rules. Each rule should use only a few variables which are partitioned by a few meaningful fuzzy sets.

4. Review on Fuzzy rule-based System

Fuzzy set methods can profitably be applied in several phases:

The business understanding and data understanding phases are usually strongly human centered and defined the goals of the knowledge discovery project, to estimate its potential benefit, and to identify and collect the necessary data. In addition, background domain knowledge and metadata knowledge is gathered. In these phases, fuzzy set methods can be used to formulate, for instance, the background domain knowledge in vague terms, but still in a form that can be used in a subsequent modeling phase.

First,an example from the induction of possible graphical models from data which complements the induction of the well-known probabilistic graphical models. The second class, fuzzy data analysis, consists of methods that use fuzzy techniques to structure and analyze crisp data, for instance, fuzzy classification for data segmentation and rule generation. In the evaluation phase, in which the results are tested and their quality assessed, the usefulness of fuzzy modeling methods becomes most obvious.

Initially, data pre-processing is performed to reduce the noise and redundant data to speedup computation. As the dataset is divided into training and testing subsets, the training subset is used for the feature extraction. In this manner, subset from given features are selected. Statistical developments with fuzzy data coming from the fuzzy perception of real-valued ones will be mainly based on propagating the associated imprecision to the distribution function, parameters, etc.

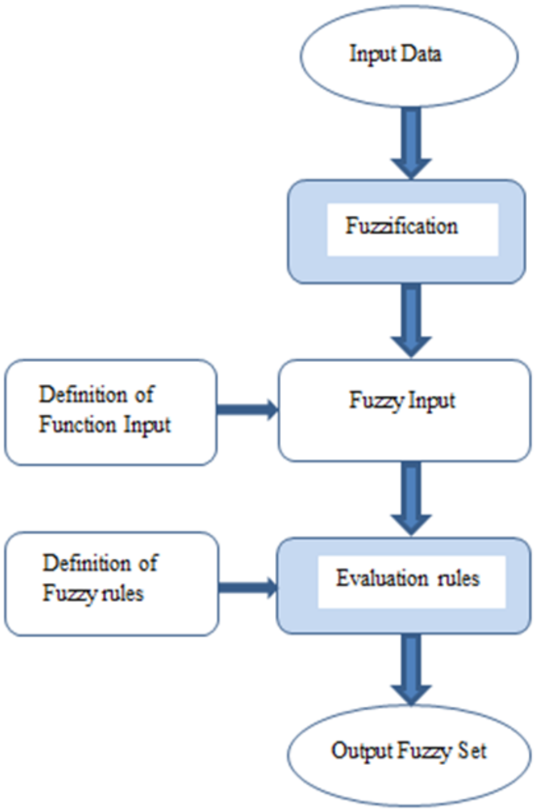

Figure 3. System flow for rule based fuzzification

4.1. Map Reduce Function

There are several tools that have been designed to address data processing in Big Data problems. Among them, the most popular one is the *MapReduce* distributed processing system. This frame work simplify the massive parallel processing of large datasets by providing a design pattern that instructs algorithms to be expressed into two primary functions, which specified by users: Map and Reduce. In the first phase, “Map” is used for each record of functional computation to produce some intermediate results. Afterwards, these intermediate results will be fed to a second phase in a “Reduce” function, which aggregates the output which applies the final results.

- Map function: The *Map* function receives a <key,value> pair as input and emits a set of intermediate <key,value> pairs as output. Then, these intermediate <key,value> pairs are automatically shuffled and ordered according to the intermediate key and will be the input to the *Reduce* function.
- Reduce function: The *Reduce* function obtains the intermediate <key,value> pairs produced in the previous phase and generates the corresponding pair as the final output of the algorithm.

1. In the Map phase, “key-value” pairs are associated to the labels of the antecedent of the rules. The antecedent is used for joining rules in the Reduce phase.

2. “Key-value” pairs are linked to the whole rule antecedent for each Map process. Therefore, the Reduce stage is devoted to build an ensemble of all learned rules.

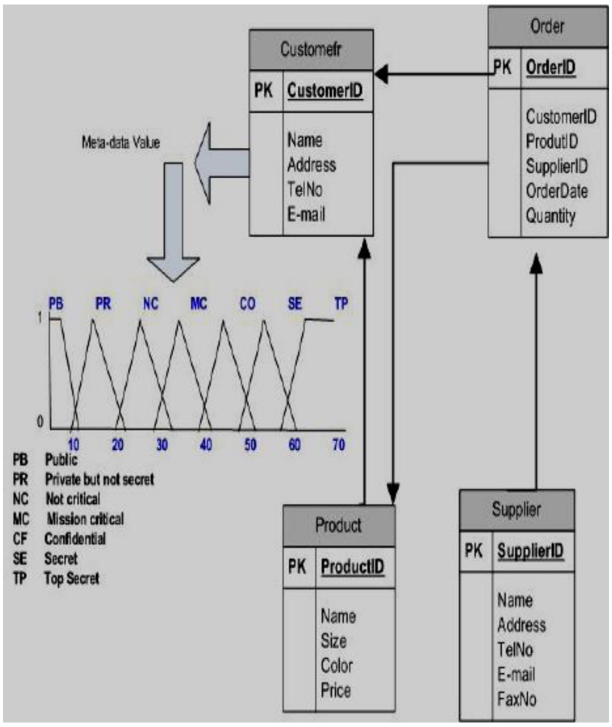

Figure 4. Apply rule based fuzzification on relational tables

4.2. Features' Characteristics

Generally, features are characterized as:

- **Relevant** : These are features have an influence on the output and their role cannot be assumed by the rest.
- **Redundant:** A redundancy exits whenever a feature can take the role of another.
- **Information:** Feature X is preferred to feature Y if the information gain from feature X is greater than that from feature Y.
- **Dependence:** The coefficient is a classical dependence measure and can be used to find the correlation of feature X with class (1) is higher than the correlation of feature Y with class (1), then feature X is preferred to Y.

The relevant features are possible, by reducing the dimensionality of the data and enables learning algorithm . Assuming the relational data:

Customer (CustomerID, Name, Address, TelNo, E-mail)
Product (ProductID, Name, Size, Color, Price)
Supplier (SupplierID, Name, Address, TelNo, E-mail)
Order (OrderID, CustometID, ProductID, SupplierID, OrderDate, Quantity)

Linguistic variable are, **TP** = top secret, **SE** = "secret", **CO** ="confidential", **MC** = "mission critical", **NC** = "not critical", **PR** = "private but not top secret", **PU** = "Public".

Table 1. Analysis on performance and number of rules

Class	Instance Number	Kddcup_DOS vs. normal data set	
		8 Maps	
		Num Rules by Map	Final Num Rules
DOS	973	RB_1 size:211	RB_R size: 301
		RB_2 size:212	
		RB_3 size:221	
		RB_4 size:216	
Normal	3883	RB_5 size:213	
		RB_6 size:210	
		RB_7 size:211	
		RB_8 size:214	

Fuzzy implication with Map Reduce model, Table.1 shows the result of rules by reducing instance number efficiently. Linguistic variable (instance) are used over mapping functions to improve performance and computational cost.

5. Conclusion

There is a constantly growing demand to keep solutions conceptually simple and understandable. In this study, a new approach for big data classification task using rule based Fuzzy was developed. The algorithm does not use gradient-descent, because the degree of fulfillment of a fuzzy rule is determined by the minimum and non-continuous membership function has been used. Instead a simple heuristic is used that results in shifting the fuzzy sets and in enlarging or reducing their support. The main idea is to create comprehensible fuzzy classifiers, by ensuring that the fuzzy sets cannot be modified arbitrarily during learning. Constraints can be applied in order to make sure that the fuzzy sets still fit their linguistic labels after learning. The fuzzy classifier always prefers the learned rule; select the rule with larger performance weight value. The mixed approach is used such as reducing mapping and learning expert rule to complete the rule based fuzzification model.

References

[1] C.-H. Jen, C.-C. Wang, B. C. Jiang, Y.-H. Chu, and M.-S. Chen, "Application of classification techniques on development an early warnin gsystem for chronic illnesses," Expert Systems with Applications ,vol. 39, no. 10, pp. 8852–8858, 2012.

[2] D Che, M Safran, Z Peng," Big data mining: challenges, issues, and opportunities", in Proceedings of the 18th International Conference on DASFAA (Wuhan, 2013), pp. 1–15

[3] F Andersson, M Carlsson, JY Tourneret, H Wendt, A new frequency estimation method for equally and unequally spaced data. IEEE Trans Signal Process 62(21), 5761–5774 (2014)

[4] F Lin, M Fardad, MR Jovanovic, Design of optimal sparse feedback gains via the alternating direction method of multipliers. IEEE Trans Automat Contr 58(9), 2426–2431 (2013)

[5] F. Pinto, C. Soares, and J. Mendes-Moreira, “A framework to decompose and develop metafeatures,” in Proceedings of the 2014 International Conference on Meta-learning and Algorithm Selection-Volume 1201.CEUR-WS. org, 2014, pp. 32–36.

[6] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” science, vol. 313, no. 5786, pp. 504–507, 2006.

[7] H Hu, Y Wen, T Chua, X Li, Analysis on scalable systems for big data analytics: a technology tutorial. IEEE Access 2, 652–687 (2014)

[8] K. Kowsari, D. E. Brown, M. Heidarysafa, K. Jafari Meimandi, , M. S.Gerber, and L. E. Barnes, “Hdltex: Hierarchical deep learning for text classification,” in IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2017.

[9] M. Fatehi and H. H. Asadi, “Application of semi-supervised fuzzy cmeans method in clustering multivariate geochemical data” Ore Geology Reviews, vol. 81, pp. 245–255, 2017.

[10] P. Brazdil, C. G. Carrier, C. Soares, and R. Vilalta, Metalearning: Applications to data mining. Springer Science & Business Media, 2008.

[11] R. Mall, V. Jumutc, R. Langone, and J. A. Suykens, “Representative subsets for big data learning using k-nn graphs,” in IEEE International Conference on Big Data. IEEE, 2014, pp. 37–42.

[12] X Wu, X Zhu, G Wu, W Ding, Data mining with big data. IEEE Trans Knowledge Data Eng 26(1), 97–107 (2014)