TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.3, March 2014, pp. 2028 ~ 2035
DOI: http://dx.doi.org/10.11591/telkomnika.v12i3.4132

Construction of E-commerce Recommendation System Based on Semantic Annotation of Ontology and User Preference

Hongsheng Xu, Ruiling Zhang*, Chunjie Lin, Wenli Gan
College of Information Technology, Luoyang Normal University
Henan LuoYang, 471022, China
***Corresponding author, e-mail: zhangruilingls@163.com**

Abstract

Ontology is an explicit formal specification of the shared conceptual model. Domain ontology can accurately describe the contents of the e-commerce site in the stage of built domain ontology. Web domain ontology can effectively improve the quality of the recommended use of the e-commerce personalized recommendation. The paper presents construction of e-commerce recommendation system based on semantic annotation of ontology and user preference. In this paper domain ontology semantic annotation for the project is established, and this ontology takes advantage of semantic similarity measure to calculate the semantic similarity between the projects, and it can predict the score of unbranded items based on user ratings. The experimental results show that the algorithm recommended more effective than traditional algorithms, to improve the quality of the recommended.

Keywords: *domain ontology, semantic annotation, E-Commerce recommendation, user preference*

1. Introduction

In recent years, Ontology is a research focus in computer and related areas of common concern, as a modeling tool for modeling the conceptual model of the information system on the semantic level. Ontology in knowledge engineering, personalized recommendations, system modeling, information processing, digital libraries, natural language understanding, has been widely used in areas such as the Semantic Web. Since the 1990s, various scholars in the field from their own application point of departure-depth study of the theory and application of the body, and some research body theory and technology has been rapid development.

Ontology initially originated in Western philosophy, from the scope of philosophy, ontology is the objective existence of a system of explanation or description of the abstract nature of the objective things. With the development of information technology, the ontology is introduced to the field of artificial intelligence, used to represent the knowledge sharing and reuse [1]. The study of artificial intelligence Niches et al ontology is defined as "given the basic terms and relations constitute the terms of the related fields, and constitute these vocabulary piracy rule definition.

E-commerce recommendation system (Recommendation System) came into being. Recommendation system to interact directly with customers, analog-store sales staff to provide customers with merchandise recommended helping customers find the necessary goods to the successful completion of the purchase process. E-commerce recommendation system is able to assist viewers into buyers, improve cross-selling of e-commerce, effective customer retention, to prevent the loss of customers and improve the sales of the e-commerce system. Therefore, with good prospects for the development and application of e-commerce recommendation technology has gradually become an important research content.

A variety of ontology representation: said that based on the concept of a simple classification level, the framework of concepts and attributes and semantic network, there is a wealth of skills and some logical reasoning function representation based on description logics (Description Logic).Description logic is a formal object-based knowledge, it inherits the main idea of the KL-ONE, is the first-order predicate logic is a decidable subset. It is the first-order predicate logic; description logic systems can provide decidable reasoning services. Description

Received August 13, 2013; Revised September 9, 2013; Accepted October 13, 2013

logic as a logical formula used to express the knowledge and reasoning concept aggregates, also known as the term logical or conceptual representation language, through a series of formulas to provide definitions of terms, atoms connected by constructing character conceptual representation of a complex concept.

Semantic annotation software agents are faster than existing software agents to extract and translate Web content in the Web Services. Web Services semantic annotation can facilitate service discovery, can also contribute to the service is configured workflow. But the practical application of this semantic restriction is in the service of the general application of a small amount of label. The paper presents construction of e-commerce recommendation system based on semantic annotation of ontology and user preference.

2. Ontology and Semantic Annotation Method

The current Internet is composed of a large number of HTML pages. The site is mainly used for reading presented to the user, the manifestations of the information is not, internal structure and expression of phase separation, so the machine does not understand the content of these information, it is difficult to be automatically processed. In response to these issues, the founder of the WWW TimBerners a Lee in December 2000 in the XML2000 meeting put forward the concept of the next generation of the Internet - the Semantic Web (Semantic Web), its hierarchy. "The Semantic Web is the semantic information on the extension of the current Web, the Semantic Web, allowing between computers and between humans and better able to work together."

Ontology representation language should have well-defined syntax, semantics, reasoning ability, as well as a wealth of semantic skills, ensuring that the user for the domain model is to write clear, formal concept description. Since the 90s of the last century, the large number of domestic and foreign scholars dedicated to the study of ontology representation language, the birth of many-ontology description language, RDF and RDF-S, DAML-ONT, OIL, DAML + OIL, OWL [2]. OWL is the W3C in February 2004 a description logic-based ontology description language, the proposal of OWL is an important milestone in the process of the development of the Semantic Web, and OWL has been widely discussed and unanimously approved. OWL based on RDF and RDFS by adding more modeling primitives to describe the nature of the class, as well as the relationship between its richer type definitions and attribute descriptions for nature, as is shown by Equation (1).

$$\xi_{ij}(k)=\left[1+\left|\frac{\Delta x_i(k)}{\sigma_i}-\frac{\Delta x_j(k)}{\sigma_j}\right|\right]^{-1} \tag{1}$$

Automatic semantic annotation, identify semantically rich OWL instance. They think, OWL lacks the instance recognition descriptive semantic OWL can not become automatic semantic annotation excellent ontology language, so they asked the atoms, vocabulary, and grammar concept to increase the semantics of OWL description to identify and contain the external representation.

At present, most of the build domain ontology or ontology editing tools to manually edit the domain ontology construction is far from engineering activities. Different ontology development group has its own design principles, design criteria and definitions. While many foreign universities, research institutes and laboratories have developed a variety of ontology editing tools, the user ontology editing, and more mature. However, domain ontology construction time-consuming, laborious and costly, and has become an accepted fact, as is shown by Equation (2) [3].

$$F=\frac{U/p}{Q_e/(n-p-1)}\sim F(p,n-p-1) \tag{2}$$

In the field of ontology construction process, how to regulate their build process in accordance with the method of engineering has become an important research content. Domain

ontology construction should follow the idea of software engineering, as a new type of software activities, learn software engineering experience. In the domain ontology provides a class hierarchy, if class A is B ancestors, or class B is A ancestors, say A and B for the same support class (Same Branch Classes), in order to facilitate the description, suppose class A is B ancestors, then A is called A and B minimum Gong Tongzu first, denoted by R (A, B); A, B between the semantic distance D (A, B) =dep (B) dep (A), where Dep (C) as a class C in the hierarchy depth.

Protégé graphical domain ontology developed by Stanford University to build tools, it provides users with a range of tools to support building a domain ontology model. It provides visual support ontology class instance edit, browse and update. Users can take advantage of standard plug-ins to extend the functionality of Protégé. Not only be able to support its own ontology language, Protégé can also support other popular RDF, RDFS and OWL ontology language, as is shown by Figure 1.

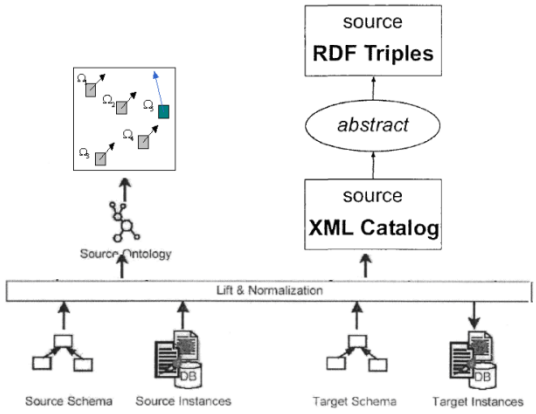

Figure 1. The Overall Model of Fuzzy Concept Lattice

A plurality of ontology is into a common interactive protocol. Here, an ontology is kept separate, but among them, at least one original body (original ontology) is used as a body overlapping portions of the conceptual model and word matching. However, the body can be in different fineness (level) to describe the area difference. Semantic interoperability is a body coordinate (ontology reconciliation) base, semantic interoperability can be used to belong to different ontology entities searching relation, so that the process for " body checking (ontology Aligning).

Typically, the body works include needs analysis, development, integration, evaluation modules. Ontology utilizes the continued expansion of the ontology, the ontology fusion and purification (maintenance), search queries, reasoning and other applications [4]. Ontology located on the fourth floor in the seven-layer architecture of the Semantic Web by Tim Berners-Lee, and its purpose is to capture the knowledge of related fields, to provide a common understanding of the domain knowledge to determine the common recognition in the field of vocabulary and give these terms, as is shown by Equation (3).

$$\frac{f'(x_1)}{2x_1} = (b^2 + a^2)\frac{f'(x_2)}{4x_2^3} = \frac{\ln\frac{b}{a}}{b^2 - a^2} x^3 f'(x_3) \tag{3}$$

The distance between items is an important relationship between the projects. Project from a [0, ∞) between the real number, a project with its own distance is 0. In general, project similarity between the distances with the project has a close relationship. Two items the greater the distance, the lower the similarity; conversely, the smaller the distance of the two items, the greater the degree of similarity. It can create a simple correspondence between distance and similarity between two items. This correspondence between the need to meet the following

conditions: the distance of the two projects, and its similarity to 1; distance of the two projects infinity, its similarity to 0; greater the distance of the two projects, and its similarity smaller.

Each ontology class has multiple different instances, the difference between them lies in the different instances of different concepts of the value of a property may contain the same attributes, so the similarity when comparing two instances, should also be considered where the similarity between the properties of the class. Set ontology class C1 instance I1, the value of its corresponding attribute P1 p1, corresponding to the value of the properties of p2 is p2, ... corresponding to the value PN of the property Pn, that the instance I1 = C1 [P], P is attribute vector (p1, p2., ... pn) the ontology concept C2 instance I2 is, its corresponding attributes of Q1 value Q1, the corresponding properties of Q2 value Q2, ..., the corresponding properties of Qn value Qn, said the instance I2 = C2 [Q], Q attribute vector (q1, q2, ..., qn).I1 = (p1, p2, ... pn) and I2 = (q1, q2, ..., qn), the similarity is calculated through its public property.

3. Ontology and Semantic Annotation Method

The recommendation algorithm consumer buying behavior is generally influenced by cultural, social, personal characteristics, and psychological factors. Only by understanding consumer behavior, in order to better recommend to users the goods of interest to the user. In this paper, based on the user's personal characteristics clustering, it is in order to solve the sparsely problem in collaborative filtering. For e-commerce the recommended websites registered user, through the user registration information can be collected to some of the data on the user's personal characteristics, such as the user's age, gender, their education degree.

Collect customer information: This includes the personal property of the customer purchase history, web page history. Personal property can be obtained from the user registration form; of course, the premise is that the user must fill in true and valid personal information. Purchase history is stored in the background transaction database in the e-commerce site; it is the details of each user of all the previous shopping records, including shopping time, the list of goods, prices, and discounts, as is shown Equation (4).

$$\begin{cases} \dfrac{dI}{dS} = -1 + \dfrac{1}{\sigma S} \\ I_0 + S_0 + R_0 = 1 \end{cases} \tag{4}$$

For e-commerce sites, such information is likely to buy than the user record is more important, can help companies analyze the target user preferences and to improve the structure and performance of the online marketing system, introduce appropriate measures to improve product click-through rate. Web page history is stored in the server's log files, record user access behavior, frequency, and content user sources [5].

In order to solve the extreme sparsely of user rating data, the user is usually not score of the project is set to a fixed default value, value: 0, scoring domain intermediate values (for example: 5 score rating is set to 3 points) or the average score for the user, these improvements help to improve the precision of collaborative filtering recommendation. However, these methods do not take into account the semantic association may exist between the different projects and upgraded items score in the practical application, the user can not be exactly. So the above method does not fundamentally solve the extreme user rating data sparse case recommended accuracy significantly reduces the problem, as is shown by Equation (5).

$$\omega(t) = \begin{cases} 0 & 0 \le t < 40 \\ \log_{10}\left(\dfrac{t-40}{3.78} + 1\right) & 40 \le t < 74 \\ 1 & t \ge 74 \end{cases} \tag{5}$$

Due to e-commerce sites have not mastered basic customer information and history information, it is necessary to collect this type of customer information in real time and make a recommendation. The main access path information collected, sequence mode, the dwell time.

The entire recommendation process from the customer to enter the site began, along with the actions of the customer site; the more accurate the classification of the recommendation system for customers, more refinement, recommended a process from course to fine results. The key to the implementation of this type of recommendation system is to improve the real-time nature [8]. The high amount of information customers is all frequent visits to customers. These customers are loyal customers of the e-commerce site, is also recommended service principal. Such customers can choose the content-based recommendation systems and collaborative filtering-based recommender system combination.

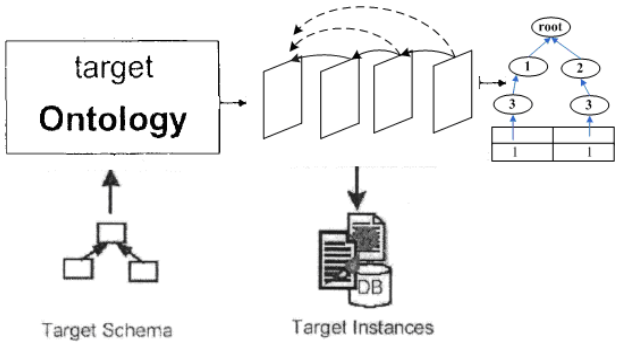

Figure 2. Semantic User Access Preferences Diagram

Recommended collection of semantic user preferences through the reverse mapping between domain ontology and Web pages, semantic user preferences file instantiated as Web pages and recommended to the current user. In addition, the use of domain ontology semantic user clustering based on easily implement more granular personalized recommendation service based on domain ontology classes and Web pages [6].

Calculated based the project collaborative filtering algorithm increases with increasing m and the number of projects the number of users n, the worst case time complexity of O (mn).Collaborative filtering recommendation algorithm based on domain ontology and user preference changes the time complexity for Part 3: User rating matrix missing values filling: Project similarity calculation between time complexities of O (n2), the user is not rated the project initial forecast time complexity of O (mn), as is shown by Equation (6).

$$\begin{cases} x_0^1(t) = (1 - c^1(t) f^1(t) \\ x_1^1(t+1) = \beta^1(t) \sum_{i=i_1}^{i_2} b_i^1(t)^* x_i^1(t) \\ x_{i+1}^1(t+1) = (1 - d_i^1(t)) x_i^1(t) \end{cases} \quad (6)$$

Due to non-registered users just browse operation, there is no data on the project's score, and collaborative filtering recommendation techniques can not provide personalized recommendations. To non-registered users personalized recommendation, it is necessary to take advantage of Web usage mining, looking for the access to the user's preferences, and then according to the current user session with user access preference matching, to generate personalized recommendation set. Mining Web-based e-commerce personalized recommendation technology, through Web transactions affairs conducted clustering analysis, the user clustering cluster page with the right to value collection, and you can export a Web document, the Web user preference analysis and generate personalized recommendations.

4. Construction of E-commerce Recommendation System Based on Semantic Annotation of Ontology and User Preference

In the traditional research of semantic search, a unified ontology is predefined, then the ontology tag network resources and the relationships between resources are existing shortcomings, it can establish a guidance learning statistical models, and dig out the social tagging semantic information, solve social tagging of polysemy and multiple word meaning problem, clear social annotation of semantic information.

Use of Web usage mining to some extent to improve the quality of recommendation, and achieved a certain effect, but due to the traditional Web usage mining related areas of semantic knowledge is not considered in the process of personalized recommendation, these personalized recommendation system can not use the object itself the semantic recommendation, leading to the discovery of user interest in the satisfaction that the accuracy rate is relatively low, and the use of cluster analysis recommended page coverage is high, excessive noise information, as is shown by Equation (7) [7].

$$W_{s3} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{\lambda_3} \overline{W}_{1-2-3} - \frac{\lambda_1}{\lambda_3} W_{s1} - \frac{\lambda_2}{\lambda_3} W_{s2}. \tag{7}$$

A clear definition of the relationship between the vocabularies, it is semantic concepts to describe the relationship between the concepts [8]. Ontology-based semantic annotation using good body support content creators to add semantic metadata to Web pages, so that its contents can be understood by humans and machines, compared to a top-down folksonomy defined by expert's taxonomy. Tagging, semantic annotation tool for a large number of existing Web information will make the contents of the Web page to be machine-readable data, which constitute the basis of the Semantic Web. The paper presents construction of e-commerce recommendation system based on semantic annotation of ontology and user preference, as is shown by Figure 3.

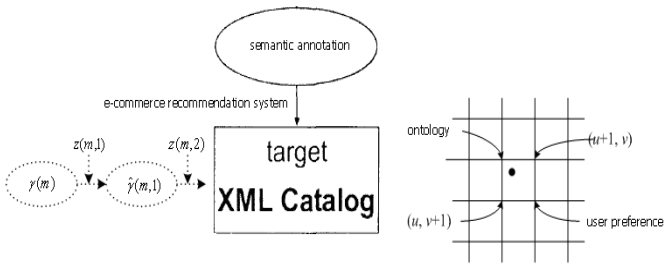

Figure 3. System Structure Diagram

Semantic annotation technology from three different angles to consider: ① The label or labeling for dynamic Web page (Web page generated by the database) from the marked objects can be divided into the static Web page; ② marked can be divided into manual, semi-automatic or automatic, manual annotation is marked by hand semantic metadata information is written directly to the source code of the Web page, semi-automatic annotation with the tools decided to use the mouse to drag and way you want to label the content with the tools the letter 10 to write Web pages, automatic annotation from concept to automatically write semantic information in Web pages, but actually write what you want with the training set of a large number of similar pages, and then to human knowledge discriminates decision (theoretically can not be classified as a fully automatic), when a small amount of, or different types of Web pages so marked an effect is not satisfactory with automatic annotation, as is shown by Equation (8).

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \mathrm{L} + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \mathrm{L} - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \theta_q \neq 0 \end{cases} \tag{8}$$

Select upgraded predictive value in the project, the top N items recommended. All rated the importance of traditional collaborative filtering algorithm [9]. Not change user preferences are above approach is reasonable, but when the user preferences change the situation is different. Forgotten the law of the user preferences, it is the importance of user early scores than the user's current ratings decreased. To determine the similarity between users based on forgetting the law, the neighbor user so that the calculated more accurately.

By domain ontology-based user access preferences on the basis of the acquired user preferences associated matrix Mt×s: discovery algorithm can get the preference patterns of different user clustering cluster user access on the semantic level [10]. Domain ontology marked the Web page; you first need to create a mapping between domain ontology and Web pages.

Semantic similarity between the domain ontology project types and attributes computing projects using KNN thought unbranded predict user score based on user ratings of the project, filling user rating matrix scarcity problem is solved to a certain extent. The use of user characteristics factors clustering. Users of the same class characteristics similar to the average preference for certain items and it is so the user clustering method can narrow the selection of the user's nearest neighbor. This can not only improve the efficiency of the algorithm can also be appropriate to reduce the rating data is sparse, the common user ratings fewer projects, as is shown by Equation (9).

$$M = \left| f(c) \right| = \left| f(0) + f'(\xi)(c-0) \right| = \left| f'(\xi)c \right| \le A \left| f(\xi) \right| c \le \frac{1}{2} \left| f(\xi) \right| \le \frac{1}{2} M \tag{9}$$

The explicit input: refers to in the e-commerce site recommendation system will take the initiative to collect user browsing and buying behavior, (1) the user will take the initiative to their needs and hobbies submitted to the recommendation system. (2) implicit input: input and display, implicit input is a passive form of recommendation system will be hidden on the site, the user is not aware of its existence, the system will record the user's browsing behavior as input, and in ordinary website browsing behavior is the same. (3) Input Keywords: information entered by the user in the site's search engine is the e-commerce recommendation system recordkeeping; collate information useful keywords, or the user browsing product attributes and information as the recommended system input. (4) User purchase history input: Recommended system user previously purchased goods analysis after finishing as the input, the recommendation system is usually that the user purchases the commodities, said they liked. (5) The user text input: user can purchase goods for the commodities evaluate recommendation system does not analyze this information, this information to the user, the user can see other users evaluation to determine the kinds of goods good or bad.

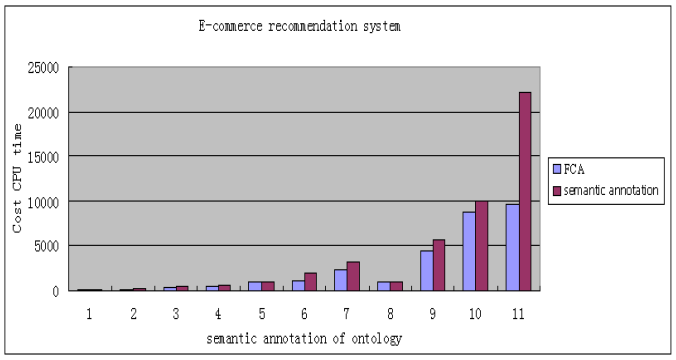

Figure 4. E-commerce Recommendation System Based on Semantic Annotation of Ontology and User Preference Compare with FCA Graph

Cluster analysis of Web user access logs, session clustering. However, how to assess the quality of clustering is an important issue in the mining Web usage. A good clustering result should be the same clustering clusters within the user's session should be as compact as

possible, with other clustering cluster distance should be as far as possible. User session clustering is completed; the next task is the evaluation for the quality of the clustering. In order to evaluate through the Web user preferences based on semantic clustering algorithm to obtain the clustering quality, the use of a particular measure - Weighted average access proportion. The paper presents construction of e-commerce recommendation system based on semantic annotation of ontology and user preference.

In the folksonomy law can be considered the user to add the label, the use of artificial intelligence and ontologies analyze the positioning of the label to the user where tree, or even knowledge of the architecture of the mesh, and the convenience of the user from the overall understanding of the problem, in order to achieve a combination of folksonomy from the bottom up and the body from the top-down. Marked developers can better organize information, users can retrieve information better. Traditional network of Web2.0 label introduced ontology, semantic annotation participation. Semantic annotation already applied research, but the semantic annotation development needs of the body deeply involved.

5. Conclusion

In ontology construction phase, most of the modeling work has focused on the abstract representation of the domain knowledge, that found that the field concept, the concept of inheritance hierarchy, the potential relationship and axioms. Generally not be considered in the modeling will be able to identify all instances involved in the body unless modeling. Instances involved in real-world applications and associated abstract ontology concept, which is what the semantic annotation of work. The paper presents construction of e-commerce recommendation system based on semantic annotation of ontology and user preference.

Web usage mining can discover the habits and patterns of common interest user groups as well as individual users will be applied to e-commerce recommendation system can effectively improve the quality of recommendation, Web usage mining has become the mainstream of the current e-commerce personalized recommendation.

Acknowledgements

This paper is supported by the National Natural Science Funds of China (61272015), and also is supported by the science and technology research major project of Henan province Education Department (13B520155) and Key scientific and technological project of Henan province (102102310058).

References

[1] Shota Sato, Yumi Asahi. A Daily-level Purchasing Model at an E-commerce Site. *IJECE*. 2012; 2(6): 831-839.
[2] Bucklin RE Sismeiro C. Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach. *Journal of Marketing Research*. 2004; 41(3): 306-323.
[3] Xin Li, Ting Li. E-commerce System Security Assessment based on Bayesian Network Algorithm Research. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(1): 338-344.
[4] Beynon M. Reducts within the variable precision rough sets model: a further investigation. *European Journal of Operational Research*.2001; 134: 592-605.
[5] Zhu LB, Dai GZ. A information security assesment model based on fuzzy fuzzy comprehensive evaluation. *Journal of Information Security and Communication Secrecy*. 2006; 8(2): 13-15.
[6] Murphy KP. A brief introduction to graphical models and bayesian networks. *Telkomnika; Indonesian Journal of Electrical Engineering*. 2008; 21(2): 164-171.
[7] LiFeng Zhang, ShuWen Yang, MingWang Zhang. E-commerce Website Recommender System Based on Dissimilarity and Association Rule. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 12(1): 154~161.
[8] Du Xiao-yong, Li Man. Research on Ontology Learning. *Journal of Software.* 2006; 17(9): 1837-1847.
[9] Nan Feng, Xue Yu. A Data-driven Assessment Model for Information Systems Security Risk Management. *Journal of Computers*. 2012; 7(12): 3103-3109.
[10] Lijuan Zhou, Hui Wang, Wenbo Wang. Parallel Implementation of Classification Algorithms Based on Cloud Computing Environment. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(5): 1087~1092.