Personalized Recommendation Method of E-commerce Based on Fusion Technology of Smart Ontology and Big Data Mining

Shengli Jiang¹,², Yi Zhang¹,² and Hongsheng Xu¹,²,*

¹Luoyang Normal University, Luoyang, China
²Henan key Laboratory for Big Data Processing & Analytics of Electronic Commerce, Luoyang, China

*Corresponding author e-mail: 85660190@qq.com

Abstract. E-commerce personalized recommendation system needs to be able to respond to user page requests in real time and provide recommendation services for users according to different recommendation models chosen by users. In this paper, based on the fusion method of intelligent ontology and big data mining, the semantic access preference of the user is captured by the user interest acquisition algorithm of the semantic cluster, and the current session of the user is matched with the acquired user preference to obtain the recommendation set. The paper presents personalized recommendation method of E-commerce based on fusion technology of smart ontology and big data mining.

Keywords: E-commerce, Personalized recommendation system, Big data, Smart ontology, Web page

1. Introduction

E-commerce website provides users with more commodity choice space and more convenient trading methods. On the other hand: for businesses, by collecting implicit and explicit ratings in the process of users visiting e-commerce websites, it is helpful for businesses to have a deeper understanding of users' demand information and shopping behavior characteristics [1]. In order to better provide personalized services to users.

Ontology is introduced into the field of network information resources organization, because the research of ontology focuses on a broader space, that is, to construct a top-level conceptual framework for human cognitive activities [2]. Ontology emphasizes the function of knowledge sharing and focuses on giving the general framework of knowledge (or domain knowledge) of human knowledge, in order to organize the information resources on the Internet into a semantic web and knowledge network.

Learning resource customization and personalized learning: how to find learning resources that meet the needs of users in the available information, and it is customization and personalized learning according to personal interests and hobbies has become a hot spot. The traditional knowledge representation technology is seriously mismatched with the way of thinking of observing problem
solving. These presentation techniques are mainly machine-oriented and have a low level of abstraction. They can not provide intuitive and easy to understand ways to represent the organization and reasoning control of problem solving.

The data in Web is usually included in pages encoded in HTML format. The data contained in these Web pages is a kind of semi-structured data. Although the data can be understood, it is not easy to be understood and processed by machines.

Ontology defines terms and relationships and reasoning rules between terms, Logic. It can provide a powerful logical language to realize reasoning, and once Logics is established, it can verify the effectiveness of resources, the relationship between resources and the reasoning results through logical reasoning.

RDFs is a simple body language that defines the RDF vocabulary provided by the w3C. It is a semantic extension of RDF, and provides a mechanism for naming and describing attributes and classes in RDF. In RDFs, the concepts of class attribute, instance and inheritance are introduced, and a group of markup with fixed semantics is defined to model the domain knowledge.

2. Intelligent recommendation technology of e-commerce based on domain ontology and user preference change

E-commerce recommendation system makes use of Web data and user-item scoring data as data source, and implements a prototype system. The electronic commerce recommendation system supports the complete recommendation management function and the unified recommendation management platform, supports many kinds of recommendation models, and realizes the combination recommendation of many kinds of recommendation models.

Ontology divides different types of entities that exist in the objective world. Concepts of universal significance are expressed in this kind of ontology, which do not depend on specific problems or subject areas [3]. In this kind of ontology, the defined knowledge can be applied interdisciplinary, which also includes vocabulary related to things, events, time, space and region. General wooden bodies can deal with the time-material properties of physical objects.

In problem solving, state space representation and problem reduction representation are the earliest knowledge representation. Knowledge-based systems emphasize expertise and general knowledge related to problem areas. The common knowledge representations are: first-order predicate logic representation, production representation, frame representation, as is shown by equation (1), where it is script representation and semantic network representation [4].

\[
x^i(t + 1) = \beta^i(t) \sum_{j=0}^{l_i} b^j_i(t) x^j(t) - s_0 \alpha N^1(t)
\]

The dialect pattern in the information source is used to match the dialect pattern to determine the ontology concept to which the dialect belongs, so as to achieve the purpose of automatically recognizing the ontology concept in the information source. The accuracy of dialect pattern determines the quality of information extraction. Chinese grammar is extremely complex, and it is reliable for domain experts to determine dialect patterns.

For a piece of data encryption, it is the key technology to realize Web trust, and the public key encryption algorithm is the basis of digital signature. Machines and software agents can use it to verify that information is provided by a particular trusted source without ambiguity.

All constraint words containing OWL in OWLDDL are divided into two categories: Object Property used to associate individuals and individuals and Datatype properties for associating individuals and specific values [5]. In addition to the Domain and Range that can define the attribute, the description and inference of the attribute are enhanced.

Web uses data Web usage data is data that describes the behavior of users accessing the site, and Web usage data is obtained by configuring the Ennib server log and the Web application server log.
Everything (thing) has an attribute (property). Things that have no attributes and all attributes must be attached to things. Attribute is equivalent to giving things a formal (form). We describe things by attribute. The world is made up of all kinds of things, and everything has its own attributes, as is shown by equation (2).

\[ x = \left( x(0) - \frac{u}{a} \right) e^{-at} + \frac{u}{a} \left( x_j, y_j \right) \quad (j = 1, 2, \cdots, n) \]

(2)

It represents knowledge through concepts and their semantic relations. A semantic network is a marked directed graph. Among them, the nodes of the directed graph represent all kinds of plums, concepts, attributes, actions, states, etc., and the directed arc represents some semantic relationship between the nodes to which it is connected, and each node can have several attributes. One of the simplest semantic network forms is a triple: (node 1, arc, node 2).

Our strategy to solve the first conflict is to select the dialect pattern with the largest support (s in quad) as the correct matching pattern of all the dialect patterns that c matches, and the data of column c is mapped to the daml: DatatypeProperty. Corresponding to the pattern.

**Definition 1.** This paper describes the Language class, which is associated with LODO:language through the attribute prefLanguage to represent the learning resource representation language that the user prefers.

In the large-scale knowledge base, there must be a considerable amount of inconsistent and unintegrated knowledge [6]. The problem of knowledge verification has been mentioned earlier. Because of the incomplete knowledge in the knowledge base, there will always be some inaccuracy in the verification. With the continuous supplement and correction of knowledge, the theory that was originally considered to be correct may be proved to be wrong, and the theory previously thought to be wrong may be considered to be correct.

After identifying the daml: Class corresponding to WRBT, the matching range of table header column names is determined, and the strategy described earlier is used to solve the first possible conflict. In most cases, a large number of conflicts occur because of the great differences in dialect patterns of different concepts.

The tagging rules based on manual rules are defined by the user, and the system dimensions the documents according to the tagging rules. Most of the manual rules are implemented by defining ‘wrapPer’. It mainly uses the context information of dimension information (including text or structure information in context) to identify the semantic information in the document.

Firstly, in the mining document set, the feature sentence extraction, feature sequence generation and sentence frequent feature pattern mining methods proposed in this chapter are used to mine the pattern and extract the rules. After the rules are obtained, they are on the experimental set of semantic tagging.

The most important thing is the Web server log [7]. The web server log records the request information for each page when the user accesses the site. Each record in the Web log includes the user's IP address, the page that the user's ID, user requests to access.

**Definition 2.** M and D are actually copies of concepts and relations in Yin's corresponding ontology. C and R in P can form a directed calibration graph, which we call the node in the ontology (ontology graph), graph as the concept in H. The edge is the relationship in R, as is shown by equation (3).

\[ H_{MD} = \frac{1}{2} \sum_{i=1}^{N} m_i(u)^2 + V \]

(3)

DeepWeb refers to the dynamic generation of Web pages from the database according to the request of the user. How to label these dynamic contents is also one of the important tasks of semantic Web tagging.
The threshold and support parameters of the feature sentence selected in this paper are biased for mining long feature patterns. The higher feature sentence threshold makes the feature sentence extraction algorithm tend to add more long sentences composed of multiple clauses to the feature sentence set, while the lower support parameters are conducive to mining long and complex patterns.

The data source of data preprocessing includes user-item scoring data and Web usage log file. The preprocessing in data mining is mainly to determine the data mining task according to the mining needs of users, extract the input of mining algorithm, check the original data according to the background knowledge and constraint rules, and reduce the data through data cleaning.

The degree of trust in knowledge is the probability that knowledge is ultimately considered to be correct. Because it is difficult to calculate this probability accurately, we only divide the trust into several rough levels, and we can not get the exact value yet. After verification and is considered to be correct knowledge, it will get a high degree of trust, verified and considered wrong knowledge, will get a lower degree of trust, unverified knowledge depends on its source.

The most important part of the Competency ontology is the Learning Skill child body. The Learning Skill sub-ontology describes the learning process and behavior that can be applied to knowledge. It is usually represented by a behavioral verb, such as application, analysis, influence, simulation, etc.

3. Ontology-based Web Information extraction and Big Data Mining Fusion method

Delete useless data from the Web log server, such as log data where the status response code begins with 4 (client error), 5 begins (server error) and part 3 begins; Delete images, audio, video files and other information independent of user access mode, such as log file suffix name GIF, JPEG, JPG, CSS, JS and so on.

A large number of nodes connect to the network by connecting with the central node, and this small number of central nodes is the core concepts in the field of knowledge. Therefore, the scope of the domain of knowledge can be calculated in two ways, one is to define a set of core concepts, and the radius distance that the domain of knowledge emits from the core concepts [8]. The node which is within the range of distance from the core concept belongs to this field of knowledge. The other is to define a set of core concepts.

Learning skills are divided into five levels to receive (Receive), response (Responding), evaluation (Valuing), organization (organizing) and self-management (self-manage), then refined in turn, divided into level 2 and level 3. We define the nine learning skills in level 2 as the difficulty level from 1 to 9.

Definition 3. Relational dependency in an integrated environment). In an integrated environment, any concept C1 belongs to ontology viewpoint P1, c2 belongs to ontology viewpoint P2, and there are relations R1, R2 and r3.

The method SVRE of the syntactic relationship between subject and predicate object and the method of analyzing the whole sentence of Chinese sentence DTRE. Compared with the SA method, both the Reall value and the F1 value are higher. The reason is that DFRE can deal with all kinds of sentences by using dependency analysis technology, while the SA method can only deal with sentences that match the rules, as is shown by equation (4) [9].

\[
\Gamma (n) = \frac{1}{M} \sum_{m} \gamma_m (n) = \text{Re} \left\{ e^{j2\pi} \frac{1}{M} \sum_{m} \psi_m (n) \right\}
\]

(4)

The rule-based learning method is suitable for tagging documents with a certain pattern structure, which is easy for users to understand, flexible in application, and supports friendly user interaction. The tagging method based on classification model has stronger learning ability than rule-based method and can represent more complex features.

By extracting the eigenvalues of item I, the Property (i), attribute set of its attribute set is calculated to determine the matching degree between an item and the recommended item. User access preference
file PropertyBased Profile (u) is defined according to the weight of keywords. T is a weight vector, and wui represents the importance of keyword ki to user u.

Synonymous terms: when the two terms are the same in semantics, then they are the same, but they represent different appellations. It is possible to use dictionaries or glossary to solve the problem, but should be aware of the differences in its possible scope (see conceptualized mismatching described above).

Definition 4. let the knowledge field Q be a finite set of a topic, and the set K of a subset of Q is called a knowledge space if and only if K contains an empty set and a complete set of topics Q. for the subset weight K1 of any topic, K2 ∈ K, their union K1 ≈ K2 ∈ K.

The middle layer ontology is constructed according to the local ontology, so the construction process of the middle layer ontology is essentially an ontology integration process. Based on the multi-view theory in demand engineering and some limitations and assumptions, we propose a multi-view based intermediate layer ontology construction method, as is shown by equation (5).

\[
p_0(t) = \frac{u}{\lambda + u} + \frac{\lambda}{\lambda + u} e^{-(\lambda+u)t}
\]

The dependency of tagging information can be used to improve the accuracy of tagging, but relational extraction is relatively difficult. In different types of documents, the characteristics of information dependence are also different, some are linear dependency, and some are nonlinear dependency [10]. Existing sequence tagging models usually only consider linear dependency, and it is but can not describe other nonlinear dependency.

On this basis, the semantic annotation of the webpage content is converted into the classification problem based on the two dimension context, and the two-stage method is adopted, namely, the block detection stage and the text labeling phase. The blocks may be rows, sections, or paragraphs, depending on the size of the block granularity. Use the vertical context to probe the blocks in the web page; use the instance in the horizontal context discovery block.

The recommendation algorithm based on association rules can be divided into two stages: recommendation model establishment stage and recommendation model application stage. In the offline stage, the data mining algorithm is used to discover the customs rules and establish the recommendation model. The model establishment stage is very time-consuming, but it can be carried out offline. The online stage provides real-time recommendation services to the users according to the state of the users of the association rules.

4. Personalized Recommendation Method of E-commerce Based on Fusion Technology of Smart Ontology

Web data is usually unstructured or semi-structured. Web content mining can be used to help users collect useful information or filter out useless information. There are various tags in web pages that imply the importance of Web page content. Reasonable use of these tags is helpful to improve the performance of Web text mining.

A typical execution of GLUE is created and the performance of each component in the system architecture is evaluated. The main components to be evaluated are different types of learning machines and relaxed label machines for classification. They use domain constraints and heuristic algorithms in order to propose better matching. The experimental results show that the performance of several hybrid classifiers with domain heuristic algorithms can be improved significantly according to the requirements of accuracy.

According to the theory of knowledge space, we can model the domain knowledge space. On the basis of the theory of knowledge space, we improve the knowledge from the original problem set to the set of knowledge points. The knowledge space of the subject domain constructs the knowledge framework of the whole subject, which includes the knowledge points in the field and the relationship model between the knowledge points.
**Definition 5:** (special relation, hasSubtype), special relation is the inverse relation of generalizing relation, which is equivalent to isSuperclassOf relation. The parent concept can be divided into several sub-object concepts according to a certain feature.

A method of generating ontology concepts is by using table data on Web pages. A table on a Web page is regarded as an instance set of concepts in ontology, and the attributes in the table are also attributes of concepts. Through the analysis of a large number of tables, the concepts and their attributes can be extracted.

Wordnet and GATE are used to mark the entity and entity relationship. For the dimension document, Artequakt first uses the natural language analysis tool (Apple Pie Parser) to analyze the syntax and dependency, then uses GATE and Wordnet to identify the named entities in the document, and then uses wordnet to analyze the meaning of the word to realize the concept relation tagging.

It provides three levels of document tagging functions, which are ontology-based content tagging, free text statement and instant document extension. Users can use these tagging features when reading and editing documents. Through adaptive information extraction for text (Amilcare), system is applied to provide semi-automatic content tagging function.

The Web page has a certain structure with respect to a plain text file. In addition, the Web page also contains the link relationship between the web pages, so the main goal in the Web structure mining is the Web mining of the Web link structure and the recognition of the authoritative Web page.

5. **Conclusion**

The pattern analysis stage mainly uses the knowledge of domain experts and other available standards to analyze these patterns and filter out those that have no application value and deviation. The patterns found are displayed in tables, pie charts, graphs, trend diagrams, rectangles, or other special forms of representation. In the personalized recommendation technology of e-commerce, the use of Web usage mining can improve the recommendation quality to a certain extent, and has achieved certain results.

**Acknowledgments**

This paper is supported by Henan key Laboratory for Big Data Processing & Analytics of Electronic Commerce, and also supported by Guiding Plan of General Project of Humanities and Social Sciences Research of Henan Education Department in 2019 (2019-ZDJH-048).

**References**

[1] Jiang Shuhao, Zhang Bole. Research on the correlation between e-commerce recommendation system and customer satisfaction. e-Business. 2016 (07):100-109.

[2] Xu Hongsheng. Novel Approach of Semantic Annotation by Fuzzy Ontology based on Variable Precision Rough Set and Concept Lattice, International Journal of Hybrid Information Technology Vol.9, No.4 (2016), pp. 25-40.

[3] Luo Ping, Tan Baoming. Research on individualized Sports Marketing recommendation system based on Electronic Commerce. The mall is modernized. 2007 (36):87-95.

[4] Xu Hong-sheng, Zhang Rui-ling. Application of ART and concept similarity in E-commerce recommendation systems. Journal of Computer engineering and Applications, (2010),46(7):213-217.

[5] Cong Jiazhi, Jiang Shuhao. Personalized e-commerce recommendation system. Channel Technology and Industry. 2017(06):286-294.

[6] Xu Hongsheng. Semantic Annotation of Ontology by Using Rough Concept Lattice Isomorphic Model, International Journal of Hybrid Information Technology Vol.8, No.2 (2015), pp.93-108.

[7] Sánchez, D., Isern, D., &Millán, M. Content annotation for the semantic web: An automatic web-based approach. Knowledge and Information Systems, 2011,27:393–418.

[8] Sun Ji Yun. Technical analysis and research of electronic commerce recommendation system. Information construction. 2016 (02):456-462.
[9] Xu Yuan, Zhang Chunhong, Ji Yang. Semantic Internet of things: starting with ontology construction. Telecommunications Network Technology, 2014, (3): 13-15.

[10] Jason J. Jung. Semantic annotation of cognitive map for knowledge sharing between heterogeneous businesses. Expert Systems with Applications, 2012, 39(5): 5857–5860.