Cluster Computing
https://doi.org/10.1007/s10586-017-1437-8

Electronic commerce commodity entity identities based on hierarchical probability model

Sheng-fu Zhang[1]

Received: 21 September 2017 / Revised: 5 November 2017 / Accepted: 21 November 2017

Abstract
How to identify the same commodity entity from the multi-source heterogeneous, autonomous, independent, diverse, and inconsistent electronic commerce data is the main challenge for the present. By analyzing the data characteristics of different platforms, an index model based on commodity attribute/value is established firstly, and then we construct the global pattern of attribute value of commodity. It comes into being a unified model, quality and efficiency of commodity information data. Then based on the hierarchical probability model, the similarity of the identity of the goods is measured. We finished commodity entity recognition. And the normalized output is to meet the same set of commodities and sort. We construct experiments based on the Hadoop platform for the 3 B2C e-commerce data sources. And the traditional methods and products are compared. The experimental results demonstrate the feasibility, accuracy and efficiency of the framework.

Keywords Hierarchical probability model · Entity identities · E-commerce

1 Introduction

With the continuous development of e-commerce, e-commerce big data for the industry and academia has brought valuable opportunities and challenges [1]. E-commerce in bringing convenience to people's life and work at the same time, more people is able to look forward to in-depth discovery and excavation of more valuable information and knowledge.

Automatic identification of all web pages that describe the same commodity entity from a variety of electronic commerce data sources is the basis for data integration and data analysis. But the multi-source heterogeneous e-commerce data has a huge number of types of goods, different modes, uneven data quality, and diverse website structure characteristics of clutter. It lacks a unified model definition and theoretical model. These have greatly influenced the analysis and application of large data of electronic commerce. In this paper, a hybrid framework is proposed based on Map-Reduce architecture, which is based on data index, data integration, entity recognition and data ordering. In the big data environment the entity recognition is achieved by machine learning. Output to meet the same data sets and related attributes, and to sort. This research can be applied to the data fusion, data search and personalized recommendation in the electronic commerce environment.

2 Related works

The research on the identity of the commodity entity in the big data is a typical branch of the data entity recognition. Altwaijry et al. [2] and Li et al. [3] took empirical knowledge to solve the problem of the entity identity. They can effectively solve the multi data tuple in the database, which describe the same real world entity. But they are not suitable to solve the same problem in multiple data sources. For the first time, Fan et al. [4] proposed the entity identity description rule. Through the problem of rule reasoning, the entity identity description rule is no longer a loose collection, but can be mutual inference and coordination. The ability to effectively improve the identity of the entity is described. Deng et al. [5] determined the identity of the entity by the method of clustering. Nuray-Turan et al. [6] took machine learning methods to determine the identity of entities with multiple attribute similarity measurements. These methods are not accurate enough in the complex Web environment. Douglas et al. [7] proposed

✉ Sheng-fu Zhang
zhangshengfuls@163.com

[1] School of Computer, Qinghai Nationalities University, Xining 810007, Qinghai, China

Published online: 09 December 2017

a similarity measure to describe the identity of entities by taking the method of Markov chain. Ribeiro et al. [8] described the similarity between XML nodes by sliding windows. They realized the entity identity on XML data and obtained good results. But it did not apply to the web big data environment entity identity [9–12].

3 Definition and description of the problem

Definition 1 (*Commodity entity*) The commodity entity refers to the abstract of the commodity data which has the concrete characteristic in the electronic commerce big data. And it indicated by symbol E.

Definition 2 (*Commodity object*) The object of goods refers to the object represented by the data recorded in the electronic commerce data. Usually the object is described by a page or a group of pages. Therefore, it is also known as the object page. And it is indicated by symbol W. Commodity object is the only specific data exists. For any one of the e-commerce data sources S_i, the data source contains a large number of commodity web pages, and these pages describe the same product information in different categories and different forms of expression. For any product page W_j, the structure information and content information are contained in the product page. Through the structure analysis, information extraction and semantic mining, the object data model is obtained as follows.

$$W_j = (C, E, B)$$

Here, the C represents the columns and structural information of the commodities. It is the site, columns, pages and pages of sub-columns abstract. The site is described as a 5-layer non-empty tree. Website data source is the root node, columns and all levels of sub columns are intermediate nodes, web pages are leaf nodes. Page W column structure information can be expressed as $C = \{c_1, c_2, c_3, c_4, c_5\}$. The C_1 is the root node, and the C_5 is the minimum node. E-commerce data source S_i all pages of the site C forms S_i column space

The E represents the commodity entity described by the commodity object. The B represents a collection of feature attributes data items that are included in this page.

Definition 3 (*Tree model based on attribute/value*) In order to unify represent all e-commerce data, we propose a model based on the key data and e-commerce features for a combination of tree structure representation. The main idea is that all data will be obtained after the initial processing into hierarchical key data element. The model includes unit and data unit related key data type. An object data is composed of a number of metadata with a hierarchical relationship, each of which is encapsulated as a four tuple.

$$Node = \langle E_i, P_i, key, value \rangle$$

Here, the E_i is the entity described in the metadata. The P_i identifies the parent object for the metadata. The *key* attribute describes the name of the object. The *value* attribute describes the variable length data. The model indicates that the model itself does not analyze the data content, and does not recognize any data structure. This enables it to handle all data types. The Web data format and data type of any electronic commerce can be converted into the model, so it has high scalability.

4 Index of commodity entities in large data

To reduce the complexity of NP-hard problem for commodity between multiple data sources pairwise comparisons, we will construct the inverted index table for the attribute value of each attribute in this paper. Then we can I effectively check the goods of the same entity [13–15].

We set up inverted index set R. Each inverted index record can be expressed as $R_i = \{A, V, Z\}$. Here, the A represents the name of the property. The V indicates its value. The Z represents a collection of any subtree for commodities containing $key = A$ and $value = V$. That is $Z = \{W_{i_1}, W_{i_2}, \ldots\}$. To traverse each data source for each commodity page, we extract the product page of each feature item key on the *Node* and configure into a hierarchical tree. If the node (*key*, *value*) is in the R, the product page is added in the Z of the record. Otherwise, we add a record in the R, that is $\{key, value, \{W_i\}\}$. Finally we form the property/value of all the data sources for the inverted index table collection R.

In the actual e-commerce platform, there are many expressions that are different but are equivalent to each other. Therefore the normalization of attributes/values is a combination of the properties and values that are equivalent to each other. Thus it is advantageous to carry on the comparison of data items and entity recognition more accurately.

According to the record of all properties/values in the inverted index set R, we establish global schema diagram $G = \langle M, N, H \rangle$. Here, the M represents a set of points that are combined by all attributes. The N represents a point set consisting of all values. The H is a set of weighted edges for the point set of connection properties and values. For any property $A \in M$ and values $V \in N$, the number of commodities Z for feature items $\langle A, V \rangle$ in R is k. Then there is a weight of k in the $\langle A,V \rangle$, the weight of it is $\omega\langle A, V \rangle$.

Definition 4 (*Equal value set*) All the elements in a set of values are equivalent to each other. These values describe the same or similar meaning. Such as the value of the collection is a combination of the value of mutual. Generally

we take the cumulative weight of the largest elements as a representative of the node. The set of values is expressed as $\bar{V} = \{V_1, V_2, \ldots\}$. The values expressed by each of these elements are equivalent to each other. Due to the complexity of e-commerce platform and other value sets can be extended to all the elements of similarity is greater than a given threshold μ. That is, for a collection of equal value $\bar{V} = \{V_1, V_2, \ldots\}$, $\forall V_i \in \bar{V}, \forall V_j \in \bar{V}$:

$$Sim_{value}(V_i, V_j) \geq \mu_1$$

The essence of the construction of the value set is to set up the clustering algorithm based on the semantic similarity of value text. The similarity of V_i and V_j is as follows.

$$Sim_{value}(V_i, V_j) = \frac{L_i + L_j}{2L_iL_j} \times (\delta \times (1 - \lambda \times (1 - S_0)) + \sum_{i \in X, i \in Y} \gamma_{ij} \times (1 - \lambda \times (1 - S_1)))$$

Here, L_i and L_j denote the value of the concept of two numbers. The λ represents the importance of word orders similarity to semantic similarity. Its value is generally below 0.5. The X and Y respectively represent two values of similarity concept sequence. The γ is called as the concept similarity of similar concepts.

The map phase will have the same value of the mode for the merger [16–18]. And the index of value is set up as key. In the reduce stage, we compare the values of the input V_{j1} with other value set. If the value which is set to a value V' of each element satisfies $Sim_{value}(V_j, V'_k) > \mu_1$, The V_{j1} is added to the value set V'. And all the attributes and relationships associated with V_{j1} are updated. A new relationship $(\langle A_{j'_1}, V', H'_{j'_1}\rangle, \langle A_{j'_2}, V', H'_{j'_2}\rangle, \ldots)$ is formed. If the value of the input V_{j1}, it cannot find a suitable set of values, the value is used as a set of equal value separately. The equivalent combination of all values is finally realized [19,20].

Definition 5 (*Equivalent set of attributes*) All the elements in the set of equivalent attributes describe the same properties of commodities. And they are equivalent to each other. That is in the equivalent set of attributes $\bar{A} = \{A_1, A_2, \ldots\}, \forall A_i \in \bar{A}, \forall A_j \in \bar{A}$:

$$Sim_{attr}(A_i, A_j) > \mu_2$$

When combined equivalent of attributes, it not only measure the semantic similarity attribute names, attribute values also analyzed two cases of orthogonal range. Therefore, the semantic identity between attribute nodes is defined as follows.

$$Sim_{attr}(A_i, A_j) = \frac{1}{2} \times Sim_{str}(A_i, A_j) + \frac{1}{2} \times Sim_{range}(A_i, A_j)$$

Here, the Sim_{str} represents the similarity of text. The Sim_{range} indicates the degree of fitting in the range of values, that is $Sim_{range}(A_i, A_j) = \frac{Q(A_i) \cap Q(A_j)}{Q(A_i) \cup Q(A_j)}$. Here, the Q represents a range of the property.

In the map stage, we will have the same value relationship mode to merge and establish a property bond index. In the reduce stage, the input of the attribute A_{j1} and each of the set for equivalence properties are compared. If the property is a set A'_k of each attribute element satisfies $Sim_{attr}(A_{j_1}, A'_k) > \mu_2$, the A_{j1} property will be added to the equivalence set A'_k and updates all data points associated with the A_{j1} edges and relationships. Then we form a new relationship $(\langle A'_{j'_1}, V', H'_{j'_1}\rangle, \langle A'_{j'_2}, V', H'_{j'_2}\rangle, \ldots)$. If the input attribute A_{j1} cannot find the appropriate set of equivalent properties, the property is solely as an equivalent set of attributes. The equivalent combination of all attributes is finally realized.

5 Entity identity recognition based on hierarchical probability model

The aim of electronic commerce commodity entity identities is to find the same product description of each commodity pages. In practical applications, it is difficult to describe exactly the same two pages. Therefore, the condition of the same entity is equivalent to two commodities: $Sim(W.B, W_i, B) > \sigma$. Here, the *Sim* is a function of calculating the similarity of two commodities.

The core of the product information is a collection of data items which are composed of multiple attributes/values. By comparing the data item sets of two commodity pages, their similarity and identity are measured. Taking into account the existence of some special attributes of goods, the impact of different attributes on the entity identification is different. Therefore, the attribute is classified and the weight is set. Attributes of the $W.B$ goods can be extended to each feature of key nodes.

$$Node_i = \langle E_i, P_i, key, value, T, P, w\rangle$$

Here, the $Node_i$ represents the i-th node of commodities. We number for traversal methods by first order traversal. The T represents the data item type. It includes general data items and special data items. The P indicates the credibility of the value of the i data item. The ω represents the weight of the attributes of the first i data item in entity recognition. The weight of the range is [0, 1].

Due to the large amount of data, the comparative efficiency is low for each commodity to traverse. Therefore, this article establishes a three-tier hierarchy probability tree. The leaves at each level of the leaves are expressed as a candidate set of product to satisfy the probability decision. The right node of each layer represents the final node that does not satisfy the condition. This greatly reduces the number of decision; improve the efficiency of the algorithm. For a given commodity W_a, we find out the three levels of each layer of the same product entity page:

(1) *Level 1* Identifying possible candidates for the same entity. Commodity W_a includes a total of k common data items. This paper will meet with the k general data items in the presence of more than φ_k of the same data item as the candidate set of commodity goods. Here, the φ is a candidate threshold,$\varphi \in (0, 1)$.
(2) *Level 2* Candidate product group similarity screening.

There may be four kinds of data item sets for commodity W_i and candidate commodity W_a.

① There is data satisfies completely equal conditions, that is.

$$W_i.B.Node_1.A = W_a.B.Node_1.A,$$
$$W_i.B.Node_1.A = W_a.B.Node_1.V.$$

② Candidate attribute is not present in the commodity trade in W_a, that is.

$$W_i.B.Node_2.A \notin W_a.B.Node.A.$$

③ Commodities W_a attribute is not present in the candidate commodities, that is.

$$W_a.B.Node_3.A \notin W_i.B.Node.A.$$

Meeting in the case of items 2 and 3 is called missing items.
④ The candidate has the same properties as the W_a, but the value is not the same:

$$W_i.B.Node_4.A = W_a.B.Node_4.A,$$
$$W_i.B.Node_4.V \notin W_a.B.Node_4.V.$$

Meeting in the data item 4 is called contradictory items.

As a result, the data item similarity between commodity W_a and commodity W_i can be measured as.

$$\begin{aligned} &Sim_{item}(W_a, B, W_i, B) \\ &= \left\{ \sum_{i \in A_1} Node_i \cdot \omega \right. \\ &\quad \left. - \sum_{i \in A_4} Node_i \cdot \omega \right\} \Big/ \left\{ \sum_{i \in A_1} Node_i \cdot \omega + \sum_{i \in A_2} Node_i \cdot \omega \right. \\ &\quad \left. + \sum_{i \in A_3} Node_i \cdot \omega + \sum_{i \in A_4} Node_i \cdot \omega \right\} \end{aligned}$$

Here, A_1, A_2, A_3, A_4 were expressed as the attribute set in ①–④. The $Node_i \cdot \omega$ is the weight of the corresponding attribute. The $-\sum_{i \in A_4} Node_i \cdot \omega$ is expressed as the apparent contradiction of the data and the similarity of the punishment value.

Finally, the candidate that does not meet the commodities $Sim_{item}(W_a, B, W_i, B) \geq \tau$ is out of the commodity set.

(3) *Level 3* Setting similarity filter items based on specific candidates for commodities.

In addition to the general property of the commodity, there are some pictures, such as title, price, details, user comments and other special items. The title and the price is the most representative.

① Title is the most direct reflection of the attributes of the content of commodities, but also is the most direct reflection of the difference between the different attributes of goods, and it should have the greatest weight. Because of different business habits and a lot of businesses in the title, they add some of the products which are not unrelated to the hot key words. The similarity of the title of the same product may be low. It is also possible that the title similarity of different commodities is very high, which seriously affects the results of the experiment. To overcome this problem, we use the word entity titles tool to extract key terms. We give the title keyword vector $K = (k_1, k_2, \ldots, k_n)$. Then, the correlation measure is carried out for each of the key words and the entire data item set of the commodity. We get the key words and the degree of the commodities. After standardization, it is used as the weight of the title key words. Then, we can get the weight vector $\omega_k = (\omega_{k_1}, \omega_{k_2}, \ldots, \omega_{k_n})$. Here $\sum_{i=1}^{n} \omega_{k_i} = 1$. Finally, we calculate the similarity of the two titles.

For commodities W_a and W_b, the key words of the title are $k_a = (k_{a_1}, k_{a_2}, \ldots, k_{a_n})$, $k_b = (k_{b_1}, k_{b_2}, \ldots, k_{b_m})$. The weight vectors are $\omega_{k_a} = (\omega_{k_{a1}}, \omega_{k_{a2}}, \ldots, \omega_{k_{an}})$ and $\omega_{k_b} = (\omega_{k_{b1}}, \omega_{k_{b2}}, \ldots, \omega_{k_{bn}})$. The key words of the two commodities are combined into $k_{ab} = k_a \cup k_b = (k_{a'_1}, k_{a'_2}, \ldots k_{a'_n}, k_{b'_1}, \ldots, k_{b'_m})$. Here, the title of the commodity W_a is the vector $(k_{a'_1}, k_{a'_2}, \ldots k_{a'_n})$. The $(k_{b'_1}, k_{b'_2}, \ldots, k_{b'_m})$ is a vector containing the key words in the W_b, but not in the W_a. That is $k_b - k_a$.

So the weight vector of commodity W_a and W_b can be extended to $a_n + b_m$, dimensional vector. If the element is not in its original vector, the weight is 0. So the similarity of the title of the two commodities can be expressed as follows.

$$Sim_{title}(W_a, W_b) = \frac{\omega k_a \cdot \omega k_b}{||\omega k_a||^2 + ||\omega k_b||^2 - \omega k_a \cdot \omega k_b}$$

We give a filter not meeting $Sim_{title}(W_a, W_b) \geq \xi$. Here, the ξ represents the threshold value of the title for the commodity.

② Price is an important characteristic of different commodities, especially for commodity and subsidiary of the same name merchandise such high similarity but do completely different product. The distance between the two commodity prices is defined as follows.

$$Sim_{price}(W_a, W_b) = \sqrt{1 - \frac{|W_a \cdot price - W_b \cdot price|}{\max(W_a \cdot price, W_b \cdot price)}}$$

We give a filtration of commodities which not satisfied with $Sim_{price}(W_a, W_b) \geq \rho$. Here, the ρ represents the threshold value of the same price of commodities.

We measure the reliability of two product pages W_a and W_b, belong to the same product.

$$Sim(W_a.B, W_b.B) = \{Sim_{item}(W_a.B, W_b.B) + Sim_{title}(W_a, W_b) + Sim_{price}(W_a, W_b)\} \Big/ 3$$

In the map phase, the results of the query for all data items to attribute/value are decomposed. In the reduce stage, for each data item W_a, we lookup product group Z which has the same data item in the inverted index table. We find out the $k$data items in the W_a commodity for collection $Z_1, Z_2, \ldots, Z_k$ in the emergence of more than ϕ_k times the commodities. We calculate the similarity of the data items of each item in Z' and W_a. The commodities W_i not meeting $Sim_{item}(W_a.B, W_i.B) \geq \tau$ are out of the candidate set Z'. We calculate the similarity of the title items of each item in Z' and W_a. The commodity W_b not meeting $Sim_{title}(W_a, W_b) \geq \xi$ is out of the candidate set Z'. We calculate the price range of W_a and each commodity in Z'. The commodity W_b not meeting $Sim_{price}(W_a, W_b) \geq \rho$ is out of the candidate set Z'. The Z' in the final product is the page that describes the same commodity entity as the W_b, and it is as the final output. And we calculate the product W_i and W_a similarity $Sim(W_a.B, W_i.B)$ as the output of *value*. This value is used to measure the identity of the two items. The output is sorted according to the value.

6 Experiments

In order to effectively test the accuracy and efficiency of the proposed method for large-scale e-commerce data parallel computing, a detailed experiment is carried out in this paper. Experimental data sets from three Chinese mainstream integrated B2C e-commerce platform for real-time data collection.

6.1 Data set

The entire commodity information collection includes 10 categories 38 sub-categories of goods 938,781. The number of different types of commodities in different platforms is shown in Fig. 1.

6.2 The evaluation criterion

In order to judge the effectiveness of the algorithm and the experiment, we test the accuracy and efficiency of the algorithm from two aspects.

(1) Accuracy

In this paper, the average accuracy rate of P, the average recall rate of R, and the average comprehensive evaluation index $F1$ are used as the criteria for the accuracy of the recognition results. Let the candidate product set is $Z = \{W_1, W_2, \ldots, W_n\}$. The k commodity which is identified as the same commodity is $W'_k = \{W'_{k_1}, W'_{k_2}, \ldots, W'_{k_m}\}$. Here, the k_m indicates the number of the same commodities identified by the k commodity. According to the actual data analysis, we can get there are k_{m1} goods identification errors. The true identity of the same number of goods is k_{m2}. The average accuracy of the algorithm is as follows.

$$P = \sum_{k=1}^{n} \frac{k_m - k_{m1}}{k_m} / n$$

Average recall rate of the whole algorithm is as follows.

$$R = \sum_{k=1}^{n} \frac{k_m - k_{m1}}{k_{m2}} / n$$

Average comprehensive evaluation index of the whole algorithm is as follows.

$$F1 = \frac{2 \times P \times R}{P + R}$$

(2) Efficiency The running time per 100 thousand data (RT) and the computation time increment speed (IS) is took

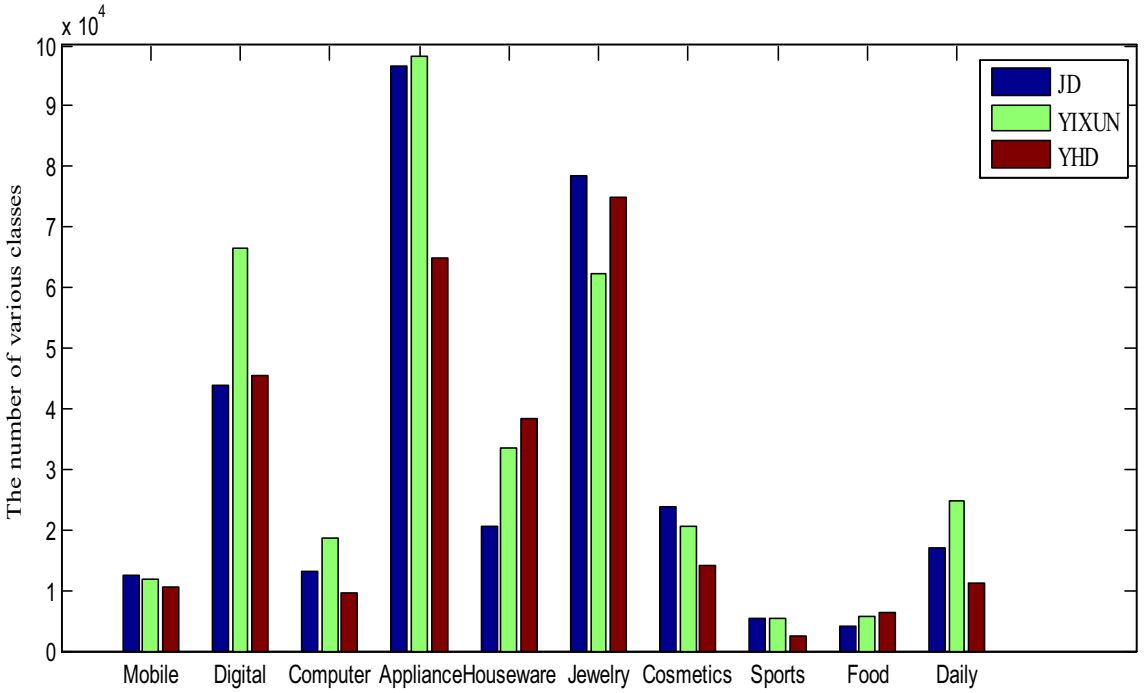

Fig. 1 The distribution for the number of various classes (Color figure online)

to measure the efficiency of the algorithm. The RT is got as follows.

$$RT = 100{,}000 \times \frac{T_1 + T_2 + T_3 + T_4}{Datasize}$$

6.3 Experimental result analysis

The experiment of this paper is to identify and cluster all the goods in the whole data set. In order to validate the method and experiment of this paper, the other three groups of experiments are constructed. The first groups of experiments with the same hardware environment are taken by the traditional multi-threaded concurrent. The software environment is calculated by using Java and Oracle. The Second groups of experiments are taken by shopping search results comparison. The third groups of experiments are taken by the shopping assistant to compare the results. Figure 2 shows the time consuming of this experiment and first groups of experiments in each phase of the algorithm.

The data in Fig. 2 shows that the efficiency of the Map-Reduce based identification algorithm is better than that of the traditional method in each stage.

Also, Fig. 3 shows that the Map-Reduce algorithm used in this paper is superior to the traditional methods in all aspects of the operational efficiency.

The experimental method is to set up a random group of goods page, in the above two platforms to obtain the same

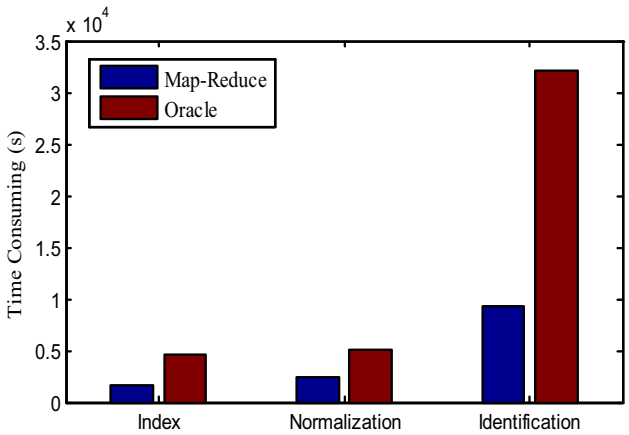

Fig. 2 The time consuming (Color figure online)

goods, and then to verify the results. The experimental results are shown in Fig. 4.

Experimental results show that the proposed method is superior to the shopping search in terms of accuracy, recall rate and comprehensive evaluation index, and the recall rate is slightly lower than the shopping assistant. But the accuracy and comprehensive evaluation index is much higher than the shopping assistant. After analysis, this is because the shopping assistant acquired entities of the same commodity list, there are many similar but not identical goods, so the recall rate can be improved but the accuracy decreases rapidly.

It can be drawn from the experiment, the algorithm as well as in the map reduce environment experiments, with

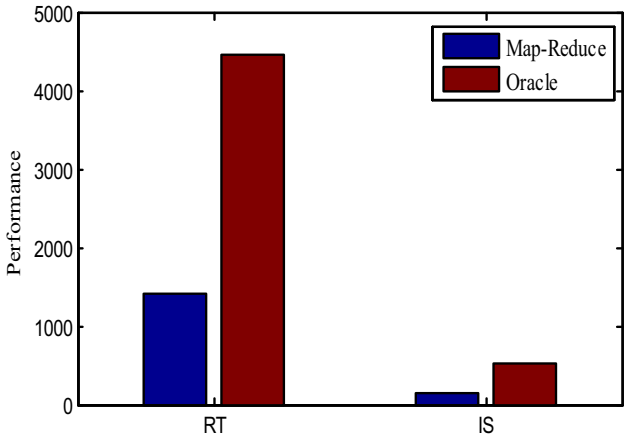

Fig. 3 The time consuming (Color figure online)

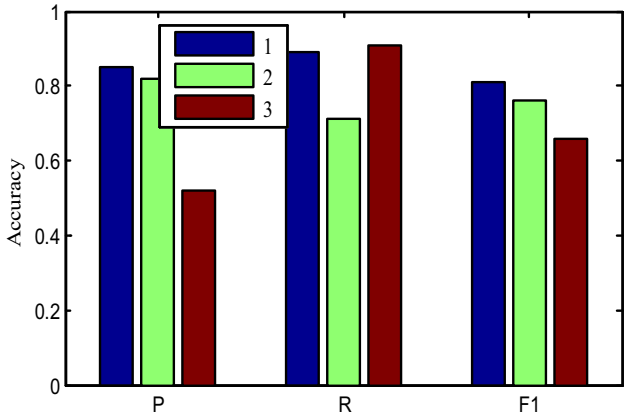

Fig. 4 The accuracy comparison (Color figure online)

high accuracy and efficiency, and the algorithm efficiency and the data set has little effect on the size of, even in larger data sets and more complex data environment algorithm in this paper have good applicability.

7 Conclusion

In this paper, an index model based on commodity attribute/value is established firstly by analyzing the data characteristics of different platforms, and then we construct the global pattern of attribute value of commodity. Experiments are constructed based on the Hadoop platform for the 3 B2C e-commerce data sources. And the traditional methods and products are compared. The experimental results demonstrate the feasibility, accuracy and efficiency of the framework.

Acknowledgements This work was supported in part by the Supported by the construct program of the key discipline in QINGHAI province.

References

1. Fang, Q., Hu, Y., Lv, S., Guo, L., Xiao, L. Hu, Y.: IIRS: A Novel Framework of Identifying Commodity Entities on E-commerce Big Data. Web-Age Information Management. Lecture Notes in Computer Science, vol. 9098, pp. 473–480 (2015)
2. Altwaijry, H., Mehrotra, S., Kalashnikov, D.V.: QuERy: a framework for integrating entity resolution with query processing. Proc. VLDB Endow. **9**(3), 120–131 (2015)
3. Li, L., Wang, H., Gao, H., Li, J.: EIF: A Framework of Effective Entity Identification. Web-Age Information Management. Lecture Notes in Computer Science, vol. 6184, pp. 717–728 (2015)
4. Fan, W., Gao, H., Jia, X., Li, J., Ma, S.: Dynamic constraints for record matching. VLDB J. **20**(4), 495–520 (2011)
5. Deng, D., Li, G., Feng, J., Duan, Y., Gong, Z.: A unified framework for approximate dictionary-based entity extraction. VLDB J. **24**(1), 143–167 (2015)
6. Nuray-Turan, R., Kalashnikov, D.V., Mehrotra, S., Yu, Y.: Attribute and object selection queries on objects with probabilistic attributes. ACM Trans. Database Syst. **37**(1), 3 (2012)
7. Douglas, B., Ronald. F., Kolaitis P.G., Popa L., Wang-Chiew, T.: A declarative framework for linking entities. In: Proceedings of the the 18th International Conference on Database Theory (ICDT 2015), pp. 25–43 (2015)
8. Ribeiro, L.A., Härder, T., Pimenta, F.S.: A cluster-based approach to XML similarity joins. In: Proceedings of the 2009 International Database Engineering & Applications Symposium, ACM, New York, NY, USA, pp. 182–193 (2009)
9. Liao, Z., Zhang, Z., Liu, Y.: Chinese named entity recognition based on hierarchical hybrid model. Pricai Trends Artif. Intell. **6230**, 620–624 (2010)
10. Chen, L., Jiang., Z.: Aircraft detection based on probability model of structural elements. SPIE/COS Photonics Asia International Society for Optics and Photonics, pp. 215–228 (2014)
11. Yang, Y., et al.: Link prediction in brain networks based on a hierarchical random graph model. Tsinghua Sci. Technol. **20**(3), 306–315 (2015)
12. Liang, X.H., et al.: Improved coherent hierarchical culling algorithm based on probability computing model. J. Softw. **20**(6), 1685–1693 (2009)
13. Gaetano, R., et al.: Unsupervised hierarchical image segmentation based on the TS-MRF model and fast Mean-Shift clustering. in: European Signal Processing Conference, IEEE, pp. 1–5 (2008)
14. Nogueira, A., et al.: Modeling self-similar traffic over multiple time scales based on hierarchical Markovian and L-System models. Comput. Commun. **33**(17), S3–S10 (2010)
15. Yan, S.R., et al.: A graph-based comprehensive reputation model: exploiting the social context of opinions to enhance trust in social commerce. Inf. Sci. **318**, 51–72 (2015)
16. Winn, J.K.: Electronic Chattel Paper Under Revised Article 9: Updating the Concept of Embodied Rights for Electronic Commerce. Social Science Electronic Publishing, 3 (2010)
17. Friedman, E.J., Halpern, J.Y., Kash, I.: Efficiency and nash equilibria in a scrip system for P2P networks. In: ACM Conference on Electronic Commerce, pp. 140–149 (2006)
18. Papyrakis, E., Gerlagh, R.: Resource-abundance and economic growth in the U.S. Soc. Sci. Electron. Publ. **51**(4), 1011–1039 (2004)
19. Wang, Y.P.: Commodity information standardization can boost the healthy development of electronic commerce. Information Recording Materials (2016)
20. Najafi, I.: Identify effective factors for improving E-trust of E-transactions in the context of E-commerce and E-government. Int. J. Comput. Trends Technol. **17**(6), 281–299 (2014)

Sheng-fu Zhang got his master degree from Qinghai Nationalities University in 2010. Now, he worked at Qinghai Nationalities University, and his research direction is ERP.