Heterogeneous Knowledge Fusion Algorithm for Minority Cultural Resources based on MapReduce

Ying Liu, Juxiang Zhou[*], and Jianhou Gan

Key Laboratory of Education Informalization for Nationalities, Ministry of Education, Yunnan Normal University, Kunming, 650500, China

Abstract

Aiming at the shortcomings of current knowledge fusion methods and understanding the knowledge fusion algorithm in the big data environment, this paper proposes a heterogeneous knowledge fusion algorithm based on MapReduce for minority cultural resources. In order to improve the performance of the fusion algorithm, the algorithm is used in the similarity calculation. It is improved on the basis of the traditional attribute similarity calculation method. Based on the Hadoop platform and MapReduce framework, the experimental platform is validated. The experimental results show that the proposed MapReduce-based heterogeneous knowledge fusion algorithm for ethnic cultural resources is effective and feasible, both in terms of accuracy and effectiveness.

Keywords: knowledge fusion; ethnic minorities; MapReduce; ontology alignment

1. Introduction

In today's big data era, there are more and more types of data, and the scale of data is increasing. In the vast ocean of data, people are often overwhelmed, and the need to extract knowledge with high utilization value from more and more miscellaneous data has become increasingly urgent. This is a topic of concern for not only the business community but also for academia [1].

In the big data environment, human exploration of knowledge services has not only been limited to traditional information and document services but has also placed more research on user behavior, large amounts of fragmented information, and users. Relationships and massive real-time data, machine data, and unstructured data have been generated by it [2]. The meaning and content of knowledge services will be more related to human needs for knowledge and will gradually shift to knowledge-predictive services, transforming big data into true intelligence.

Knowledge fusion is a cross-disciplinary subject, mainly derived from knowledge science and information fusion. Multi-source heterogeneous cultural resources of ethnic minorities are characterized by diversity, closure, and dispersion. These resources are distributed in different places such as the China National Network, Chinese National Religious Network, and Encyclopedia. In addition, most of these resources are non-structured and difficult to use directly, let alone integrated. The core purpose of heterogeneous knowledge fusion of ethnic cultural resources is to extract hidden, valuable, and undiscovered knowledge from scattered and heterogeneous minority cultural resources. The choice of fusion algorithm will affect the level and connotation of knowledge after knowledge fusion.

The more classical knowledge fusion algorithms are based on the DS theory [3], fuzzy set theory [4], topic map [5], semantic rules [6], and knowledge fusion algorithm. Yan proposed an ontology-based multi-source text knowledge. The fusion algorithm uses the ontology conceptual framework to structure the text knowledge and fuse the conceptual framework [7]. Zhang et al. rationally transplanted the existing data fusion algorithms into knowledge fusion and designed and constructed the big data environment [8]. Shen proposed a multi-objective artificial bee colony algorithm based on

_____

* Corresponding author.
E-mail address: zjuxiang@126.com

evolutionary knowledge fusion [9]. Fang proposed a knowledge fusion method based on key attributes and applied the method to the data integration processing system, improving the rationality of automatic merge processing of heterogeneous data [10]. Tang proposed a knowledge fusion method based on data fusion and an improved algorithm based on the granularity of knowledge sources. The decomposition-merging algorithm was used to obtain the knowledge as reliably as possible [11]. In the study of ethnic minorities, Xia constructed the knowledge fusion algorithm KFA based on the genetic algorithm theory. In order to further optimize the solution space based on the ant colony optimization algorithm, a knowledge fusion algorithm KFA-ACO based on user demand was proposed, and the algorithm was analyzed in combination with the simulation experiment [12].

In summary, scholars at home and abroad are more enthusiastic about the research on knowledge fusion algorithms, but there are still few kinds of literature on the integration of cultural resources of ethnic minorities. Aiming at the problem of slow speed and low accuracy of massive ethnic cultural resource fusion algorithms in the traditional mode, this paper proposes a heterogeneous knowledge fusion algorithm for minority cultural resources based on attributes and rules in the Hadoop platform and MapReduce parallel computing framework. The validity is verified. The algorithm proposes a similarity calculation method based on the improved concept in the similarity calculation. The extracted trinity knowledge of the minority cultural resources is merged by the ontology to identify the contradiction knowledge and redundant knowledge, thus achieving a better fusion result.

2. Knowledge Fusion Algorithm

2.1. Improved Similarity Calculation Method

The key step in knowledge fusion is the calculation of similarity. The traditional semantic similarity calculation mainly includes semantic distance-based, information content-based, attribute-based, and mixed semantic similarity calculation [13]. This paper is based on reference [14]. The similarity algorithm of the conceptual attribute is improved, wherein the similarity sequence in the conceptual attribute is jointly calculated by the word forest and the pinyin. The following sections provide the synonym word forest, the similarity calculation method based on the concept attribute, and the improved based on the concept attribute. The similarity algorithm is also introduced.

2.1.1. Similarity Calculation Method based on Concept Attribute

The conceptual entity X of m object attributes is represented as $X_{attr} = (a_1, a_2, \cdots, a_m)$. The conceptual entity Y of n object attributes is represented as $Y_{attr} = (b_1, b_2, \cdots, b_n)$. The similarity calculation is performed for each attribute of concept X and concept Y, and an $m \times n$ similarity matrix is obtained, wherein the similarity calculation is calculated by using the synonym word forest and pinyin. After obtaining the similarity matrix, it is arranged in reverse order. The top is the maximum value found in this paper. The maximum value is deleted. The above operation is repeated until the matrix is empty. In this way, a maximum similarity sequence $A_1, A_2, \cdots, A_k$ can be obtained, and the similarity calculation algorithm is $\sum_{i=1}^{k} A_i / k$. For the same reason, in the case of data type attributes, a maximum similarity sequence $B_1, B_2, \cdots, B_t$ can be obtained, and the similarity calculation algorithm is $\sum_{j=1}^{t} B_j / t$. Finally, the two similarity calculation algorithms are combined to obtain the attribute similarity calculation algorithm in Equation (1).

$$Sim(X_{attr}, Y_{attr}) = \frac{1}{2}\left(\frac{1}{k}\sum_{i=1}^{k} A_i + \frac{1}{t}\sum_{j=1}^{t} B_j\right) \tag{1}$$

2.1.2. Semantic Similarity Calculation Algorithm based on Improved Concept Attributes

In this paper, based on Equation (1), the entity value, attribute, and entity value are weighted to calculate the concept of entity similarity. Ontology concept attributes are the basis for people to distinguish different ontology concepts, including object type attributes and data type attributes. The object type attribute is used to represent the relationship between the concept nodes, the relationship between the data type attribute concept nodes, and the data type or RDF type in the XML. The similarity calculation method of the concept attribute in Equation (1) is improved, and node X is obtained. The conceptual property similarity calculation of Y is shown in Equation (2).

$$Sim_{PV}(X_{PV}, Y_{PV}) = \left(\frac{2k}{X_{AL}+Y_{AL}}\sum_{i=1}^{k} A_i + \frac{2t}{X_{BL}+Y_{BL}}\sum_{j=1}^{t} B_j\right) \tag{2}$$

$$Sim_P(X_P, Y_P) = \alpha Sim_{PV}(X, Y) + \beta Sim(X_{pN}, Y_{pN}) \tag{3}$$

Where $Sim_{PV}(X_{PV}, Y_{PV})$ is the attribute value similarity, X_{PV} represents the attribute value of concept X, Y_{PV} represents the attribute value of concept Y, and X_{AL} is the concept entity X. Y_{AL} is the length of the maximum similarity sequence of the conceptual entity Y object attribute. Then, the weighted similarity calculation can be obtained using Equation (3), where $Sim_P(X_P, Y_P)$ is the attribute similarity value, X_P represents the attribute of concept X, Y_P represents the attribute of concept Y, X_{pN} is the attribute name of conceptual entity X, and Y_{pN} is the conceptual entity and the attribute name of Y.

Based on the weighted comprehensive semantic similarity calculation of ontology, the concept attribute name, attribute value, and entity concept are introduced when calculating the similarity of minority cultural resource ontology concepts. The weighted similarity is calculated using Equation (4), where $\rho + \sigma = 1$. X_E represents the entity name of concept X, Y_E represents the entity name of concept Y, X_{eN} represents the entity name of concept X, Y_{eN} represents the entity name of concept Y, and $Sim_E(X_E, Y_E)$ indicates the similarity of the concept entities X and Y.

$$Sim_E(X_E, Y_E) = \rho Sim_P(X_P, Y_P) + \sigma Sim(X_{eN}, Y_{eN}) \tag{4}$$

2.2. Ontology Alignment of Ethnic Cultural Resources

Ontology alignment is also called ontology mapping [15] or ontology matching [16]. The process of mapping the elements (concepts, attributes, and relationships) of the ontology is established by comparing the similarity of the ontology. The ontology is used to solve the heterogeneity between the ontology. This article mainly studies entity alignment and attribute alignment. The following is an example from the Baidu Encyclopedia, interactive encyclopedia, and Sogou Encyclopedia Minority.

2.2.1. Resources Entity Alignment Process

The entity alignment judges whether the entities from different places are the same entity. If they represent the same entity, the alignment relationship is established and merged, and the entity attribute information is obtained from the encyclopedia, minority website, and knowledge base to realize the source. The alignment and fusion of different entities, as shown in Figure 1, align and merge the Wa entities from the Baidu Encyclopedia, interactive encyclopedia, and Sogou Encyclopedia to obtain the Wa.
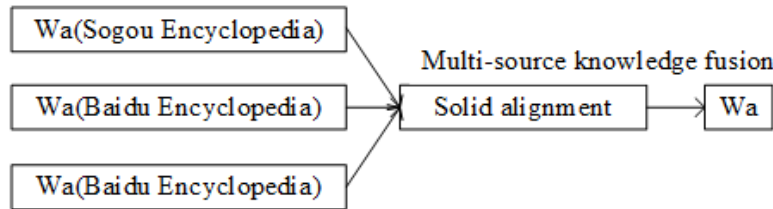

Figure 1. Alignment of the Wa entity

Taking the three encyclopedias' data as an example, Equation (5) shows the process of alignment of the datasets of the three encyclopedia entities.

$$\begin{aligned} Align_{entity}(S_1, S_2, S_3) &= Align_{entity}\left(S_1(s_{11}s_{12}\cdots s_{1m}), S_2(s_{21}s_{22}\cdots s_{2n}), S_3(s_{31}s_{32}\cdots s_{3p})\right) \\ &= \left\{ <S_{1i}, S_{2j}, S_{3k}> | S_{1i}\epsilon S_1, S_{2j}\epsilon S_2, S_{3k}\epsilon S_3 \right\} \end{aligned} \tag{5}$$

Where S_1, S_2, S_3 respectively represent the Baidu Encyclopedia, interactive encyclopedia, Sogou Encyclopedia, and minority key related entities, S_1 has m concept entities, S_2 has n concept entities, and S_3 has p concept entities. For example, the "Hui" of the Sogou Encyclopedia is an entity. The S_{1i} entity comes from S_1, S_{2j} comes from S_2, and S_{3k} comes from S_3, where $i = 1, 2, 3, \cdots, m$; $j = 1, 2, 3, \cdots, n$; $k = 1, 2, 3, \cdots, p$. Because an encyclopedia or website owned by an encyclopedia or website does not necessarily have it, "i, j, k" may not be equal. For example, the Sogou Encyclopedia has an entity that "takes Lutheran Festival", but the interactive encyclopedia does not exist. This paper establishes the relationship between the Baidu Encyclopedia, interactive encyclopedia, and Sogou Encyclopedia minority resources through the ontology alignment algorithm, namely S_{1i}, S_{2j}, S_{3k}. A relationship is established to obtain $<S_{1i}, S_{2j}, S_{3k}>$.

Figure 2 presents is the alignment process of the Baidu Encyclopedia, interactive encyclopedia, and Sogou Encyclopedia entity. In the figure, each circle represents an entity. The dotted line indicates the entities that can be aligned in the three encyclopedias. The data format of this paper exists in the form of triples (entities, attributes, attribute values). Entity alignment finds entities that characterize the same in libraries from different sources. As can be seen from Figure 2, there are still no aligned data in the three libraries. This is because the entity sets contained in different knowledge bases are inconsistent. Entities from the Baidu Encyclopedia do not necessarily exist in the interactive encyclopedia. This is a phenomenon that causes some entities to fail to align.
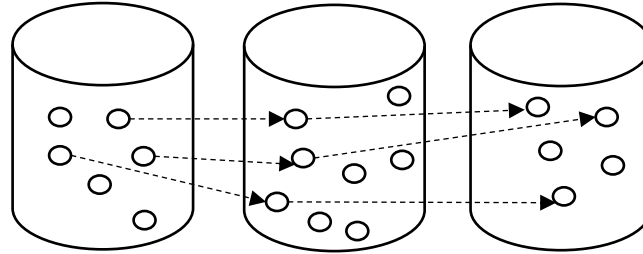

Figure 2. Entity alignment of three knowledge bases

2.2.2. Attribute Alignment Process

After the entity is aligned, the attribute and the attribute value of the entity need to be aligned. The attribute alignment aligns the attribute with the same attribute name or the same meaning. The process of aligning the attribute of the Wa is shown in Figure 3, with the Baidu Encyclopedia, interactive encyclopedia, and Sogou as examples. The encyclopedic entities are composed of two groups: attributes and attribute values. The results of aligning attributes from different sources are stored in rectangular boxes, such as "national festivals" and "populations". In the interactive encyclopedia, religious beliefs, "Chinese name", and "language" correspond to "national festival", the Sogou Encyclopedia corresponds to "festival", and the Baidu Encyclopedia corresponds to "ethnic festivals". The attribute alignment is used to align these three attributes for more accurate knowledge.
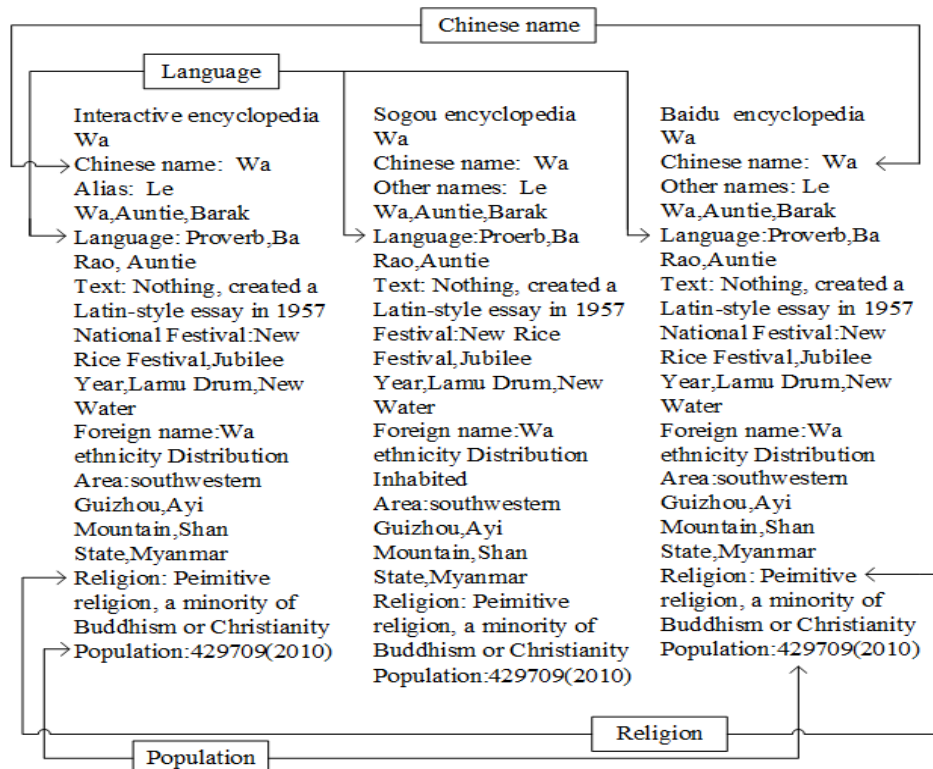

Figure 3. Schematic diagram of the attribute alignment process

Attribute alignment can be expressed by Equation (6).

$$Property = Align(x, y, z) \tag{6}$$

After the attribute is aligned, it is represented by a property. "x, y, z" are the attributes of the entities after the interactive encyclopedia, Sogou Encyclopedia, Baidu Encyclopedia are aligned, and the attributes in the three knowledge bases are traversed to find the attributes with the same name or semantic similarity. Equation (7) gives the m attributes of the entity fusion, and "$Property_m$" is the aligned attribute set.

$$P_T = \left(prop\left(Property_1, Property_2, \cdots, Property_m \right) \right) \tag{7}$$

Figure 4 provides a schematic diagram of the attribute alignment process, including attribute alignment and fusion. Equations (6) and (7) represent attribute alignment and attribute fusion, respectively.

Interactive Encyclopedia, Sogou Encyclopedia and Baidu Encyclopedia properties				
Encyclopedia	$Property_1$	$Property_2$	…	$Property_n$
Interactive Encyclopedia	A_{11}	A_{12}	…	A_{1n}
Sogou Encyclopedia	A_{21}	A_{22}	…	A_{2n}
Baidu Encyclopedia	A_{31}	A_{32}	…	A_{3n}

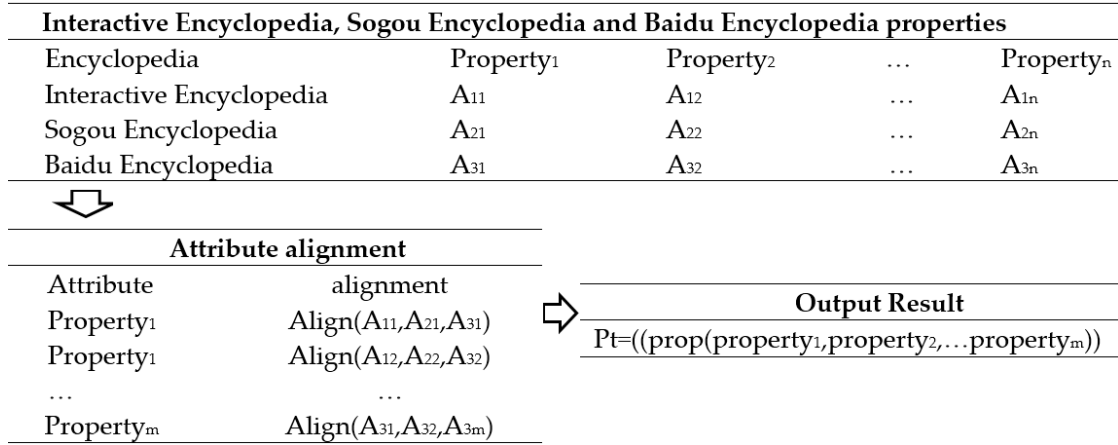| Attribute alignment			Output Result
Attribute	alignment		
$Property_1$	$Align(A_{11}, A_{21}, A_{31})$		$Pt = ((prop(property_1, property_2, \ldots property_m))$
$Property_1$	$Align(A_{12}, A_{22}, A_{32})$		
…	…		
$Property_m$	$Align(A_{31}, A_{32}, A_{3m})$		

Figure 4. Schematic diagram of the attribute alignment process

Based on the Hadoop platform and MapReduce framework, this paper adopts a fusion algorithm based on attributes and fusion rules. The process of knowledge fusion is mainly divided into three parts. Each part is a MapReduce process. In the first part, the Map process divides the extracted knowledge triples. The Reduce step calculates the similarity of the knowledge triple entity and generates the candidate set. In the second part, the Map process divides the candidate set, and the Reduce step performs body alignment according to the similarity. In the third part, the Map process divides the extracted knowledge triples, and the Reduce stage combines and deduplicates the knowledge triples. The specific operation process is shown in Figure 5.
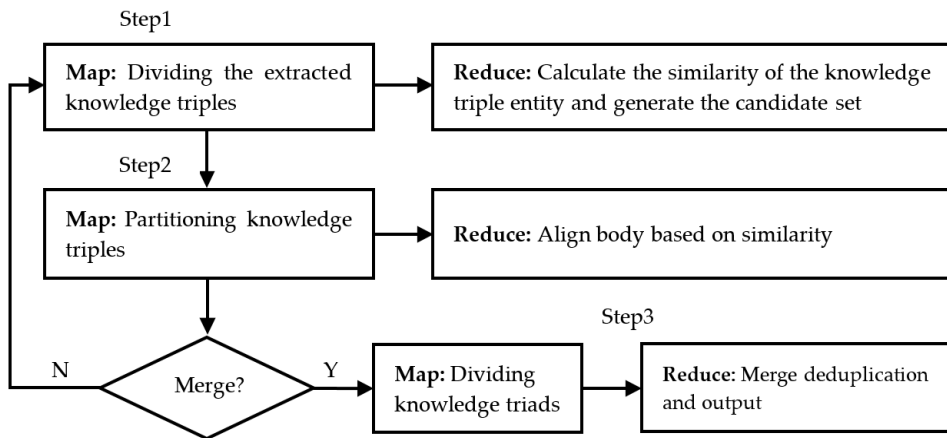
Figure 5. The implementation process of fusion algorithm based on the Hadoop platform

The knowledge of ethnic minorities with different sources is mostly in the form of (entity, attribute, attribute value). Because of its large number, this paper chooses to build the Hadoop platform and integrates minority triad knowledge under the MapReduce framework. Because these knowledge sources are different, there will be some grammatical semantic heterogeneity. The important part of knowledge fusion is the entity link. The entity link includes entity identification and entity disambiguation. Entity recognition has been completed in knowledge extraction. It mainly performs entity

disambiguation and solves the redundancy and conflict problems of the same entity with different sources. The pseudo code of the algorithm is as follows:

Algorithm: Knowledge fusion based on attributes and rules

Input: *Source data and data to be fused*

Output: Fusion results

Step 1 *Get the entity name, attribute and attribute value of a single entity and save it in the entitiesMap*

    Step 1.1 *entry1.getKey().toString()//Get the entity name*

    Step 1.2 *Traversing the attribute collection of entity entities - contains a collection of attribute names and attribute values*

    Step 1.3 *Get the property and property values propertyMap.put(property, propertyValueList)*

    Step 1.4 *Parse out the entity triplet entitiesMap.put(entitys, propertyMap)*

Step 2 *Calculate the similarity of any two strings strOfSimilarity()*

Step 3 *Traverse the source library, find the entity set with the similarity of the name of the currently merged entity greater than ω and the set smaller than ω. If less than ω, it is directly added to the library, and the processing is greater than ω.*

    Step 3.1 *Calculate the similarity between the merged entity name and the entity name of the source library.*

    Step 3.2 *Obtain a source entity set candidateMap with entity name similarity greater than ω as a candidate set*

    Step 3.3 *Calculate similar pairs of candidate entity sets with similarity greater than ω, and sort them according to entity similarity, and store the results in eSimMap.*

Step 4 *Align the attributes of each entity in the set whose entity name is similar to ω*

    Step 4.1 *Get the attribute with the highest similarity in the first value set*

        Double msMaxSimilarity = topSortMap.get(msMaxProperty)

    Step 4.2 *Find attribute pairs that match according to the highest similarity*

        msPropertyTwainMap.put(msMaxPropertyArr[0], msMaxPropertyArr[1])

    Step 4.3 *Remove the attribute from the msPropertyList to the element dit.remove() with the name equal to msMaxProperty*

    Step 4.4 *Remove the attribute from the msPropertyList to the element whose name is equal to msMaxProperty*

    Step 4.5 *Find an entity with the highest similarity from the entity with similarity greater than the ω candidate set, and generate a set of common attribute names of the attributes of the entities to be merged, and store them in unionPNameSet.*

        Step 4.6 *Generate a set of attributes and attribute values of the entities that need to be merged, stored in temPmap*

Step 5 *The final fusion result is stored in sourceMap in sourceMap.put(meName, matchingMap.get(meName))*

Step 6 *Train a model suitable for minority cultural resources and conduct model testing*

---

3. Analysis of Experimental Results

3.1. Experimental Environment

Because the experimental triple data is relatively large, the experimental choice of this paper relies on the Hadoop platform and the MapReduce framework to build a knowledge fusion environment. The method of dividing and conquering a large amount of data and using parallel computing is convenient for completing a large number of tasks in a short time. The cluster environment mainly includes three PCs, all of which are CentOS7 systems. Two of them act as slave nodes, which are responsible for the operation of the DataNode and NodeManager processes. The other acts as the master node and is responsible for the NameNode and the resource manager (ResourceManager). In the operation of the process, the nodes are connected through the local area network. The information of the three hosts is shown in Table 1.

Table 1. MapReduce cluster details

IP	CPU name	Effect
10.10.115.66	Master	NameNode JobTracker
10.10.115.65	Slave1	DataNode TaskTracker
10.10.115.8	Slave2	DataNode TaskTracker

The JDK installed in the MapReduce experimental cluster environment is JDK-8u151-Linux-x64.tar.gz, and the Hadoop version is Hadoop-2.7.3. After configuring SSH without password login and installing the Hadoop cluster, MapReduce is installed on this basis. After the above software is installed on three machines and the corresponding configuration is completed, the MapReduce distributed cluster can be initiated. The process of the master node after the cluster begins is shown in Figure 6. The Hadoop process is viewed using jps.
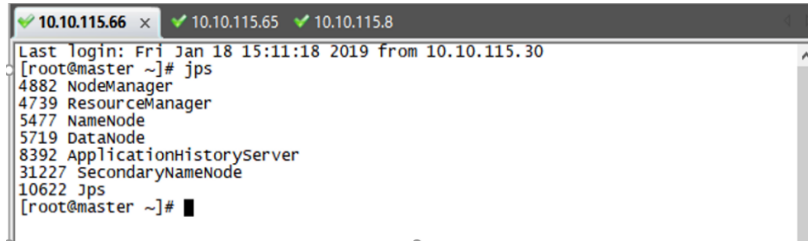
Figure 6. The process diagram after the master node is started

3.2. Experimental Data

This paper uses the knowledge extraction method to obtain the data needed for the experiment and stores it in the form of RDF/XML. The data mainly includes some structured databases, as well as semi-structured minority data extracted from the Baidu Encyclopedia and Sogou Encyclopedia. The existing text resources of the laboratory (107.9733 million words) were extracted and used for fusion experiments. Based on the raw data from different sources above, more than 20,000 sets of triple data were extracted to test the performance of the fusion algorithm.

3.3. Analysis of Experimental Results

3.3.1. Experimental Evaluation Index

The experimental accuracy rate, recall rate, and F-measure are used as evaluation indicators. The specific calculation formulas are shown in Equations (8) to (10).

$$\text{Precision } (P) = \frac{\text{Correctly merging the number of triples}}{\text{The number of fused triples}} \tag{8}$$

$$\text{Recall } (R) = \frac{\text{Correctly merging the number of triples}}{\text{The number of triples that should be merged}} \tag{9}$$

$$\text{F-measure} = \frac{\left((\beta^2+1)\times P\times R\right)}{(P+R)}, \ \beta = 1 \tag{10}$$

3.3.2. Experimental Results

The experiment is mainly divided into three parts. Firstly, after the improved algorithm is trained by the model, a set of parameter values is obtained. At this time, the accuracy of the fusion is maximized. The second part is a comparison between the traditional algorithm and the proposed algorithm. The last part is a comparison between stand-alone and MapReduce distributed under the same scale data.

As shown in Figure 7, when the parameters $\omega = 0.9$ and $\varphi = 0.5$, the accuracy P can reach a maximum of 0.989, where ω is the entity name similarity threshold and φ is the parameter when the word forest and pinyin are used together. When the weight of the pinyin is 0.5, the similarity threshold of the entity name is 0.9, and the accuracy of the fusion is the highest.
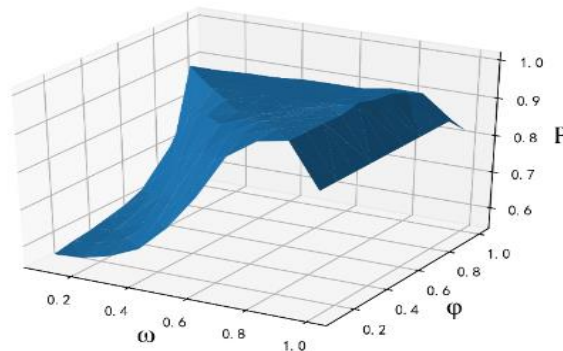

Figure 7. Relationship between accuracy P and threshold ω and φ

Figure 8 compares the traditional concept-based similarity calculation method with the improved algorithm. From Figure 8, it can be seen that the accuracy, recall rate, and F value of the algorithm are improved.
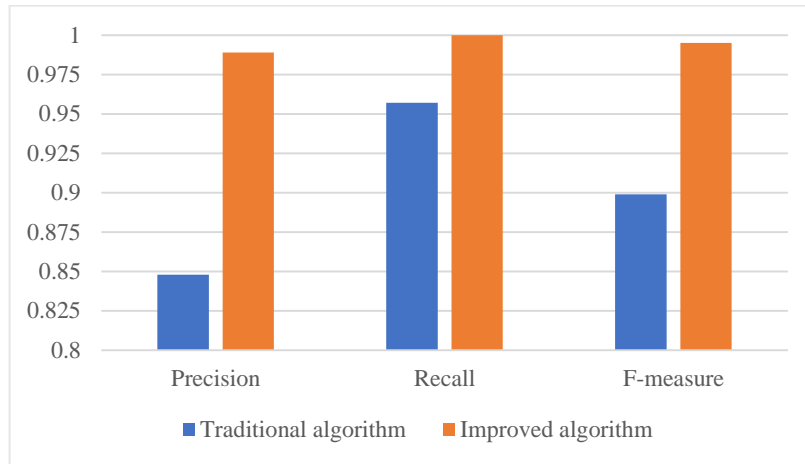

Figure 8. Comparison of traditional algorithm and algorithm fusion results

Figure 9 compares the effectiveness of single-machine and distributed-environment operations in different dataset modes. The algorithm's run-time results in the stand-alone mode and MapReduce parallel mode were improved. By comparing the runtimes of different modes under the same data scale, it can be seen that the time under the MapReduce-based distributed operation is the shortest under the same data scale. It can also be found that as the data size increases, the time-consuming growth under the MapReduce framework becomes smoother, and the computational efficiency advantage is even more evident.
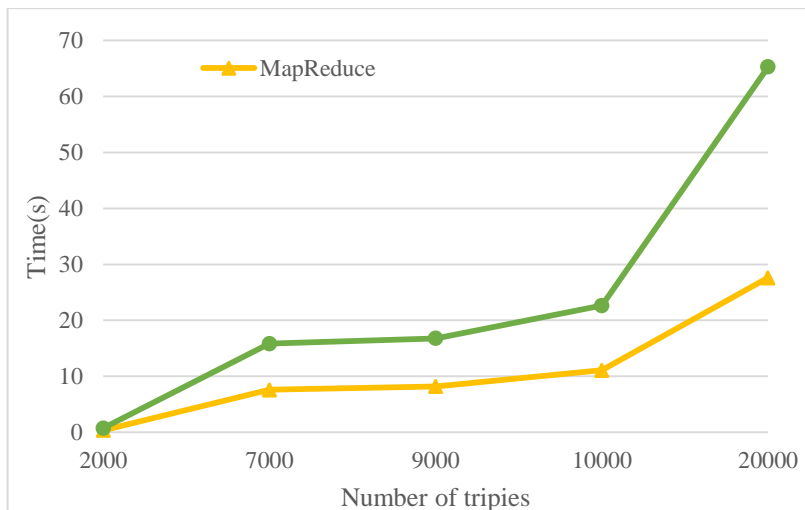

Figure 9. Comparison of single-machine and MapReduce parallel mode under different data scales

4. Conclusions

Based on the MapReduce framework, this paper proposes a heterogeneous knowledge fusion algorithm for ethnic cultural resources. This algorithm improves the traditional concept-based similarity calculation method in the similarity calculation, and it achieves the improvement of this paper under the stand-alone and MapReduce distributed platform. The experimental results show that the algorithm has a certain improvement in fusion accuracy and computational efficiency. At the same time, knowledge reasoning, knowledge recommendation, and more detailed text knowledge extraction can also be carried out, which will be the key direction of future research.

Acknowledgements

References

1. X. N. Su, "Knowledge Organization Theory and Method for Knowledge Service," Science Press, Beijing, March 2014
2. X. B. Tang and W. Wei, "Knowledge Fusion: The Growth Point of Knowledge Services in the Age of Big Data," *Library Science Research*, Vol. 5, pp. 9-14, May 2015
3. P. Valin and P. Djiknavoria, "A Pragmatic Approach for the Use of Dempster-Shafer Theory in Fusing Realistic Sensor Data," *J. of Advances in Info. Fusion*, Vol. 5, No. 1, pp. 32-40, 2010
4. N. Werro, "Fuzzy Classification of Online Customers," Fuzzy Management Methods, pp. 256-263, 2015
5. J. Lu, J. Ma, and G. Zhang, "Theme-based Comprehensive Evaluation in New Product Development using Fuzzy Hierarchical Criteria Group Decision-Making Method," *IEEE Transactions on Industrial Electronics*, Vol. 58, No. 6, pp. 2236-2246, December 2011
6. K. Okoye, A. Tawil, U. Naeem, R. Bashroush, and E. Lamine, "A Semantic Rule-based Approach Supported by Process Mining for Personalised Adaptive Learning," *Procedia Computer Science*, Vol. 37, pp. 203-210, 2014
7. W. Yan and Y. X. Lei, "Analysis of Multi-Source Text Knowledge Fusion Algorithm," *Software Guide*, Vol. 17, No. 5, pp. 62-64, May 2018
8. Y. Zhang, Y. Y. Li, and T. Wei, "Research on Multi-Source Heterogeneous Knowledge Fusion Algorithm under Big Data," *Computer Technology and Development*, Vol. 27, No. 9, pp. 12-16. September 2017
9. Y. X. Shen, J. Chen, and D. H. Wu, "A Multi-Objective Artificial Bee Colony Algorithm based on Evolutionary Knowledge Fusion," *Control and Decision*, Vol. 32, No. 12, pp. 2176-2182, December 2017
10. L. F. Fang, "Research on Heterogeneous Data Integration and Fusion based on Ontology," University of Science and Technology of China, Pattern Recognition and Intelligent Systems, May 2010
11. W. Tang, "Research on Heterogeneous Knowledge Fusion Method in the Big Data Environments," Shaanxi Normal University, July 2016
12. Y. L. Xia, "Research on Knowledge Fusion Algorithm of Multi-Source Heterogeneous Ethnic Information Resources," Yunnan Normal University, June 2014
13. H. Zhang and C. Y. Wen, "Improved Ontology-based Semantic Similarity Calculation," *Computer Engineering and Design*, Vol. 36, No. 8, pp. 39-45, August 2015
14. J. L. Tian and W. Wei, "Calculation Method of Word Similarity based on Synonym Word Forest," *Journal of Jilin University* (*Information Science Edition*), Vol. 28, No. 6, pp. 602-608, November 2010
15. W. Q. Li and X. Sun, "A Semantic Similarity Calculation Method for Ontology Concepts," *Acta Automatica Sinica*, Vol. 38, No. 2, pp. 229-235, February 2012
16. P. Shvaiko and J. Euzenat, "Ten Challenges for Ontology Matching," *Lecture Notes in Computer Science*, Vol. 5332, pp. 1164-1182, May 2008

Ying Liu is a master's student in the Key Laboratory of Educational Informatization for Nationalities (Ministry of Education) at Yunnan Normal University. Her research interests include the informatization of ethnic education resources.

Juxiang Zhou received her Ph.D. in 2019 from Dalian University of Technology. She is currently an assistant research fellow from the Key Laboratory of Educational Informatization of Nationalities (Ministry of Education) at Yunnan Normal University. She specializes in education informationization in digital resource processing.

Jianhou Gan received his Ph.D. in computational metallurgy from Kunming University of Science and Technology in 2016. He is currently the vice director of the Key Laboratory of Educational Informatization for Nationalities at Yunnan Normal University. His research interests include knowledge engineering and educational informatization for nationalities.