

Design of Good Bibimbap Restaurant Recommendation System Using TCA based on BigData

Suk-jin Kim¹, Yong-sung Kim²

Department of Computer Engineering, Chonbuk National University, 664-14 DuckJin-Dong, DuckJin-Gu, Jeonju, Korea.

¹ hanuri00@jbnu.ac.kr, ² yskim@jbnu.ac.kr

² Corresponding Author

Abstract. This paper designs the model of bibimbap restaurant recommending system in BigData(collected from Twitter's data related to bibimbap). We suggests a TCA(Termite Colony Algorithm) k-means algorithm for clustering BigData, TCA algorithm that used the habits of termites. Through the TCA, finding the appropriate initial clustering needed for the K-means clusters is the goal. We recommend good Bibimpop restaurant to user, using "Bibimbap Restaurant database of Korea(2012)" and an "Taste Adjective Dictionary for the Globalization of Korean Food" for Ranking Algorithm.

Keywords: temite colony algorithm, TCA, recommendation system, Bigdata, Bibimbap, k-means

1 Introduction

Recently, although information can be shared through twitter, the characteristics of a twitter cannot be found with simple keyword searches of these twitter with different characteristics. As such, in this paper, to find a good bibimbap restaurant that suits the specific topics that an individual seeks among the countless number of interspersed twitter, the words in each twitter were categorized by their frequency, and the levels of similarities were measured, and, based on these measurements, similar twitter were clustered to propose a solution for finding good bibimbap restaurant characteristics clusters. In addition, within the cluster groups, this paper suggests a similar good bibimbap restaurant recommendation system that makes rankings based on the levels of similarities in accordance with specific topics.

The main goal of this paper is to make rankings per the similarities of specific topics, through a good bibimbap restaurant recommendation system. Therefore, in order to achieve the goal of this paper, clustering needs to be enabled facilitated for finding groups with similar items from BigData. The clustering techniques and the number of clusters need to be determined according to the purpose of clustering.

2 Related Work

2.1 BigData

Gartner is defined that Bigdata is "In real time a large amount of data would constantly come in many forms"[1]. IBM was defined that Bigdata is a vast scale (Volume), types of diversity (Variety), data processing threw (Velocity). Oracle(2012) on the Value of IBM's 3V is defined properties to be added to the 4V.[2]

BigData analytics engine to be open source R and Mahout. R is a programming language for statistical computation and visualization development environment and provides the basic statistics from mining to implement and take advantage of Bigdata is possible through this package. Also, java, C, Phthon programming languages and can also be connected such as a portable, multiple OS environments increases processing. R is Hadoop environments through RHIFE possible treatment. Oracle, IBM's Netezza, SAP R engine, and even the BigData analysis is adopted, the representative in the R statistical software SPSS or SAS interfaces for interworking with offer.

This paper presents data collected from the big stores, statistics, and graphical capabilities of R and the underlying open-source programs built-in functions, such as through the implementation of the package.



Fig. 1. Representation of the properties od BigData

2.2 Clustering

Clustering is classifying without the prior knowledge of the samples that are to be classified in an unsupervised learning fashion based on the similarity levels; groups of close samples gathered are also referred to as clusters; the process of grouping is referred to as clustering. The similarity levels, which are the basis for clusters, use the Minkowski distance, Mahalanobis distance, Lance-Williams distance, Hamming distance, etc [3].

Clustering techniques [4] were classified into hierarchy and partitioning algorithm. [5] was classified into hierarchical clustering, partitioning clustering, neural networks, and statistical search technique. First, as a hierarchical clustering technique,

dendrogram shows cluster results in hierarchy; second, as partitioning clustering technique, an operation that assigns each sample to clusters. Third, as neural network, there are SOM and ART techniques; fourth, as statistical search technique, there are simulated annealing, genetic algorithm, etc [4] [5].

The hybrid method that was created by appropriately using these 4 methods is recently being published as studies. Most of all, the right clustering method and the number of clusters that are appropriate for the characteristics of clusters are important for clusters; it helps to understand clusters if clusters can be well expressed visually and classified.

3 Design for a Good Bibimbap Restaurant recommendation system

3.1 Behavioral habits of termites

In this paper, the Termite Colony Algorithm (TCA) that uses a good Bibimbap restaurant search behavior of termite colony that corresponds to the colony behavior algorithm is proposed. Additionally, the behavioral habits of termite colony are a good Bibimbap restaurant search behavior studied in [6], and based on this, a cluster initialization in accordance with sample density is obtained through a termite a good Bibimbap restaurant search simulation.

Fig. 2 is the tunneling results obtained in a termite tunneling experiment after 1,000 Formosan subterranean termites and 100 soldier ants were added to the center of the grid, after a 0.2cm height 105x105 grid was filled with sand [6].

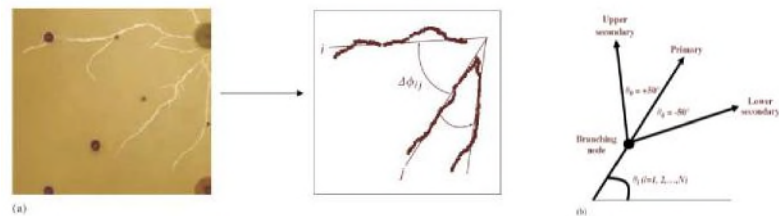


Fig. 2. Termite tunneling results

In Fig. 2, (a) shows the actual experimental results; measurement of the Primary tunnel on the right side showed that the angle measurement of Primary tunnel was ; the $53.1 \pm 8.5^\circ (\theta_k = k \Delta \phi)$ number of tunnels (N) of the Primary tunnel appeared as $6.78 \pm 1.01 (\Delta \phi = 360/N)$

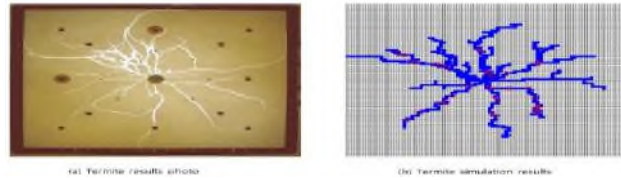


Fig. 3. Termite result photo and simulation results

In Fig. 3 (a) is the actual experimental results, and Fig. 3 (b) is results of the simulation. The behavioral habits of termites, in a two-dimensional grid, achieve the most optimal search path in finding randomly-place a good bibimbap restaurant in the grid. The number of primary tunnel was about six and the angle between each tunnel was about 60 degrees. These structures indicate a good bibimbap restaurant search structure of termite colony habits that has a minimum area of overlap [5].

3.2 Termite Colony Algorithm

TCA is designed based on the probability value of termite's behavior of action. Through the TCA, finding the appropriate initial clustering needed for the K-means clusters is the goal. Through the termite searches, the sample density in the cluster can be estimated, and the estimated initialization that matches the density affects the performance of clustering

3.3 Similar twitter Search Algorithm

The client receives query with words in the web environment. Applicable twitter are found in the word table. In other words, by using the primary key value of applicable a good Bibimbap restaurant, cluster group list that applies to the cluster number is obtained from BigData. The distance between twitters of applicable group is calculated, and raking is made in the order of closest distance twitters.

3.4 Ranking and Recommendation

Fig. 5 shows our system, BRRS(Bibimbap Restaurant Recommendation System). We find restaurant name from the "Bibimbap Restaurant database of Korea(2012)" to make TCA DB, and using "Taste Adjective Dictionary for the Globalization of Korean Food" for weight calculations. [Fig. 5] Show our objective system.

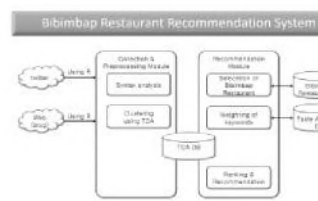
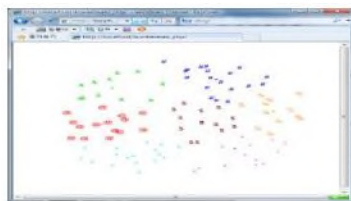


Fig. 4.
TCA
cluster
screen
Recommendation
System)

4

Conclusions

The existing twitter recommendation systems have limitations in terms of finding twitter for the wanted interested subject based on popularity or recommendation types. In addition, even with the search of keywords, the disadvantage exists in that the search is limited to twitter with many documents.

In this paper, a good bibimbap restaurant recommendation system based on BigData level by finding twitters that have similar characteristics as the keywords was proposed. The proposed system is a good bibimbap restaurant recommendation system that ranks the specific groups of the twitters to be searched by clustering based on similarity level of specific words of twitters. For better clusters, TCA k-means algorithm that used the habits of termites was proposed. Based on the experimental evaluation of the proposed algorithm, the following achievements were attained.

A future paper challenge: the acquired twitter were of English language good bibimbap restaurant data, therefore, for acquiring words in the case of Korean language, analyses of morpheme are necessary. Since the 2-dimensional data compression to see cluster forms resulted in a loss of data, minimize data loss by clustering in 3-dimensional data forms, and studies would need to be undertaken for methods that express cluster forms in an easy-to-see fashion.

References

1. Smith, T. F., Waterman, M. S.: Identification of Common Molecular Subsequences. *J. Mol. Biol.*, vol. 147, pp. 195-197, (1981).
2. May, P., Ehrlich, H. C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) *Euro-Par 2006. LNCS*, vol. 4128, pp. 1148–1158. Springer, Heidelberg (2006).
3. Foster, I., Kesselman, C.: *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, San Francisco (1999).
4. Garter, <http://www.gartner.com/it-glossary/big-data/>.

5. Oracle (2012). Oracle: Big Data for the Enterprise, Oracle White Paper, <http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf>
6. Xu R. and Wunsch II D.: Survey of Clustering Algorithms. IEEE Transactions on Neural Networks, vol. 16, no. 3, pp. 645-678, May 2005.
7. Su, N.-Y., Stith, B.M., Puche, H., Bardunias, P., 2004. Characterization of Tunneling Geometry of Subterranean Termites (Isoptera: Rhinotermitidae) by Computer Simulation. Sociobiology, vol. 44, no. 3, pp. 471-483.
8. Lee S.-H., Su N.-Y. and Bardunias P.: Optimal Length Distribution of Termite Tunnel Branches for Efficient Food Search and Resource Transportation. Biosystems, vol. 90, pp. 802–807, (2007).
9. Segaran T., Programming Collective Intelligence, O'Reilly, pp. 77-80, (2007).
10. <http://kiwitobes.com/clusters/feedlist.txt>.
11. <http://www.feedparser.org>.
12. Segaran T., Programming Collective Intelligence, O'Reilly, pp. 77-80, (2007).
13. Economy Censu 2011, Statistics Korea. <http://www.kostat.go.kr>.