International Journal of Engineering & Technology, 7 (3.33) (2018) 51-53

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



A Study on the Clustering of Bidding Company Trends Using Machine Learning Based G2B Data

MinSun Kim[1], EunSoo Choi[2]*, Min Soo Kang

[1,2] Department of Medical IT Marketing, Eulji University, Korea[3]
[3]Department of Medical IT, Eulji University, Korea
**Corresponding author E-mail: miic@eulji.ac.kr*

Abstract

KONEPS is the National Comprehensive Electronic Procurement System of the Public Procurement Service. If KONEPS can know the bidding possibility and trend before bidding, it will be more efficient for companies to bid. In this paper, we used in the experiment was the data of "Progress Bidding Classification" of the Procurement Information Open Portal. And preprocessing process was performed to facilitate prediction model learning. Prior to learning, preprocessed 1,158 data sets were normalized to match the range of data or to make the distribution similar. After normalization we select the number of cluster. As a result of K-Means Clustering, Biddropping is 77 ~ 80%, Budget Allocated is about 2 billion Won(₩), Biddropping is 83 ~ 87%, Budget Allocated is about 1 billion won, bid dropping is 87 ~ 90% Budget Allocated is distributed around 500 million won. And can be confirmed that the cluster is divided based on the number of enterprise 58. Through the results, it is possible to study the tendering trends through the community by learning the prediction models of the bidder companies, the number of bidders, and the tendency of the bidding business, and it will help KONEPS to develop the next generation ISP.

Keywords: *Bidding, Clustering, G2B Data, K-Means Clustering, Unsupervised Learning*

1. Introduction

KONEPS (Korea On-line E-Procurement System) is the National Comprehensive Electronic Procurement System of the Public Procurement Service. Procurement services, which have been handled mainly by existing documents, have a great need for electronicization. It was promoted as one of the 11 e-government projects to provide integrated services that jointly utilize a single system of all departments. It is possible to process all the procurement procedures such as bidding announcement, company registration, bidding and winning bidder, contract signing, payment, etc. Bidding information of all the demanding organizations is announced, and the company can participate in any bidding by registering once in KONEPS. Providing services in conjunction with 77 agency systems, it is possible to omit the submission of documents that are repeatedly submitted during bidding and contracting [1].

If KONEPS can know the bidding possibility and trend before bidding, it will be more efficient for companies to bid. In addition, KONEPS plans to develop a next-generation ISP, and plans to create a user-oriented system by applying cloud, block chain, AI, and big data technology [2].

In this paper, the tendency of bidding is examined by learning the prediction model of bidder's bid rate, the number of bidders, and the tendering rate of bidding business through K-Means clustering algorithm.

2. A Study on Clustering of Bidding Company Trends Using Machine Learning Based G2B Data

2.1. Clustering

The cluster model can be roughly divided into two models in terms of clustering algorithm. The first is a flat or partition based clustering, and the second is a hierarchical clustering model[3]. K-Means, K-Medoids, and DBSCAN are representative of partition-based clustering models. Hierarchical clustering are distinguished by agglomerative clustering and dividing clustering.

2.2. K-Means Clustering

K-Means Clustering is an algorithm that groups given data into clusters, minimizing the variance of distance between clusters. K-Means is one of the simplest unsupervised learning algorithm that clustering unlabeled data. Since K-Means does not know how many clusters exist in unlabeled data, the number of clusters to be classified is determined in advance.

The K-Means algorithm proceeds in the following order [5].

First, after determining the number of clusters (k = n), an arbitrary center n is set. Second, All the data calculates the distances to the center of n, and sets the nearest center as its cluster center. Third, the average of the coordinate values of the learning data for each cluster is calculated and set as a new center. Forth, repeat steps 2 and 3 until the cluster of objects belonging to each data does not

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

change. At last, if there is no case in which the cluster to which the learning data belongs, the learning is completed.

3. Experiment

3.1. Data Preprocessing

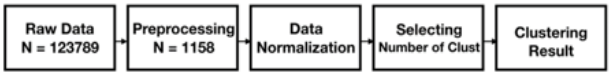


Fig. 1: Prediction Model Training Process

Fig 1 shown Prediction Model Training Process. Data preprocessing, data normalization, selecting number of cluster, and clustering results are presented in order of prediction model process. The data used in the experiment was the data of "Progress Bidding Classification" of the Procurement Information Open Portal [6]. A preprocessing was performed to facilitate prediction model learning. There are a total of 1158 data to make the prediction model through preprocessing. The bid contract method was limited to general bid based on data preprocessing, and the successful bidder rate was 72 ~ 100%. Prior to learning, preprocessed data sets were normalized to match the range of data or to make the distribution similar [7].

3.2. Experiment Environment

We used the R Studio tool for experiments [8].
The number of clusters were selected using the nbClust library to select the number of clusters [9]. The K-Means algorithm is applied based on the number of selected clusters.

3.3. Experiment Result

Hubert and Dindex methods were used to select the number of clusters [10-11]. The number of clusters was selected through the majority rule.

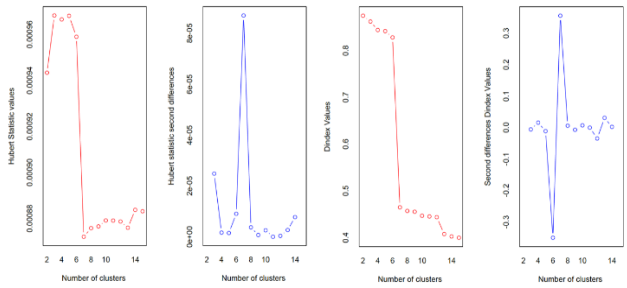


Fig. 2: Budge Allocated & Bid dropping Optimal number of clusters for two attributes

Figure 2 shows that the optimal number of clusters K for the two properties of Budget Allocated and Bid dropping is 7.

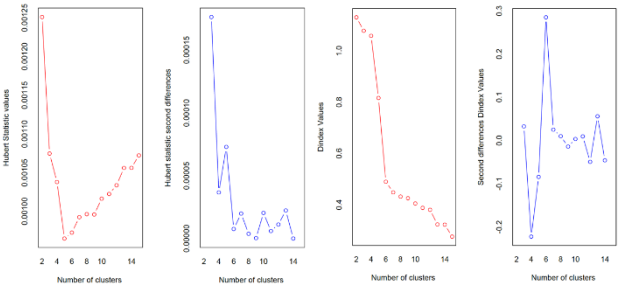
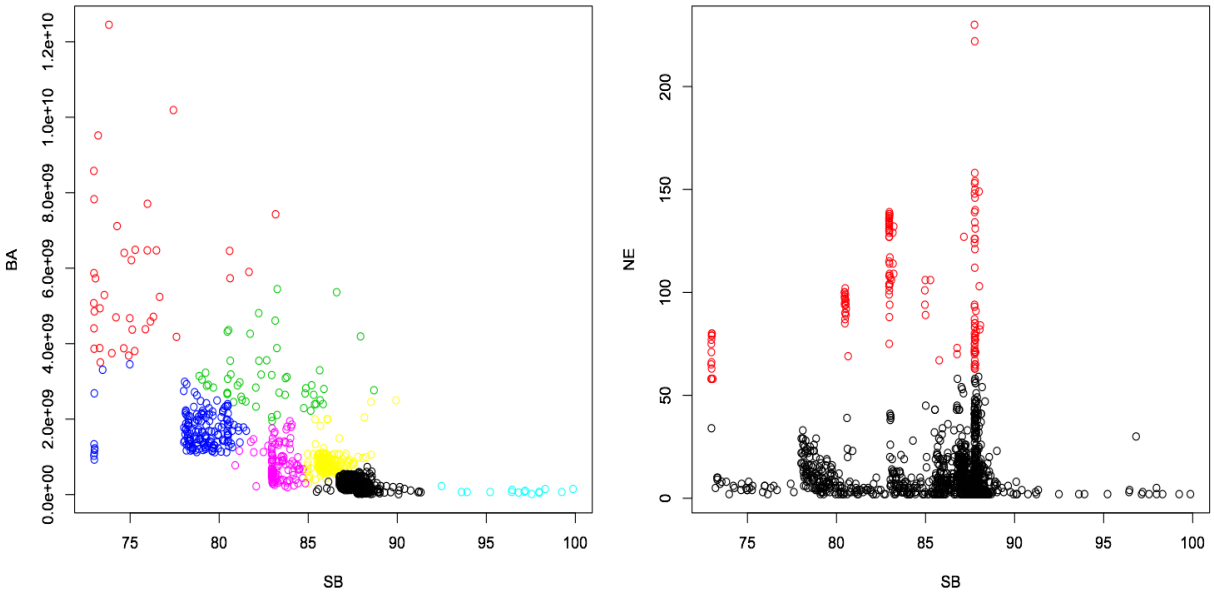


Fig. 3: Number of enterprise & Bid dropping Optimal number of clusters for two attributes

Figure 3 shows that the number of optimal clusters K for the two properties of Number of enterprises and Bid dropping is 2.

Figure 4 shows the results of learning the number of clusters optimized by the Hubert and Dindex methods. Figure (a) shows, Bid dropping is 77 ~ 80%, Budget Allocated is about 2 billion won, Bid dropping is 83 ~ 87%, Budget Allocated is about 1 billion won, Biddropping is 87 ~ 90% Budget Allocated is distributed around 500 million won. Figure (b) shows, it can be confirmed that the cluster is divided based on the number of enterprise 58.



(**a**) Budget Allocated and Bid dropping Clustering Result (K=7)

(**b**) Number of enterprise and Bid dropping Clustering Result (K=2)

Fig. 4: K-Means Clustering Result

4. Conclusion

KONEPS is the National Comprehensive Electronic Procurement System of the Public Procurement Service. It is possible to process all the procurement procedures such as bidding announcement, company registration, bidding and winning bidder, contract signing, payment, etc. Bidding information of all the demanding organizations is announced, and the company can participate in any bidding by registering once in KONEPS. If KONEPS can know the bidding possibility and trend before bidding, it will be more efficient for companies to bid. In this paper, we used in the experiment was the data of "Progress Bidding Classification" of the Procurement Information Open Portal. And preprocessing process was performed to facilitate prediction model learning. Prior to learning, preprocessed data sets were normalized to match the range of data or to make the distribution similar. After normalization, Hubert and Dindex methods were used to select the number of clusters.

As a result of K-Means Clustering, Biddropping is 77 ~ 80%, Budget Allocated is about 2 billion won, Bid dropping is 83 ~ 87%, Budget Allocated is about 1 billion won, Bid dropping is 87 ~ 90% Budget Allocated is distributed around 500 million won. And can be confirmed that the cluster is divided based on the number of enterprise 58.

Through the results, it is possible to study the tendering trends through the community by learning the prediction models of the bidder companies, the number of bidders, and the tendency of the bidding business, and it will help KONEPS to develop the next generation ISP.

References

[1] http://www.g2b.go.kr/

[2] http://www.zdnet.co.kr/news/news_view.asp?artice_id=20180329092725&lo=zv41

[3] James, Gareth, et al. An introduction to statistical learning. Vol. 112. New York: springer, 2013.

[4] MinSoo Kang, EunSoo Choi, “Start Machine Learning with MicroSoft AZURE ML ”, Hanti Media, 2018.03

[5] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1), 100-108.

[6] http://data.g2b.go.kr:8275/pt/main/index.do

[7] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

[8] Team, R. Core. "R language definition." Vienna, Austria: R foundation for statistical computing (2000).

[9] Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2012). NbClust Package: finding the relevant number of clusters in a dataset. UseR! 2012.

[10] Hubert LJ, Arabie P (1985). \Comparing Partitions." Journal of Classi_cation, 2(1), 193-218

[11] Lebart L, Morineau A, Piron M (2000). Statistique Exploratoire Multidimensionnelle. Dunod, Paris.