ICCEPM 2020
The 8th International Conference on Construction Engineering and Project Management
Dec. 7-8, 2020, Hong Kong SAR

Development of a Collaboration Recommendation Model between Global Consulting Firms using Link Prediction

Young-su Yu[1], Bon-sang Koo[2]*

[1] *Graduate student, Department of Civil Engineering, Seoult National University of Science and Technology, Korea,* E-mail address: youngsu@seoultech.ac.kr

[2] *Ph.D, Professor, Department of Civil Engineering, Seoult National University of Science and Technology, Korea,* E-mail address: bonsang@seoultech.ac.kr

Abstract: Global construction and engineering consulting (E&C) firms are actively seeking entry into overseas markets based on loan projects from multilateral development banks to provide a basis for entry into overseas markets and sustainable growth. Bids on these projects are competitive between global top firms in terms of the technical level and price due to the limited number of projects; thus, developing a successful partnership to complement competence has become an essential element to win bids. In this regard, many studies have analyzed enterprises through characteristic analyses or the derivation of influential factors from the past social networks based on social network analysis (SNA). However, few studies have been conducted to reflect the process of changes to analyze collaborative relationships. Thus, this study aims to identify dynamic changes in past social networks and develop a model that can predict changes in the relationships between E&C firms based on similarities or differences between firms, presenting a methodology to target firms for appropriate collaboration. The analysis results demonstrate that the sensitivity of the developed prediction model was 70.26%, which could accurately predict 163 out of 232 actual cooperative cases.

Key words: international cooperative strategies, World Bank ODA, link-prediction, extreme gradient boosting

1. INTRODUCTION

The construction orders for infrastructure have increased, focusing on rapid urbanization and an increase in official development assistance (ODA) investment in developing countries [1]. Based on these circumstances, global engineering and construction consulting (E&C) firms are actively searching for entry to overseas markets and have participated in multilateral development bank (MDB) ODA projects as a means to enter major overseas markets.

However, because bids on MDB ODA projects are competitive with leading global firms, engaging in a consortium or successful partnership with global and local firms that hold various project execution experience is more advantageous for winning projects [2].

In this regard, several studies have analyzed features of past social networks or firms with a high topology in a social network using social network analysis (SNA) for recommending appropriate business partner candidates. However, few studies have been conducted to reflect the process of changes to analyze collaborative relationships with E&C firms. In addition, previously conducted SNA-focused studies were node-oriented analyses, which were focused only on the topology of the firms, and few studies have analyzed the differences or similarities between firms.

This study aims to develop a predictive model of the collaborative relationship reflecting the dynamic changes of the past network based on the successful bid data of the ODA E&C projects ordered by the World Bank. To this end, a 'link prediction' technique was applied that predicts future networks based on link representing relationships between nodes, rather than focusing on the nodes that represent each firms.

2. RESEARCH BACKGROUND

2.1. Social network analysis

Social network analysis (SNA) is a methodology that analyzes the relationship in social network based on a graph of actors in point (node) and their relationships in line (link). In particular, it has the advantage of being able to visualize potential structural types among actors. In addition, the SNA can determine the main actors in the network by using the centrality indicator, and in this study, four centrality indicators (degree centrality, closeness centrality, betweenness centrality, eigen-vector centrality) were used as variables for machine learning.

As an example of SNA application in construction industry, [3] used SNA to analyze the process of changes in cooperative networks within construction projects. In addition, [2] developed a cooperative relationship strategy by utilizing SNA centrality indicators as machine learning variables. As such, SNA is useful in deriving the topology and role of actors in the network. However, there is a limitation in analyzing changes in organic networks, which has recently led to the emergence of link-prediction techniques to reflect organic changes.

2.2. Link-prediction

Link-prediction predicts links that are newly added or removed in the future network based on the current given network, which is used to analyze dynamic changes in the network. This technique is based on the edge or link that represents a relationship between nodes rather than the network centrality indicator of nodes that is primarily used in SNA. The link-prediction analysis methodology is divided into node-based, topology-based, and learning-based methods. This study utilized the learning-based method and used the network topology presented in Table 1 and the difference between nodes for variables for learning.

As an example of link-prediction, [4] explored key factors in building inter-enterprise partnerships by developing a link-prediction system based on Support Vector Machine (SVM). In addition, [5] used link-prediction to analyze manufacturing-service convergence to predict the emerging technology convergence relationship. As such, link-prediction has an advantage in predicting new relationships to be formed through analysis of past network changes. However, studies that have attempted link-prediction in the existing construction industry have been found to be very rare.

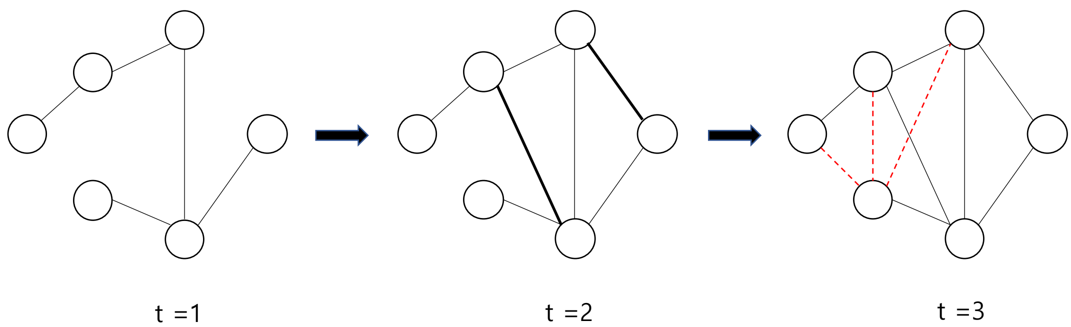

Figure 1. Link-prediction

Table 1. Network topology variables and definitions.

No.	Topology	Abbreviation
1	Preferential Attachment	PA
2	Adamic/Adar	AA

No.	Topology	Abbreviation
3	Resource Allocation	RA
4	Common Neighbors	CN
5	Sorensen Index	SI
6	Salton Cosine Similarity	SC
7	Leicht-Holme Nerman	LHN
8	Hub Promoted	HP
9	Hub Depressed	HD
10	Jaccard's Coefficient	JC

3. LINK-PREDICTION LEARNING MODEL

This study targeted the World Bank's ODA consulting project in three Asian countries (Vietnam, Indonesia and Bangladesh). The data used 'World Bank Finance', an open database provided by the World Bank, and the scope was limited to data that participated in the project with cooperative form because the main purpose of this study was to explore network forming factors. There are 182 cases from 273 companies that participated 471 times in the form of cooperation, and the data were split into three periods based on the number of project participation by country, such as Table 2, in order to establish a learning model that reflects the time flow.

Table 2. Period division based on the number of project participation by country

Country		Period 1	Period 2	Period 3
Vietnam	Year	2003–2010	2011-2014	2014-2017
	Project Participation	77	76	60
	Enterprises	51	46	49
Indonesia	Year	2003-2005	2006-2011	2012-2017
	Project Participation	61	66	59
	Enterprises	41	56	42
Bangladesh	Year	2003-2009	2010-2012	2013-2017
	Project Participation	29	20	23
	Enterprises	19	17	21

Then, data has been transformed into the link unit, which is the basic unit of the link- prediction learning model. In other words, the link data indicates whether individual nodes (companies) are cooperating (0 and 1) and as a result, it has been converted to 37,128 units, or ${}_{273}C_2$, a combination of 273 companies.

In order to establish a link-prediction learning model, 10 variables based on network topology presented in Table 1 and 10 variables based on data presented in Table 3 were utilized. The data-based variables are based on the differences in network centrality between individual nodes used in the SNA and the differences in information contained in the World Bank database.

Table 3. Data-based variables

	No.	Varibles
Difference in information (WB database based)	1	Established year difference
	2	Cooperative form (global-local)
	3	Position difference (leader / member)
	4	ODA experience difference
	5	Enterprise size difference
	6	Average contract size difference
Difference in network centrality (SNA based)	7	Degree centrality difference
	8	Closeness centrality difference
	9	Betweenness centrality difference
	10	Eigen-vector centrality difference

As mentioned in Chapter 1, the learning model in this study was intended to reflect changes over time. Thus, as shown in Fig. 3, the model was trained with the input variables of period 1 and target variables of period 2, and the model was evaluated with the input variables of period 2 and target variables of period 3 respectively. The number of cooperation between training data and test data was 134 and 232 out of 37,128 links, respectively.

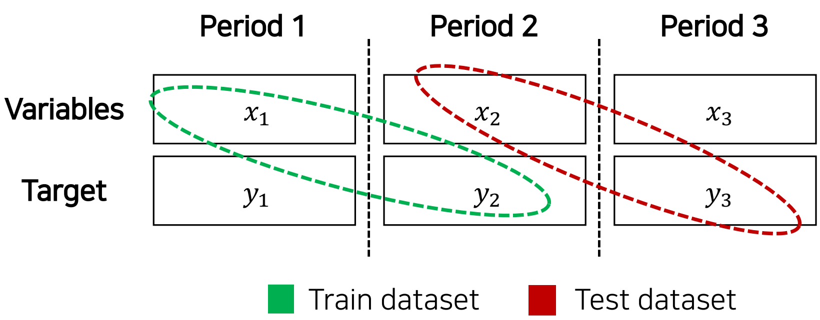

Figure 2. Data combination for link-prediction model

However, when the number of data to be observed is very small compared to the total number of data, an imbalanced data problem occurs in which the model is learned in the direction of increased accuracy, ignoring the actual examples of cooperation. In this study, a small number of data were interpolated using synthetic minority oversampling technology (SMOTE) [6], and the final established training and evaluation data are as shown in Table 4.

Table 4. Final dataset for link-prediction model

Dataset	Linked	Unlinked	Total
Train	36,994	36,994	73,988
Test	232	36,896	37,128

The final data were used to explore the optimal learning model by utilizing three different machine learning models: random forest, multilayer perceptron, extreme gradient boosting (XGBoost). In particular, in this study, sensitivity was selected as the main model evaluation index to predict actual cooperative cases, unlike the general model evaluation method. As a result, the sensitivity of xgboost model was the highest at 70.26% (correctly predicts 163 out of 232 linked cases), which was selected as the final link-prediction model.

The deployed XGBoost model can interpret the portion of individual variables to each forecast cases through log-odds-based probabilities calculations. In particular, this process can be interpreted through visualization through the waterfall chart in 'xgboost expander' package of the R program, and a representative example is given in Figure 3.

The Fig. 3 is a cooperative case between France's EGIS and Indonesia's local company INDEC INNUSA in the management and technical assistance projects of the water resources sector ordered from Indonesia in 2013. In this case, the link-prediction model predicted a 60.12% (log-odds: 0.41), with a large difference in degree centrality of 2 (log-odds: 2.58), and a large ODA experience difference of 3 (log-odds: 1.59) as factors that increase the probability of cooperation. On the other hand, the large difference in the size of the main engagement project (log-odds: -1.62), the five-year difference in general experience in the construction industry (log-odds: -0.83), and the global-local partnership (log-odds: -0.31) contributed to reducing the probability of cooperation.

In summary, the greater the difference in topology (degree centrality) in the network in the case of cooperation between global and local companies, the less likely it is to cooperate with companies with similar general experience in the construction industry. In addition, the wide gap in ODA experience is also a factor that increases the chances of cooperation, and the relationship between global and local companies is a factor that reduces the probability of cooperation.

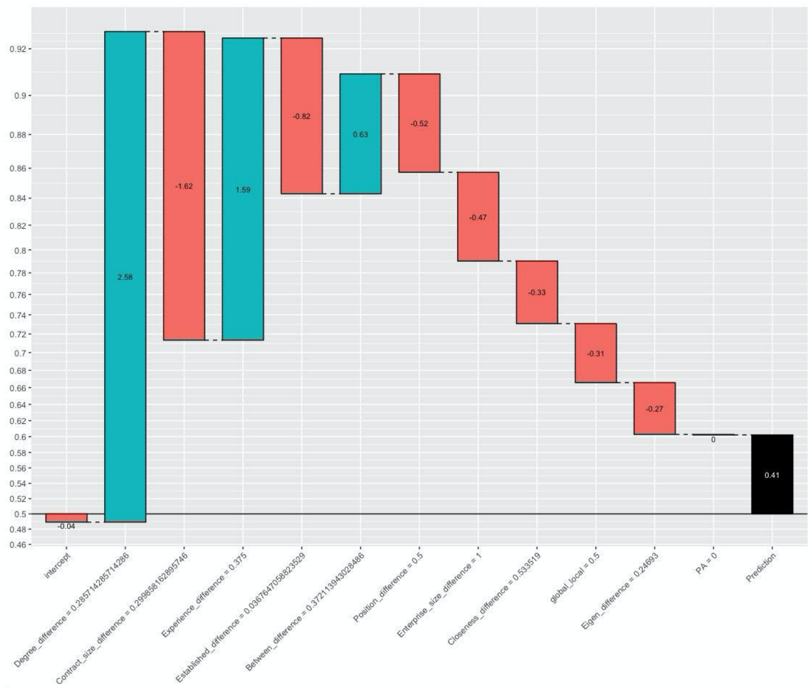

Figure 3. EGIS-INDEC Internusa cooperative case

4. CONCLUSION

This study has developed a link-prediction model with a sensitivity of 70.26% that correctly predicts 163 out of 232 actual cooperative cases, reflecting the dynamic change in the winning company network participating in the World Bank ODA project. In this process, three machine learning algorithms were used to explore suitable learning models for link-prediction, and a methodology was presented to enable E&C firms to explore appropriate business partners. In particular, this model presented the process of deriving the probability of cooperation of individual links, which is expected to provide customized information that recommends specific partners for each firm.

However, other cases predicted by the link-prediction model showed that even if the variables were the same value, the effect on the cooperative factors changed as the values of the other variables changed. This means that detailed forecasting of individual cases is possible, but it is difficult to generalize the effects of each variable, making it difficult to provide universal information. Therefore, additional factor analyses will be conducted in the future to provide consistent information about the variables used to build the model.

ACKNOWLEGEMENTS

This research was supported by a grant [grant number: 20AUDP-B127891-04] from the Architecture & Urban Development Research Program funded by the Ministry of Land, Infrastructure and Transport of the Korean government.

REFERENCES

[1] Global Infrasturcture Hub, “Global Infrastructure Outlook”, CWA update, pp.1-164, 2018.

[2] Lee, J. S., Lee, J. H., Han, S. H., and Kang, S. Y., “Partnering Strategy for Bidding Success in World Bank’s Vietnam Consulting Projects through Country-level Competitive Analysis: A Vietnam Case Study”, vol. 18, no. 2, pp.1021-1028, 2018.

[3] Liu, L., Han, C., and Xu, W., "Evolutionary analysis of the collaboration networks within National Quality Award Projects of China.", International journal of project management, vol. 33, no.3, pp. 599-609, 2015.

[4] Mori, J., Kajikawa, Y., H., and Sakata, "Machine learning approach for finding business partners and building reciprocal relationships." Expert Systems with Applications, vol. 39, no. 12, pp. 10402-10407, 2012.

[5] Seo, H. B., “A Deep Learning based Approach to Prediction of Technological Convergence”, Master thesis, Seoul National University of Science and Technology, 2017, print.

[6] Barua, S., Islam, M. M., and Murase, K., "A novel synthetic minority oversampling technique for imbalanced data set learning.", International Conference on Neural Information Processing. Springer, Berlin, Heidelberg, 2011.