Enhancing Similar Business Group Recommendation through Derivative Criteria and Web Crawling

Min Jeong LEE[1], and In Seop NA[2*]
[1] Department of Computer Science, Chosun University
Gwangju, Republic of Korea
[e-mail: lmjc2008@naver.com]
[2] Division of Culture Contents, Chonnam National University
Yeosu, Republic of Korea
[e-mail: ypencil@hanmail.net]
[*]Corresponding author: In Seop NA

Abstract

Effective recommendation of similar business groups is a critical factor in obtaining market information for companies. In this study, we propose a novel method for enhancing similar business group recommendation by incorporating derivative criteria and web crawling. We use employment announcements, employment incentives, and corporate vocational training information to derive additional criteria for similar business group selection. Web crawling is employed to collect data related to the derived criteria from 'credit jobs' and 'worknet' sites. We compare the efficiency of different datasets and machine learning methods, including XGBoost, LGBM, Adaboost, Linear Regression, K-NN, and SVM. The proposed model extracts derivatives that reflect the financial and scale characteristics of the company, which are then incorporated into a new set of recommendation criteria. Similar business groups are selected using a Euclidean distance-based model. Our experimental results show that the proposed method improves the accuracy of similar business group recommendation. Overall, this study demonstrates the potential of incorporating derivative criteria and web crawling to enhance similar business group recommendation and obtain market information more efficiently.

1. Introduction

Small and medium-sized enterprises (SMEs) in Korea often face difficulties in obtaining the market information necessary for their growth and sustainability [19]. While job seekers benefit from access to customized recruitment services, companies currently lack access to tailored information provision services [20]. Consequently, companies are compelled to navigate multiple providers repeatedly to access information related to job announcements, employment incentives, and corporate vocational training, all of which are vital for their growth and success [21]. Moreover, the information required by companies is often scattered across various websites, resulting in a cumbersome and time-consuming process to obtain the desired information.

To address these challenges and improve the effectiveness of similar business group selection, we propose a novel system that incorporates new classification variables into the existing traditional criteria for classifying similar companies [22]. Our approach utilizes derivative criteria derived from information on employment incentives, corporate vocational training, and corporate recruitment to recommend similar companies that are highly relevant to their specific interests [23]. To enhance the efficiency of market information acquisition, our system leverages web crawling techniques to collect data related to the derived criteria from 'credit jobs' and 'worknet' sites [24].

Our study extensively analyzes and compares diverse datasets and machine learning techniques, including XGBoost, LGBM, Adaboost, and SVM [25]. Additionally, we introduce a cutting-edge Euclidean distance-based model tailored for selecting similar business groups [26]. Through this paper, we demonstrate the efficacy of our proposed system in providing customized and efficient market information to companies, thereby facilitating their growth and enhancing the employment environment [27]. The study highlights the potential of integrating derivative criteria and web crawling to refine similar business group selection and improve market information acquisition, revolutionizing the way SMEs access essential information and surmount information barriers to achieve sustainable growth and success.

2. Machine Learning Algorithm for Business Group Recommendation

A recommendation system is a specialized subset of information filtering systems that places paramount importance on personalization, tailoring recommendations to the unique interests and preferences of individual users [1-6, 19-23].

The early attempts at computer-based recommendations included Grundy, which was a librarian program that grouped users into stereotypes based on an interview and generated recommendations based on hard-coded information [3,7]. Collaborative filtering became a solution for dealing with information overload in the early 1990s, with Tapestry as a manual collaborative filtering system and GroupLens as an automated system. Interest in recommender systems led to the development of systems for various domains such as music, movies, and jokes [3]. Commercial deployment of recommender technology began in the late 1990s, with Amazon.com being the most widely-known application. Recommender technology has been integrated into many e-commerce and online systems to increase sales volume [2,3,7]. Recommender techniques have grown to include content-based approaches, bayesian inference, and case-based reasoning methods, and hybrid recommender systems have emerged. In 2006, the Netflix Prize competition sparked a flurry of activity in academia and hobbyists to improve the state of movie recommendation [3,7]. Although the roots of recommender systems can be traced back to various fields, they emerged as an independent

research area in the mid-1990s. Since the widespread adoption of deep learning, which was introduced by Geoffrey Hinton in ImageNet [8] in 2012, machine learning techniques have been actively utilized in recommendation systems and have shown outstanding performance. The most common formulation of the recommendation problem is to estimate ratings for yet unrated items, which allows recommendations to be made to the user.

There are many machine learning algorithms that are commonly used in recommender systems, including XGBoost [9,16-18], AdaBoost [10,16-18], LightGBM [11], SVM [12,16-18], Linear Regression [13,16-18], K-NN [14,16-18], and SHAP [15]. XGBoost and LightGBM are popular gradient boosting frameworks that have been shown to have high accuracy and scalability, making them effective for handling large datasets. Adaboost is another boosting algorithm that can be useful for ensemble learning, but may not perform as well with high-dimensional data. SVM is a powerful algorithm for classification and regression, but can struggle with large datasets due to its computational complexity. Linear regression is a simple yet effective algorithm for regression tasks, but may not be as robust as other models when dealing with noisy data. K-NN is a non-parametric algorithm that can work well with small datasets, but may not scale well to larger ones. Finally, SHAP is a newer algorithm that can provide insights into feature importance, helping to explain the recommendations made by other algorithms. Overall, the choice of algorithm will depend on the specific needs and constraints of the recommendation system in question, such as the size of the dataset, the type of data being analyzed, and the desired level of accuracy and interpretability.

3. Similar Business Group Recommendation System

3.1 Architecture of Proposed Recommendation System

The system architecture, illustrated in **Fig. 1**, comprises data acquisition, data preprocessing, optimal model selection, feature variable extraction, and modeling of similar business groups.
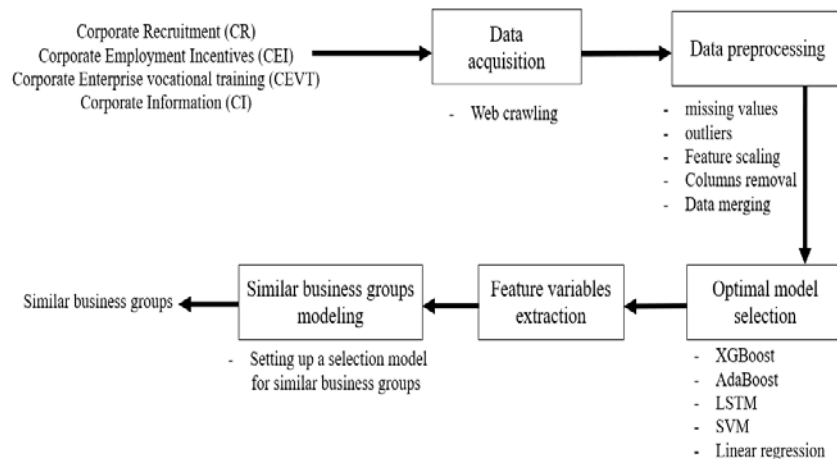

Fig. 1. Architecture of Proposed Recommendation System

Proposed system utilizes various datasets, including Corporate Information (CI) data, Corporate Employment Incentives data (CEI), Corporate Enterprise Vocational Training data (CEVT), Corporate Recruitment data (CR), and 'WORKNET' and 'HRD-NET' open APIs, for data preprocessing. We then apply optimal model selection to extract the most influential

feature variables and incorporate them into the classification criteria of similar companies. By measuring the distance between customer companies and each company or using clustering, our system recommends market information of similar companies based on the newly created classification criteria. This allows our system to distinguish regions and industries from companies seeking market information, offering a more tailored and effective service.

3.2 Data Acquisition

We collected various types of data to construct our dataset, including corporate information (CI) data, Corporate Employment Incentives (CEI) data, Corporate Enterprise vocational training (CEVT) data, and Corporate Recruitment (CR) data. To supplement our dataset, we also crawled recruitment information using the WORKNET's open API and corporate vocational training information using HRD-NET's open API. For some companies with missing values in certain columns such as revenue and number of employees, we utilized the 'KREDIT JOB' service to automatically search for the missing information and collect it through code. With all this data, we constructed a comprehensive dataset for our analysis. We conducted the following procedures to perform web scraping: For 'worknet' and 'hrd net,' we obtained the authentication keys for their open APIs, which allowed us to access the data. We then selectively crawled the necessary data from the accessed information. For 'credit jobs,' we crawled only the required data from the website to fill in the missing values in the corporate information. In all cases, we utilized the BeautifulSoup library, which is a data extraction tool, to parse the HTML tags and extract the relevant data.

3.3 Data Preprocessing

To prepare for model learning, we preprocessed the CI and CR data by excluding private businesses with missing or exceptional values in the IPO (Initial Public Offering) column, resulting in only registered corporations, external auditors, securities markets, general corporations, KONEX, and KOSDAQ markets being included. Additionally, only limited liability companies, limited companies, corporations, general partnerships, and limited partnerships were included in the corporate type column, excluding individual companies. As we are now focusing on classifying similar companies and providing market information, all non-normal data in the corporate and registration status columns were excluded. To reduce the dimension of the financial-related columns for training, we added derivatives that could be generated from the data, resulting in the addition of six variables: total asset turnover, sales operating return, total asset growth, equity capital ratio, ROA, and ROE. These variables were generated using the following equations.

$$TAG = TAEC/TAEP * 100 - 100 \qquad (1)$$

$$ECR = TC/(TL + TC) * 100 \qquad (2)$$

$$TAT = S/ATA \qquad (3)$$

$$SOR = BP/S * 100 \qquad (4)$$

$$ROA = (NP/TA) * 100 \qquad (5)$$

$$ROE = (NP/ASC) * 100 \qquad (6)$$

(1) **TAG**: Total Asset Growth, **TAEC**: Total Assets at the End of the Current term, **TAEP**: Total Assets at the End of the Previous term
(2) **ECR**: Equity Capital Ratio, **TC**: Total Capital, **TL**: Total Liabilities
(3) **TAT**: Total Asset Turnover, **S**: Sales, **ATA**: Average Total Assets
(4) **SOR**: Sales Operating Return, **BP**: Business profits, **S**: Sales
(5) **ROA**: Return On Asset, **NP**: Net Profit, **TA**: Total Assets
(6) **ROE**: Return On Equity, **NP**: Net Profit, **ASC**: Average Self-Capital
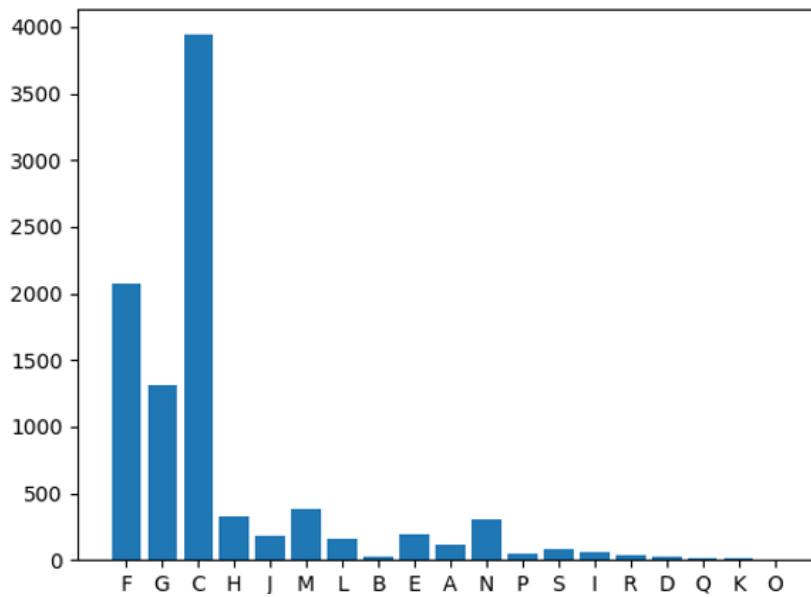
Fig. 2. Distribution of companies by industry

---

1. Employment-related column dimension reduction:
   - Two employment-related variables added: employment rate and retirement rate
2. Distribution of companies by industry bias:
   - **Fig. 2** shows a bias towards specific industries: C (manufacturing), F (construction), G (wholesale and retail)
   - Minority data in remaining industrial codes cannot be accurately learned and predicted
   - To improve service quality, decision to classify similar companies for the three largest industries C, F, G only
3. Feature scaling to unify feature range and unit
   - Prevent machine learning model bias against specific data
4. Data merging
   - CEVT data and CI data merged based on business registration number
   - CEI and CI data merged based on business registration number
   - CR data and CI data merged based on business registration number
5. CR data preprocessing
   - Data less than the minimum wage column removed from the monthly average wage column
   - Columns with missing value ratio of 50% or more removed
   - Outliers removed and feature scaling performed

To reduce the dimension of the employment-related column, we added two employment-related variables: the employment rate and retirement rate. However, **Fig. 2** shows that the distribution of companies by industry is biased towards specific industries (C, F, G) and the remaining industrial codes have relatively small data. This presents a challenge for accurate learning and prediction of these minority data. Therefore, we decided to classify similar companies for the three largest industries (C, F, G) instead of all industrial codes to improve the quality of our services. To prevent machine learning models from being biased against specific data, we standardized the value of each feature into one range size through feature scaling. We merged the CEVT data and CI data based on the business registration number and merged the CEI and CI data based on the same number. We removed the data less than the minimum wage column from the monthly average wage column in the 'WORKNET' employment data. Additionally, we removed columns with a missing value ratio of 50% or more, and other preferential or desired contents, industrial complex codes, employment registration purposes, vacant jobs, reasons for vacant jobs, and employment announcement numbers. Finally, we removed the outliers and performed feature scaling on each data. The processed CR data and CI data were merged based on the business registration number.

3.4 Optimal Model Selection

1. Utilizing various machine learning algorithms with preprocessed data to establish new criteria for classifying similar companies
2. Using CI data as independent variables and upper policy name of employment incentives and KECO code of corporate vocational training as dependent variables for each machine learning model
3. Selection of dependent variable based on the purpose of the recommended service for each information provided
4. Incentive type selected as machine learning target value for employment incentives and type of training for corporate vocational training
5. Difficulty in predicting recruitment data due to multiple factors affecting the recommendation purpose
6. Recruitment information data decided based on company characteristics and selection of new similar company classification criteria
7. Selection of model with highest accuracy among trained machine learning models
8. Extraction of feature variables

We utilized various machine learning algorithms such as XGBoost, LGBM, and Adaboost with preprocessed data to establish new criteria for classifying similar companies by predicting the types of employment incentives and corporate vocational training received by companies. For each machine learning model, CI data (business days, registered corporations, total asset growth rate, equity capital ratio, ROA, ROE, etc.) were used as independent variables, and the upper policy name of employment incentives and the KECO code of corporate vocational training were used as dependent variables. The selection of the dependent variable was based on the purpose of the recommended service for each information provided. Since the purpose of the employment incentives service is to increase the benefit rate, the incentive type was selected as the machine learning target value. Similarly, since the purpose of the corporate vocational training service is to increase the participation rate of corporate vocational training, the type of training was selected as the machine learning target value. However, predicting the recruitment data was challenging as there are multiple factors affecting the recommendation purpose of the recruitment information data. Therefore, the recruitment information data was

decided based on the characteristics of the company and the selection of the new similar company classification criteria. We will select the model with the highest accuracy among the various trained machine learning models to extract feature variables.

3.5 Feature Variables Extraction

Model Selection -> SHAP Visualization -> New Criteria Selection -> Financial Characteristics Consideration

1. Select the model with the highest accuracy:
   - Based on the company's characteristics
   - Best predicts types of employment incentives and KECO code of vocational training
2. Visualize variable influence through SHAP:
   - Extract new criteria for classification of similar business groups
   - Consider financial and scale characteristics
3. Incorporate financial characteristics for similar company classification criteria:
   - Activity
   - Profitability
   - Growth
   - Stability

We selected the model with the highest accuracy to predict the types of employment incentives and the KECO code of the company's vocational training, based on the company's characteristics. Using SHAP, we visualized the influence of variables that affect each prediction in a graph and extracted three new criteria for classifying similar business groups, taking into account financial and scale characteristics. Since CR data cannot be used for supervised learning, we selected the classification criteria based on the company's four financial characteristics: activity, profitability, growth, and stability, instead of extracting variables through SHAP.

3.6 Similar Business Groups Modeling

1. Three models considered: simple group matching, clustering, and Euclidean-based k-ranking
2. 'k' value select for clustering and k-ranking method
3. check the limitations of the simple group matching model
4. Exclusion of the simple group matching model from the selection process
5. Selection of the clustering and Euclidean distance-based k-ranking model as the model for selecting similar companies
Use of cluster variables for CR information
Use of k-ranking model for CEI and CEVT information

We utilized various models, including simple group matching, clustering, and k-ranking based on Euclidean distance to select similar companies. The clustering model divides companies into 'k' clusters and recommends market and business information received by companies in the same cluster, providing a way to classify companies by type. The k-ranking model calculates the distance between new company information and existing companies using Euclidean distance and selects 'k' similar companies based on ascending distance values to recommend business and market information. This method has the advantage of always referring to a constant number of companies, showing high accuracy with increasing data, and

2816
LEE et al.: Enhancing Similar Business Group Recommendation
through Derivative Criteria and Web Crawling

being relatively strong against data noise. The simple group matching model is easy to understand and apply, but may fail to recommend information if no corresponding group exists. Therefore, we excluded this model and chose the clustering and k-ranking model as the selection method for similar companies. The CR information was clustered using the average value to generate cluster variables and select similar companies. For the CEI and CEVT information recommendation services, we used the k-ranking model to select the closest 'k' companies as similar companies. Choosing the appropriate 'k' value is crucial for accurate recommendations.

4. Experiment and Results

4.1 Datasets

To extract similar companies, we utilized various datasets including corporate information, employment incentives, and employment training history data [1]. The CI dataset consisted of 268,044 raw data tuples in 44 items, covering basic corporate information, scale, and financial data. The CR dataset had 158,174 registration raw data tuples in 36 items, while the CEVT dataset included 1,330,025 participation history raw data tuples in 11 items on corporate vocational training. The CEI dataset comprised 2,760,759 beneficiary history raw data tuples, containing 36 items on employment incentives and the beneficiary history of each company. In addition, we gathered data from multiple sources such as 92,380 company data tuples with 59 items from WORKNET open API, 1,207 company data tuples with 25 items from HRD-NET open API, and 18,770 company data tuples with 10 items from KREDIT JOB crawling data, as shown in **Table 1**. Following the preprocessing stage, we were left with a total of 8,418 tuples for CI, 15,543 tuples for CR, 175,224 tuples for CEVT, and 35,211 tuples for CEI.

Table 1. Change of industrial code

Data	Tuples (Raw Data)	Tuples (Source Data)	Items
Corporate Information (CI)	268,044	8,418	44
Corporate Recruitment (CR)	158,174	151,543	36
Corporate Enterprise vocational training (CEVT)	1,330,025	175,224	11
Corporate Employment Incentives (CEI)	2,760,759	35,211	36
WORKNET's open API	92,380	92,380	59
HRD-NET's open API	1,207	1,207	25
'KREDIT JOB' crawling data	18,770	18,770	10

4.2 Experimental Results

We used various machine learning algorithms, including XGBoost, LGBM, and AdaBoost, to predict the types of CEI and CEVT that companies benefited from. The independent variable was the corporate information data, including business days, registered corporations, total asset growth rate, equity capital ratio, ROA, ROE, etc., while the dependent variables were the upper policy name of CEI and the KECO code of CEVT. XGBoost exhibited the highest accuracy with 78.38% in CEI and a relatively high accuracy of 62.56% in CEVT. Therefore, we chose the XGBoost Classifier's results as a model for extracting decision variables for

similar companies, as displayed in **Table 2**. The hyperparameters of XGBoost, which include learning_rate, n_estimators, max_depth, subsample, gamma, objective, and disable_default_eval_metric, are shown in **Table 3**. Any parameters not shown in the table were set to default values.

Table 2. Predicted Accuracy of types of Employment Incentives

Model	Accuracy (CEI)	Accuracy (CEVT)
XGBoost	**78.38**	**62.56**
AdaBoost	55.68	52.9
LGBM	41.72	54.69
SVM	41.83	54.29
Liner Regression	40.92	56.00
KNN	31.72	51.90

Table 3. Hyperparameters of XGBoost

Parameter	Value
learning_rate	0.1
n_estimators	100
max_depth	3
subsample	0.6
gamma	5
objective	'multi:softprob
disable_default_eval_metric	True
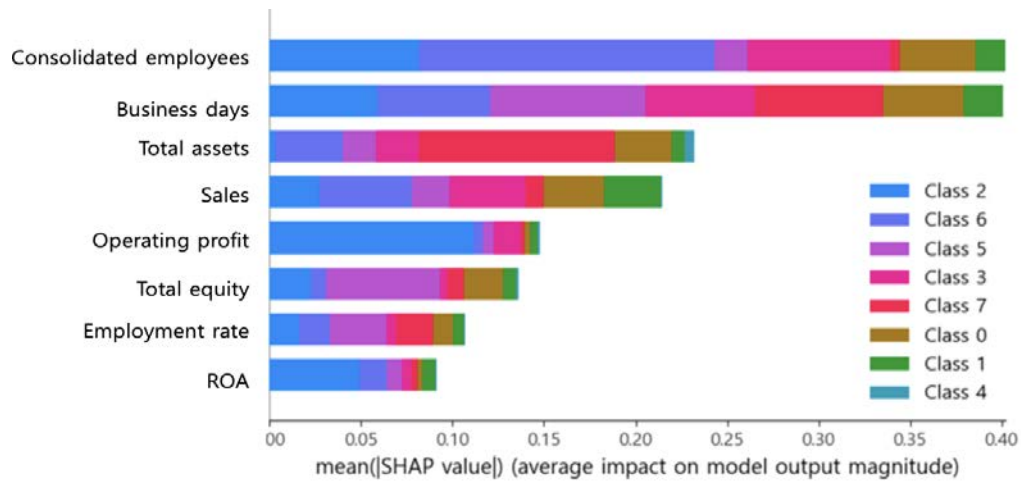
Fig. 3. Visualization of the SHAP values for CEI.

To extracting feature variables, we visualized the best performing models for predicting CEI and CEVT KECO codes using SHAP, highlighting the variables that most influence each prediction. The SHAP value for CEI prediction is presented in **Fig. 3**, and the SHAP value for predicting the KECO code of CEVT is shown in **Fig. 4**. Based on the financial and scale characteristics of the company, we identified the top three variables from the graphs and

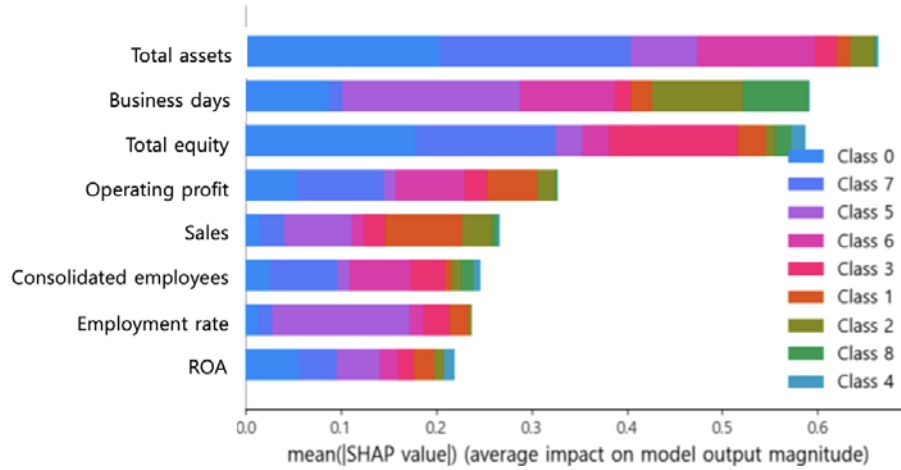selected them as new criteria for classifying similar business groups.

Fig. 4. Visualization of the SHAP values for CEVT.

The SHAP VALUE analysis revealed that the number of integrated employees, business days, and total assets were the most influential feature variables for CEI, while the total assets and number of integrated employees and business days were the most influential for CEVT. Based on these findings, a total of four variables were selected for the new classification criteria for similar companies, which included total asset turnover, operating profit ratio, total asset growth rate, and capital adequacy ratio.

The model used to select similar business groups for the recruitment recommendation service is a clustering model. If a company's profitability, activity, growth, and stability exceed the average value, it is labeled as 1, otherwise as 0, and these labels are merged to create a new cluster variable. Companies with the same cluster variable are then clustered together to select a group of similar companies, as shown in **Fig. 5**.
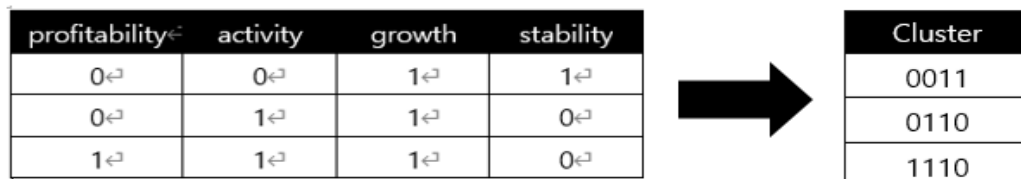

Fig. 5. Cluster of employment

The model used to select similar business groups for the CEI and CEVT recommendation services is a k-ranking model based on Euclidean distance. The model calculates the distance using new classification variables and generates an ascending list, and then selects the closest 'k' companies to the information of new companies as similar companies. Selecting the appropriate 'k' value is crucial, and 20 and 29 KECO codes were chosen as 'k' values for CEVT, considering the time when all the incentives and training received by similar companies are released one by one for each type and the number of high-level employment incentives.

The process of transitioning from the existing criteria to the new criteria involves two steps: clustering and k-ranking. The clustering step involves labeling companies as 1 or 0 based on whether their profitability, activity, growth, and stability exceed the average values. The labels

are then merged to create a new cluster variable, which is used to cluster similar companies. In the k-ranking step, the Euclidean distance between new companies and existing companies is measured using the new classification variables. The closest 'k' companies are selected as similar companies, with appropriate 'k' values chosen based on the number of high-level employment incentives and KECO codes for CEVT, as shown in **Fig. 6**.
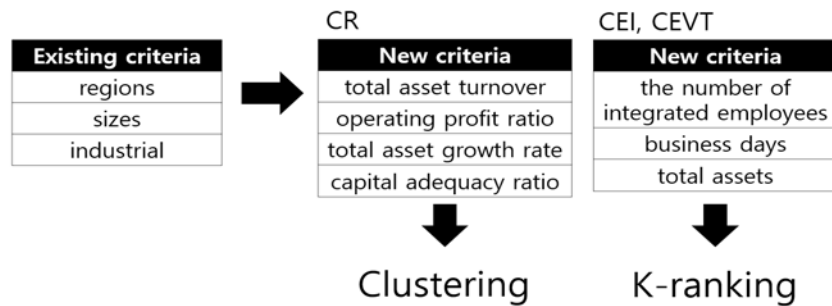

Fig. 6. The process of transitioning from the existing criteria to the new criteria

5. Conclusion

In conclusion, our proposed Similar Business Group Recommendation System offers a comprehensive solution to the challenges faced by small and medium-sized enterprises in Korea. By incorporating derivative criteria and web crawling, we were able to enhance similar business group selection and provide customized and efficient market information to companies. Our model overcomes the limitations of the traditional classification criteria for similar companies by including new selection criteria that are relevant to each type of information. The use of clustering and K-ranking modeling further improves the accuracy of our system.

The benefits of our proposed system are significant. Companies can receive missed or unknown incentives, obtain tailored vocational training, and improve productivity and employment. Customized recruitment notices can also be sent to increase employee satisfaction and lower turnover and resignation rates. The potential impact of our system on the growth and success of small and medium-sized enterprises cannot be overstated.

Future work includes conducting satisfaction surveys to further refine our services and to evaluate the effectiveness of our newly extracted variables for classifying similar companies. We believe that our proposed system can provide a valuable service to companies in Korea and elsewhere, and we hope that our study can inspire further research and development in this area.

Acknowledgement

References

[1]   M.J. Lee, M.S. Park, W. H. Cho, I. S. Na, "Similar Business Group Recommendation System using XGBoost and Derived Variables," in *Proc. of The Proc. of 14th International Conference on Internet (ICONI) 2022*, pp. 159-161, 2022.

[2]   Adomavicius, Gediminas, and Alexander Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE transactions on knowledge and data engineering*, vol. 17, no. 6, pp.734–749, 2005. Article (CrossRef Link)

[3]   J. Konstan, L. Terveen, and J. Lui, "Evaluating Collaborative Filtering Recommender Systems," *ACM Transactions on Information Systems*, vol. 22, pp. 5-53, 2004. Article (CrossRef Link)

[4]   P. Resnick, H.R. Varian, "Recommender Systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56-58, 1997. Article (CrossRef Link)

[5]   James Bennett, and Stan Lanning, "The netflix prize," *Proceedings of KDD cup and workshop*, vol. 9, no. 2, pp. 5-52, 2007. Article (CrossRef Link)

[6]   Davidson, James, et al., "The YouTube video recommendation system," in *Proc. of Proceedings of the fourth ACM conference on Recommender systems*, pp.293–296, 2010. Article (CrossRef Link)

[7]   N. N. Qomariyah, "Pairwise Preferences Learning for Recommender Systems," Doctoral dissertation, University of York, 2018. Article (CrossRef Link)

[8]   ImageNet, Accessed on: March 23, 2023, [Online] Available: https://en.wikipedia.org/wiki/ ImageNet

[9]   Tianqi Chen and Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp.785–794, August 2016. Article (CrossRef Link)

[10] Yoav Freund and Robert E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol.55, no. 1, pp. 119-139, 1997. Article (CrossRef Link)

[11] Guolin Ke, Qi Meng, Thomas Finley et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Proc. of 31st Conference on Neural Information Processing Systems (NIPS 2017)*, vol. 30, pp.1-9, December 2017. Article (CrossRef Link)

[12] Bernhard E. Boser, Isabelle M. Guyon, Vladimir N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. of the fifth annual workshop on Computational learning theory*, pp.144-152, July 1992. Article (CrossRef Link)

[13] Hilary L. Seal, "The historical development of the Gauss linear model," *Biometrika*, vol. 54, no. 1-2, pp. 1-24, 1967. Article (CrossRef Link)

[14] Evelyn Fix and J. L. Hodges, Jr., "Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties," *International Statistical Review / Revue Internationale de Statistique*, vol. 57, no. 3, pp. 238-247, 1989. Article (CrossRef Link)

[15] Scott M. Lundberg, Su-In Lee, "A Unified Approach to Interpreting Model Predictions," in *Proc. of the 31st International Conference on Neural Information Processing Systems*, pp. 4768-4777, December 2017. Article (CrossRef Link)

[16] M. H. Na, W. H. Cho, S. K. Kim, I. S. Na, "Automatic weight prediction system for Korean cattle using Bayesian ridge algorithm on RGB-D image," *Electronics*, vol. 11, no. 10, 1663, 2022. Article (CrossRef Link)

[17] I. S. Na, W. H. Cho, S. K. Kim, M. H. Na, "Fruit Ripeness Prediction Based on DNN Feature Induction from Sparse Dataset," *CMC-Computers, Materials & Continua*, vol. 69, no. 3, pp. 4003-4024, 2021. Article (CrossRef Link)

[18] K. O. Lee, M. K. Lee, I. S. Na, "Predicting Regional Outbreaks of Hepatitis A Using 3D LSTM and Open Data in Korea," *Electronics*, vol. 10, no.21, 2668, 2021. Article (CrossRef Link)

[19] W. Shafqat, and Y. C. Byun, "A Recommendation Mechanism for Under-Emphasized Tourist Spots Using Topic Modeling and Sentiment Analysis," *Sustainability*, vol. 12. no. 1, 320, 2020. Article (CrossRef Link)

[20] E. A. Si, and G. J. Lee, "Association of perfectionism with Job Search Behavior and Career Distress among Nursing Students in South Korea," *J. Korean Acad. Psychiatr. Ment. Health Nurs. ournal of Distribution Science*, vol. 31, no. 1, pp. 27-35, 2022. Article (CrossRef Link)

[21] I. A. Jibril, and M. Yesiltas, "Employee Satisfaction, Telent Mangement Practices and Sustainable Competitive Advantage in the Northern Cyprus Hotel Industry," *Sustainability*, vol. 14, no. 12, 7082, 2022. Article (CrossRef Link)

[22] P. Stamolampros, N. Korfiatis, K. Chalvatzis, and D. Buhalis, "Job satisfaction and employee turnover determinants in high contact services: Insights from Employees's Online reviews," *Tourism Management*, vol. 75, pp.130-147, 2019. Article (CrossRef Link)

[23] E. Ko, Y. Kwon, W. Son, J. Kim, and H. Kim, "Factors Influencing Intention to Use Mobility as a Service: Case Study of Gyeonggi Province, Korea," *Sustainability*, vol. 14, no. 1, 218, 2022. Article (CrossRef Link)

[24] D. Zeng, J. Zhao, W. Zhang, and Y. Zhou, "User-interactive innovation knowledge acquisition model based on social media," *Information Processing & Management*, vol. 59, no. 3, 102923, 2022. Article (CrossRef Link)

[25] S. F. Ahamed, A. Vijayasankar, M. Thenmozhi, S. Rajendar, et. al., "Machine learning models for forecasting and estimation of business operations," *The Journal of High Technology Management Research*, vol. 34, no. 1, 100455, 2023. Article (CrossRef Link)

[26] S. Scardapane, R. Altilio, M. Panella, and A. Uncini, "Distributed spectral clustering based on Euclidean distance matrix completion," in *Proc. of 2016 International Joint Conference on Neural Networks (IJCNN)*, Vancouver, BC, Canada, pp. 3093-3100, 2016. Article (CrossRef Link)

[27] E. Mallinguh, C. Wasike, and Z. Zoltan, "Technology Acquisition and SMEs Performance, the Role of Innovation, Export and the Perception of Owner-Managers," *Journal of Risk and Finalcial Management*, vol. 13, no. 11, 258, 2020. Article (CrossRef Link)

Min Jeong Lee is currently pursuing her undergraduate degree in computer engineering at Chosun University. Her research interests encompass machine learning and computer vision.

In Seop NA obtained his B.S., M.S. and Ph.D. degrees in Computer Science from Chonnam National University, Korea, in 1997, 1999, and 2008, respectively. From 2012 to 2018, he served as a research professor at Chonnam National University, Korea. In 2018, he held the position of a visiting scholar at the University of California, Merced, USA. From 2018 to 2023, he worked as an assistant professor at Chosun University, Korea. Since March 2023, he has been serving as an associate professor at Chonnam National University. His research interests encompass visual intelligence, artificial intelligence, image processing, pattern recognition, object detection/segmentation/recognition and tracking, as well as human emotion recognition.