Available online at 2019.creative-construction-conference.com/proceedings/

CCC 2019

Proceedings of the Creative Construction Conference (2019) 002

Edited by: Miroslaw J. Skibniewski & Miklos Hajdu

https://doi.org/10.3311/CCC2019-002

Creative Construction Conference 2019, CCC 2019, 29 June - 2 July 2019, Budapest, Hungary

Application of artificial intelligence to automate construction materials data classification

Weerayut Chunyaem[a], Chayakrit Charoensiriwath[b], Kriengsak Panuwatwanich*[c], Shigeki Saito[d], Patai Padungtin[e]

[a]Sustainable Energy and Resources Engineering, Sirindhorn International Institute of Technology, Thammasat University, Thailand
[b]National Electronics and Computer Technology Center, National Science and Technology Development Agency, Thailand
[c]School of Civil Engineering and Technology, Sirindhorn International Institute of Technology, Thammasat University, Thailand
[d]Department of Transdisciplinary Science and Engineering, School of Environment and Society, Tokyo Institute of Technology, Japan
[e]BUILK, Thailand

Abstract

The applications of Big Data analytics and Artificial Intelligence (AI) have gained a widespread attention in the construction industry in recent years following the promulgation of Industry 4.0. In the realm of construction research, AI has been utilised widely in areas such as structural design optimization, resource and equipment planning, and project scheduling. The research presented in this paper is aimed to utilise AI to assist with the automatic classification of the large volume of construction material orders created by users through an online marketplace website. Such big data of material orders contained numerous errors (e.g. typographical errors and incorrect units) that were extremely time consuming to correct before the datasets can be used to for further business intelligence analysis. In this research, the dataset was obtained from a business-to-business e-commerce company in Thailand, namely BUILK. The data from BUILK was the construction materials purchase orders created by BUILK's customers through its website, which contained hundreds of thousand unorganized records. In this study, Artificial Neural Networks (ANNs (was applied to automate the categorization of approximately 220,000 records of reinforcement steels orders. The ANNs model was developed and trained using over 32,000 records, with approximately 92 percent of prediction accuracy. The model automatically categorized the steel reinforcement data into 11 groups; Deformed Bars, Round Bars, Wire mesh (Deformed Bars), Wire mesh (Round Bars), Stirrup, Anchor, Material and Others. The outcome of this research helped the company to easily analyze the data to generate insights for its business management and development.

Keywords: artificial nearal network; big data; business intelligence; construction materials; purchase orders; Thailand

*Corresponding author: kriengsak@siit.tu.ac.th

1. Introduction

A number of small and medium sized contractors in Thailand has grown rapidly. More than 10,000 construction materials purchase orders per day were made by these contractors using tradition transaction systems that are difficult to manage and verify. In various construction projects, the cost of materials consists of approximately 50% to 60% of the total cost of each project [1]. Therefore, the better control of construction cost materials plays a significant role in the financial management of a construction project. In this case, an enterprise resource management system in a construction project becomes very essential to understand the usage and flow of construction materials in the project. In addition, having readily access to the supply of construction materials at the prices that are consistent with those estimated in the Bills of Quantities (BOQ) can further improve the project delivery to help decrease the risk in the construction business by streamlining the project supply chain. In Thailand, BUILK is a company that provides Software-as-a-Service (SaaS) for a business-to-business (B2B) e-commerce for the construction industry. The company offers an Enterprise Resource Planning (ERP) solution with recent electronic marketplace service that allows the contractors to easily procure construction materials and products directly from the suppliers. One of the most common services that customers use is creating an online purchase order (PO). Each PO consists of various construction materials or products requested by the contractors. Apart from fulfilling the customers' orders, some of these PO items (POIs) have useful business implications that BUILK can capitalize on. However, the value of data would not be fully exploited, if the construction material PO data are unorganized or contain many errors. Given the widespread applications of Big Data analytics and Artificial Intelligence (AI) in the construction industry in recent years, the research presented in this paper was aimed to apply AI to help organize and cleanse the large volume of PO data stored in BUILK database, using Artificial Neural Networks (ANNs). ANNs is a classification technique in machine learning applied to categorize the construction materials into each group through the extraction of specific words from the material descriptions [2]. The results from this research help the company to visualize their data to reveal interesting trends of material prices and insights into the construction market. In addition, BUILK could use the vast amount of such organized data to increase the efficiency of their electronic marketplace platform.

2. Material and methods

The aim of this research was primarily to apply ANNs to automatically categorize big data of material orders, which contained numerous errors (e.g. typographical errors and incorrect units) so the datasets can be used to for further business intelligence analysis. This was achieved using "Classification", which is a part of supervised learning technique in Machine Learning (ML). Particularly, "Text Classification Process" was employed to prepare the data set prior to being further processed through a model developed using ANNs. In doing so, words in the datasets were converted to numerical values and split into specific proportion of test and train subsets. ANN is a widely-used technique to solve classification problems. For this research, ANN was applied to classify unorganized reinforcement steel data set using the Keras library running on TensorFlow, which has been developed by Google for deployment and operation of large scale machine learning models. The Keras library allows users to easily create neural networks that can be used for deep learning purposes. The following sections provide more details of the abovementioned research tools.

2.1 Text Classification Process

Text Classification Process (Fig. 1) consists of the following steps [3]: (1) Conversion of the original text file into the CSV format and importing it to Python [4]; (2) Word tokenization - words were split into letters and stored in the so-called "bag-of-words", each labelled with a number (commonly known as "features") [5-7]; (3) Stemming - the removal of the misspelled words or words with the same stem and then retaining the stem or the most common of them as a feature. In this paper, stemming process was not used because letters were used instead of words; (4) Identifying "Stopwords" – repeated words or letters were removed in the bag of words to reduce the amount of features. In this research, the same letters were removed from the bag of words with only unique letters retained; (5) Vector representation of text – the predetermined construction material categories were labelled (called a labelled set) using a series of binary (0 or 1) vectors to represent the description of each category [8]; (6) Feature selection (or feature transforming) - the counting of how many times each unique letter occurred in the description of each construction material, and normalizing the frequency of each unique letter with respect to the length of the construction material description (i.e. the total number of letters appear in the description); (7) Machine learning algorithm development - ANNs was employed to develop a text classification algorithm.

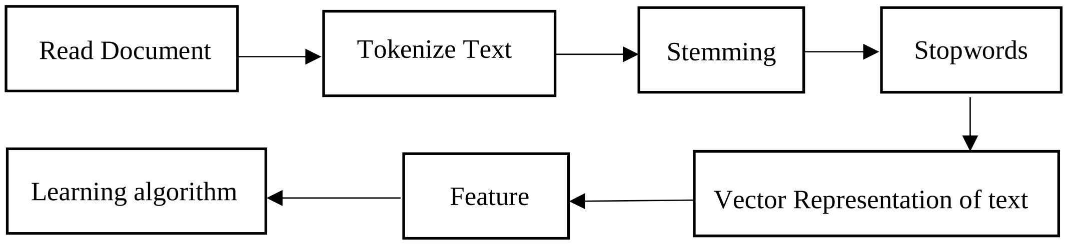

Fig. 1. Text Classification Process.

2.2 Data preparation for Text Classification Algorithm Development

The original data set of BUILK construction material descriptions was divided into train and test sets. The train and test sets were required in the development and measurement of the overall model performance of the text classification algorithm. Prior to the algorithm development, each of the 32,380 construction material descriptions in the original data set was manually grouped into 11 categories (activities, anchor, deformedbar, material, roundbar, screw, steel tube, stirrup, wiremesh deformedbar, wiremesh roundbar and other). Table 1 shows an example of the 11 construction material categories for the selected material descriptions [9]. Each category was then converted to numerical values shown in Table 2. These categorized datasets were then used in the subsequent development of ANNs-based text classification algorithm.

Table 1. Construction material dictionary.

Construction Material Description	Categories
ค่าขนส่งเหล็ก	activities
พุกเหล็ก 1/2	anchor
เหล็กข้ออ้อย 6 มม. SD-40 12 มม.*10 ม. โรงใหญ่ (พับ/ตรง)	deformedbar_db
ใบตัดเหล็ก Dwalt 4" หนา 1 MM	material
ชอล์คหินขีดเหล็ก	other
เหล็กเส้นกลม 6 มม SR 24*10 ม บกส	roundbar_rb
สกรูยึดโครงเหล็ก	screw
เหล็กกล่อง 1*1	steel_tube
เหล็กปลอก 20 x 65 (RB 9)	stirrup
ตะแกรงเหล็กข้ออ้อย 6.0 มิล 15x15 (3.0x4.0)	wiremesh_db
เหล็กตะแกรงไวร์เมชกลม 4.00 m @ 20 x 20 cm. ขนาด 2.50 x5.0 m	wiremesh_rb

Table 2. Construction material categories

Key	Type	Size	Value	Vector
other	int	1	0	[1,0,0,0,0,0,0,0,0,0,0]
deformedbar_db	int	1	1	[0,1,0,0,0,0,0,0,0,0,0]
material	int	1	2	[0,0,1,0,0,0,0,0,0,0,0]
roundbar_rb	int	1	3	[0,0,0,1,0,0,0,0,0,0,0]
activities	int	1	4	[0,0,0,0,1,0,0,0,0,0,0]
steel_tube	int	1	5	[0,0,0,0,0,1,0,0,0,0,0]
screw	int	1	6	[0,0,0,0,0,0,1,0,0,0,0]
anchor	int	1	7	[0,0,0,0,0,0,0,1,0,0,0]
stirrup	int	1	8	[0,0,0,0,0,0,0,0,1,0,0]
wiremesh_rb	int	1	9	[0,0,0,0,0,0,0,0,0,1,0]
wiremesh_db	int	1	10	[0,0,0,0,0,0,0,0,0,0,1]

.23 Artificial Neural Networks

Artificial Neural Networks or ANNs are widely used for performing functions related to learning, classification, relating, determination specialty, generalization and optimization similar to those performed by a human brain [10]. In the construction industry, ANNs can be used for classification of information such as construction materials, types, and suppliers. The muti-layer perceptron (MLP), which is the most common type of ANN, was used in this research to model the potentially nonlinear response of the unorganized construction material description data. The MLP is a collection of connected processing elements called nodes or neurons, arranged together in three layers: input layer, hidden layer (intermediate) and output layer [11]. The raw data in the MLP neural network is normalized and fed into the input layer. It is then passed into to the hidden layer and subsequently into the output layer as shown in Fig.2.

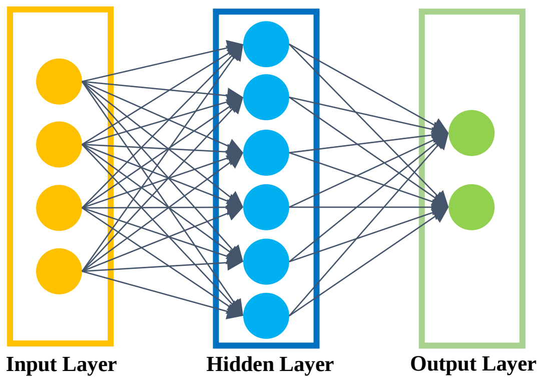

Fig. 2. Diagram of a multilayer feed forward neural network.

For this research, the input layer consists of 196 unique letters extracted from the material descriptions as carried out in the feature selection (feature transformation) process explained in Section 2.1. The output layer consists of the 11 construction material categories stored in binary vectors (Table 2). The datasets were divided into train and test sets based on the ratio of 70:30 [12]. The 70% data portion was trained using the Keras framework in TensorFlow [13]. The results of the ANNs is determined by the nature of its activation function. Activation functions are an extremely important feature of the ANNs. They decide whether a neuron should be activated or not, i.e. whether the information that the neuron is receiving is relevant for the given information or should it be ignored. It is the non-linear transformation that is performed over the input signal. This transformed output is then moved to the next layer of neurons as input. In this paper, Rectified Linear Unit (ReLU) and SoftMax are the commonly used activation functions that were used in the developed model [14, 15].

3. Results and discussion

As summarized in Table 3, the result showed that the developed model (based on the original datasets of 32,380 construction material records) was able to achieve the training cross validation and test accuracy values of more than 90%. In particular, the model achieved the following test accuracy metrics: Precision 0.92, Recall 0.90, and F1-score 0.91. Table 4 shows the reduction in losses and increase in percentage of accuracy across a 10-fold training cross validation. These results indicate that the developed model is of acceptable performance.

Table 3. Accuracy metrics

Metrics / Models	ANNs Model
Training cross validation	≈ 98%
Test accuracy	92.33%
Precision	0.92
Recall	0.90
F1-score	0.91

https://doi.org/10.3311/CCC2019-002

Table 4. Training accuracies and losses per fold in 10-fold training cross validation

Fold #	Loss	Accuracy (x100%)
1	0.5078	0.8434
2	0.2371	0.9231
3	0.1754	0.9416
4	0.1332	0.9568
5	0.1009	0.9672
6	0.0809	0.9748
7	0.0658	0.9797
8	0.0500	0.9852
9	0.0431	0.9870
10	0.0359	0.9896

The predictive performance of the developed ANNs model was then tested using the Confusion Matrix, which shows how well the predicted value match with the true value as indicated by the intensity of the color of the squares alinging diagonally across the matrix. As shown in Fig. 3, the ANNs model had the high number of correct predictions as indicated by the high intensity of the color of the squares.

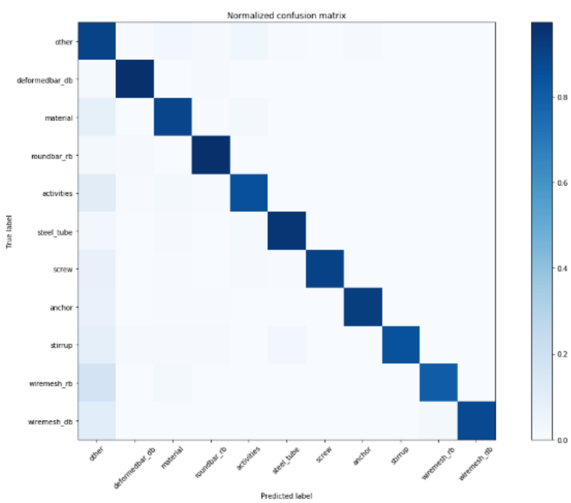

Fig. 3. Normalized confusion matrix of the construction materials labelled

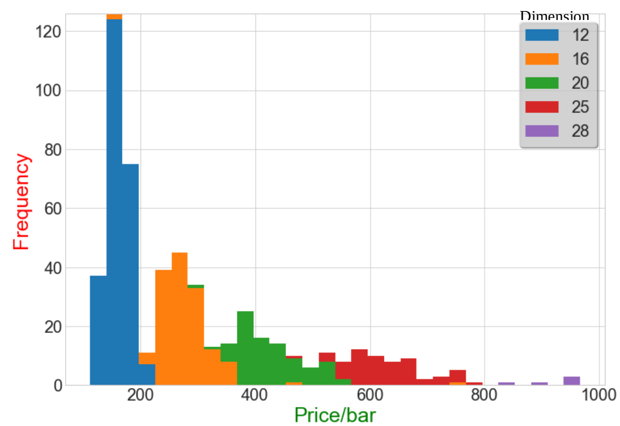

Fig. 4. Deformed Bar priced comparing in various dimension.

The applicability of the developed model was then tested using the new dataset obtained from the company. This new dataset consisted of approximately 220,000 records of purchase orders. The model was applied to categorise the reinforcement steel bars (rebars) into 11 categories. For the company, this model has a benefit in assisting the business analyst to quickly organize the large volume of purchase orders data in order to help extract some insights for their business. For example, as shown in Fig. 4, the results of the categorisation can be used to illustrate the prices and demand of steel rebars across different diameters. This type of information can help the analyst to quickly understand the price distributions of steel rebars based on different sizings. Addition analysis of the categorized data may help reveal other important information such as the price fluctuation of rebars across different region at different time points.

4. Conclusion

The research presented in this paper demonstrates a business application of AI in the construction industry apart from the technical application such as in areas such as structural design optimization, resource and equipment planning, and project scheduling. In this research, AI was applied using ANNs-based text classification technique to automate the categorization of unorganized big data of construction material orders. The ANNs model was developed using the actual purchase order data from the construction e-commerce company in Thailand, namely BUILK. The was developed and tested using the data of 32,380 construction material purchase order records. The developed model demonstrated an acceptable performance and was subsequent used to automatically categorize the larger set of 220,000 records of purchase orders. Such model helped the company to easily organize and analyze the data to generate insights for its business management and development. The future application of such model could also further benefit the online construction material searching service of BUILK whereby the model could make search recommendation to users in the same way as that seen in Google Search Engine.

References

[1] Safa, M., et al., *Supplier selection process in an integrated construction materials management model.* Automation in Construction, 2014. **48**: p. 64-73. https://doi.org/10.1016/j.autcon.2014.08.008
[2] Moselhi, O., T. Hegazy, and P. Fazio, *Neural networks as tools in construction.* Journal of ConstructionEngineering and Management, 1991. **117**(4): p. 606-625. https://doi.org/10.1061/(ASCE)0733-9364(1991)117:4(606)
[3] Sebastiani, F., *Machine learning in automated text categorization.* ACM computing surveys (CSUR), 2002. **34**(1): p. 1-47.
[4] McKinney, W., *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. 2012: " O'Reilly Media, Inc.".
[5] Al Qady, M. and A. Kandil, *Concept relation extraction from construction documents using natural language processing.* Journal of Construction Engineering and Management, 2009. **136**(3): p. 294-302. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000131
[6] Zhang, J. and N. El-Gohary, *Extraction of construction regulatory requirements from textual documents using natural language processing techniques*, in *Computing in Civil Engineering (2012)*. 2012. p. 453-460. https://doi.org/10.1061/9780784412343.0057
[7] Zhang, J. and N.M. El-Gohary, *Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking.* Journal of Computing in Civil Engineering, 2013. **30**(2): p.04015014. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000346
[8] Leopold, E. and J. Kindermann, *Text categorization with support vector machines. How to represent texts in input space?* Machine Learning, 2002. **46**(1-3): p. 423-444. https://doi.org/10.1023/A:1012491419635
[9] Dixon, L., et al. *Measuring and mitigating unintended bias in text classification*. in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. 2018. ACM. https://doi.org:/10.1145/3278721.3278729
[10] Baykal, H. and H.K. Yildirim, *Application of artificial neural networks (ANNs) in wine technology.* Critical reviews in food science and nutrition, 2013. **53**(5): p. 415-421.
[11] Kourou, K., et al., *Machine learning applications in cancer prognosis and prediction.* Computational and structural biotechnology journal, 2015. **13**: p. 8-17. https://doi.org/10.1016/j.csbj.2014.11.005
[12] Ghiassi, M., et al., *Automated text classification using a dynamic artificial neural network model.* Expert Systems with Applications, 2012. **39**(12): p. 10967-10976. https://doi.org/10.1016/j.eswa.2012.03.027
[13] Rampasek, L. and A. Goldenberg, *Tensorflow: Biology's gateway to deep learning?* Cell systems, 2016. **2**(1): p. 12-14.
[14] Gold, S. and A. Rangarajan, *Softmax to softassign: Neural network algorithms for combinatorial optimization.* Journal of Artificial Neural Networks, 1996. **2**(4): p. 381-399.
[15] Agarap, A.F., *Deep learning using rectified linear units (relu).* arXiv preprint arXiv:1803.08375, 2018.