

Intellectual Property Creation based on Construction of Patent Evaluation Model

1st Shigeaki Onoda

Department of Information Systems Engineering
Toyama Prefectural University
Toyama, Japan
t855005@st.pu-toyama.ac.jp

2nd Koji Okuhara

Department of Information Systems Engineering
Toyama Prefectural University
Toyama, Japan
example@okuhara.com

3rd Ota Mamoru

Department of Information Systems Engineering
Toyama Prefectural University
Toyama, Japan
example@ota.com

Abstract—Nowadays, the web includes enormous data such as patent. This research find the rule of textual patent data and create new model.

Index Terms—patent, big data, DEA, natural language processing

I. INTRODUCTION

A needs of informatization in firms or publics sector is increasing. In Ministry of Public Management in Japan, the transparency, reliability and accessibility is as agenda for boosting Open-Data strategy [1]. For example, some agency conduct making kind of hazzard map as API [2]. Then, Japan Patent Office release a API service named "Patent information Platform" which includes Patent documents or utility models. Somebody can search any patent in Japan.

However, such Open-Data is not organized well for analysis as Big Data. The reason is the documents in the service is saved as PDF. PDF data is not format for scraping. Moreover, it doesn't include citation number of patents. Therefore, I will apply DEA: Data Envelopment Analysis) to some patent datas to evaluate this data mathematically, and I will suggest patent evaluation model.

II. METHODOLOGY

A. DEA

DEA is a a nonparametric method coined by Charnes, and it is used to apply units evaluation of a firm in default of absolute scale [3][4]. Concretely, measure productivity of DMU: Decision Making Unit in a organization as a target by using of promotion of profits and assets. Efficiency of a branch named "k" is following formula.

Identify applicable funding agency here. If none, delete this.

$$\begin{aligned} \max \quad & \theta_k = \frac{\sum_{n=1}^N v_{kn} y_{kn}}{\sum_{m=1}^N u_{km} x_{km}} \\ \text{subject to} \quad & \frac{\sum_{n=1}^N v_{kn} y_{sn}}{\sum_{m=1}^N u_{km} x_{sn}} \leq 1 \quad (s = 1, 2, 3, LK) \\ & u_{km} \geq 0 \quad (m = 1, 2, 3, LM) \\ & v_{kn} \geq 0 \quad (n = 1, 2, 3, LN) \end{aligned} \quad (1)$$

x and y show inputs and outputs, and u and v is show each weight of direction, so s is a number of target DMU. This formula shows multi-input/multi-output, weighted DMU to maximize efficiency for k. To sum up, each DMU advantageously evaluate myself.

B. CCR Model

To solve (1) as liner problem, I convert a denominator or a numerator to scalar number "1". This method is called as CCR. Now, I set a numerator to 1 and formulate as following (2).

$$\begin{aligned} \max \quad & \theta_k = \frac{\sum_{n=1}^N v_{kn} y_{kn}}{\sum_{m=1}^N u_{km} x_{km}} \\ \text{subject to} \quad & \sum_{m=1}^N u_{km} x_{sn} - \sum_{n=1}^N v_{kn} y_{sn} \geq 0 \\ & \sum_{m=1}^N u_{km} x_{sn} = 1 \quad (s = 1, 2, 3, LK) \\ & u_{km} \geq 0 \quad (m = 1, 2, 3, LM) \\ & v_{kn} \geq 0 \quad (n = 1, 2, 3, LN) \end{aligned} \quad (2)$$

Solving liner problem for (2), could be emitted each efficiency. The DEA concept figure 1 be shown below. the frontier in the figure indicate Production UTF2013possibility frontier. The closer DMU is, the more efficient DMU is.

Figure1 The concept of DEA

III. COLLECT PATENT DATA FROM OPEN DATA

A. Collecting Method

In this study, I use Google Patent as a resource to analysis patent steadily. Google Patent browse html format, it is easier for gathering data than non constructed data like PDF.

By the way, Google Patent API is closed in 2015, I take unique way to collect data. Firstly, I have collected some data which includes 1.patent, 2.inventor, 3.title, 4.grant date, 5.citation number, 6.cited by number, 7.words in the patent documentation. Moreover, I also extract only noun from the documents as well as Tumura [5].

B. DataBase

I have to construct data base to analysis. Collection data could divides 7 type, and each word data is different during sorts of words. I apply NoSQL as DB because of scalability of it.

IV. RESULTS

A. Data, Observations

Applying DEA method to patent data is studied by Hyeonju [6]. Hyeonju find new business area from patent data, but I will shave the input data to word frequency matrix. That can predict a valued patent and valued word in patent. This method aim at finding core words in a domain "patent". Next, I will show concrete rule to apply DEA.

Rule1: I use word: w_i in patent i as a input, and a input sets scalar Z if no word: w_i occur in patent i

Rule2: I set frequency distribution as inputs. It is called Plain input.

B. Further Analysis

under construction.

V. CONCLUSION

A. Interpretation of Results/Discussion

This study make evaluation model of patent area. Thanks to that, We will conduct trends analysis in a technological field. The last, this study will help us to do multi-modal learning.

B. Future Word

As a subject in the future, I would like to do stemming terminology or condense dimension of word vector by NMF method to decrease computational complexity.

REFERENCES

- [1] Ministry of Public Management, "Boost Open Data Strategy", http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/, Accessed: May 5, 2018.
- [2] NTT Communications, "Report of Deploying Information Infrastructure of Hazzard Data", "No. 1.0, 2013.
- [3] Charnes, A., Cooper, W.W., Rhodes, E., "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research*, Vol. 2, , pp. 429-444, 1978.
- [4] ToshiYuki Sueyoshi, ""DEA: Business Analysis method", Asakura, 2001.
- [5] Hyonju Seol, Sungjoo Lee, Chulhyun Kim, "Identifying new business areas using patent informatin: A DEA and text mining approach", *Expert Systems with Applications*, vol. 38, pp. 2933-2941, 2011.
- [6] Takumi Tsumura, Fumiaki Saito, Shohei Ishizu, "Knowledge Extraction from Textual Patent Data Using a Random Forest", *J Jpn Ind Manage Assoc*, vol. 68, No. 3, pp. 161-170, 2017.
- [7] Akira Gto, Kiminori Genba, Jun Suzuki, Schumpeter Tamada, "Classification Index of Important Patent", *RIETI Discussion Paper Series*, 06-J-018, 2018.