Technology Marketing using PCA , SOM, and STP Strategy Modeling

Sunghae Jun

Department of Bioinformactics and Statistics, Cheongju University
Cheongju, Chungbuk 360-764, Korea

Abstract

Technology marketing is a total processing about identifying and meeting the technological needs of human society. Most technology results exist in intellectual properties like patents. In our research, we consider patent document as a technology. So patent data are analyzed by Principal Component Analysis (PCA) and Self Organizing Map (SOM) for STP(Segmentation, Targeting, and Positioning) strategy modeling. STP is a popular approach for developing marketing strategies. We use STP strategy modeling for technology marketing. Also PCA and SOM are used to analyze patent data in STP modeling. To verify improved performance of our study, we make experiments using patent data from USPTO.

Keywords: *Technology Marketing, Principal Component Analysis, Self Organizing Map, Segmentation, Targeting, Positioning.*

1. Introduction

This paper proposes a technology marketing method using machine learning algorithms and marketing engineering tools. We use principal component analysis (PCA) and self organizing map (SOM) for machine learning algorithms [6],[10]. STP (Segmentation, Targeting, and Positioning) strategy modeling is considered for marketing engineering tool in our research [9],[12]. Technology marketing is a total processing about identifying and meeting the technological needs of human society. Most technology results exist in intellectual properties like patents [17]. Recently, the researches for patent analysis have been published [1],[8],[11],[16],[18]. On our research, we consider patent document as a technology. So, patent documents are analyzed by text mining [2], PCA, SOM, and STP strategy modeling. Original patent documents are transformed into document–term matrix as structured data by keyword extraction of text mining [3]. And then, using PCA, the document–term matrix is reduced suitable dimension size to be analyzed [4]. STP is a popular approach for developing marketing strategies [9],[12]. We use STP strategy modeling for technology marketing. Also PCA and SOM are used to analyze patent data in STP modeling. To verify improved performance of our study, we make experiments using patent data from United States Patent and Trademark Office (USPTO) [15].

2. Related Works

2.1 Technology Marketing

Technology marketing (TM) is at variance with general business marketing. TM is defined as total behavior of technology transfer for planning, selling, and sales promotion from technology developer to technology demander. According to development of industrial structure, the importance of technology marketing has been increased. We have to consider the following issues for technology marketing. They are organization of TM, selection of selling techniques, technology packaging, planning of TM, technology contract, post evaluation, and feedback.

2.2 SOM, PCA, and Text Mining

SOM: Self organizing map (SOM) was introduced by T. Kohonen [10]. SOM is an unsupervised learning algorithm for clustering. Also SOM is called as a neural networks model based on competitive learning. It has two layers which are input and feature layers. We can cluster all elements by feature map with two dimensions. Firstly SOM performs clustering with input vector x and weight matrix M. The data point x_i is treated one at a time. Also the closest m_j to x_i is found by Euclidean distance, and then m_j is updated as the following [7].

$$m_k = m_j + \alpha(x_i - m_j) \quad (1)$$

where m_j and m_k are current and new weights. So m_k moves to x_i. This learning is repeated until given conditions such as change rate of weights and the number of repeat.

PCA: Principal component analysis (PCA) is a dimension – reduction method [6],[7]. The aim of PCA is to transform input vector with high dimension (p) into feature vector with low dimension (k). Though the k is smaller than p, PCA has the minimum information loss.

Text mining: Text mining is a data mining technique for finding hidden and useful patterns from a large text database [3]. General text mining methodology consists of document preprocessing and indexing [2]. There are collection creation, document parsing, document segmentation, and text summarization in the document preprocessing [14]. Also keyword extraction, phase extraction, morphological analysis, stop-word filtering, term association, and term clustering are the methods for indexing. In addition, we consider topic clustering and mapping for our research. The methods of topic clustering are term selection, document categorization, and cluster title generation. Trend, query, aggregation, and zooming maps are used in topic mapping.

2.2 STP Strategy Modeling

Segmentation, targeting, and positioning (STP) is a popular strategy model in marketing [9],[13]. Segmentation clusters customers to similar groups by their wants and needs. In targeting, marketers determine one or two groups for their marketing approaches. Positioning explains the competition power of company's product to target segments.

Segmentation: Segmentation has five phases which are segmenting markets, describing market segments, evaluating segments attractiveness, selecting target segments and allocating resources to segments, and finding targeted customers [13]. Using factor analysis to reduce the data and forming segments by cluster analysis are popular methods for segmentation. Also a perceptual map is a good approach for segmentation. This map shows a visual representation of competitive alternatives and customers' preference. In this paper, we use recency – frequency (RF) chart as a perceptual map.

Targeting: Targeting is the approach how to assess the attractiveness of segments, determine one or two segments to serve, and identify target customers [13].

Positioning: Positioning is the behavior of planning company's offering and image for occupying a distinctive place in target customers [9]. The goal of targeting is to posit the product or brand in customers' mind and maximize the potential benefit of the company.

3. Technology Marketing using PCA, SOM, and STP

We propose a technology marketing approach by analysis of patent data. Patent data has complete information about technology. To discover knowledge in patent documents, we use text mining, PCA, SOM, and STP. Our proposed approach has six steps as the following.

Step1: Searching patent data
Using keywords equation in title, abstract, …
: Retrieved patent documents

Step2: Document – term matrix
Using text mining
: Document(n) – term(k) matrix, ($n<<k$)

Step3: Document – PC matrix
Using SVD – Principal Component Analysis
: Document(n) – PC(p) matrix, ($n>p$) and ($k>>p$)

Step4: Segmentation
Using Self Organizing Map (SOM)
: Patents clustering

Step5: Targeting
Using top 5 keywords
: Defining one or two clusters for target market

Step6: Positioning
Using target result
: Increasing benefit and contribution of company

In the step 1, we retrieve patent documents from USPTO. A searching formula is needed for patent retrieval. Our study uses the searching formula from abstract in patent document. The number of terms (k) is very larger than the number of documents (n) in step 2. k and n are the numbers of variables and observations respectively. But in general PCA model, the number of variables has to be extremely smaller than the number of observations. So we are not able to do PCA to this document – term matrix directly. To settle this problem, we consider singular value decomposition (SVD). We can construct SVD of document – term matrix X [5],[7].

$$X=UDV^t \quad (2)$$

Many computational algorithms for this standard decomposition have existed. Where U is an orthogonal matrix and its columns u are called left singular vectors. V is also an orthogonal matrix with right singular vectors v. D is a diagonal matrix and its diagonal elements are $d_1 \geq d_2 \geq \ldots \geq d_p \geq 0$ called singular values. According to SVD

– PCA in step 2, we get document – PC matrix in step 3. The number of PCs (p) is very smaller than n. Also we select the optimal number of PCs by screeplot with 10% or more variance. To get efficient segmentation, we use SOM as a clustering method. SOM has provided many results in diverse data types such as text, document [15]. After SOM clustering, each group is represented its detailed technology by extracted top five keywords. We can define the technology fields of all segments and determine target market. Finally we try positioning approach from determined target market. The following figure shows the process of our proposed method for technology marketing.

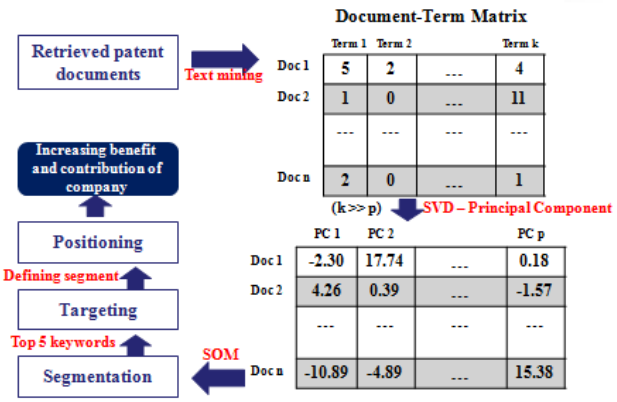

Fig. 1 Proposed technology marketing process.

Increasing company's benefit and contribution are effective results of our study. Also we can expect the diverse contributions for marketing approach.

4. Experimental Results

We apply our research to marketing engineering technologies. A company, developing business models for marketing engineering, wants to know which technologies are needed for 'marketing engineering' field. In this case, we analyze patent data of marketing engineering for solving above problem. The real data for our experiments are patent documents about marketing engineering. So, we retrieved patent data from United State Patent and Trademark Office USPTO). The keywords equation for searching patents is the following.

Abstract = Marketing * Engineering

We selected patents with marketing and engineering as keywords in their abstracts from 1948 to 2010. Also the time point of patent retrieval was November 9, 2010. Total number of selected patent documents was 80. The following table shows all searched patents.

Table 1: Total number of selected patent documents

Years	# of Patents
1948 – 1970	0
1971 – 1980	4
1981 – 1990	1
1991 – 2000	25
2001 – 2010	50
Total	80

Most patents of the technology of marketing engineering have been shown after 1991. So we are able to think this technology has enormous potential for research and development. Also we know the total number of registered patents through 1948 – 2010 is 80.

Using keywords extraction of text mining, we got document – term matrix. This matrix has 80 documents and 2018 terms. The number of columns (2018) is very larger than the rows (80). So we are not able to do PCA directly. To solve this problem, we use SVD. And then we have another problem of PCA. This is how many principal components should be retained for analysis. There is no clear answer but a couple of popular rules. One rule is to consider only those with variance over 10% of each principal component.

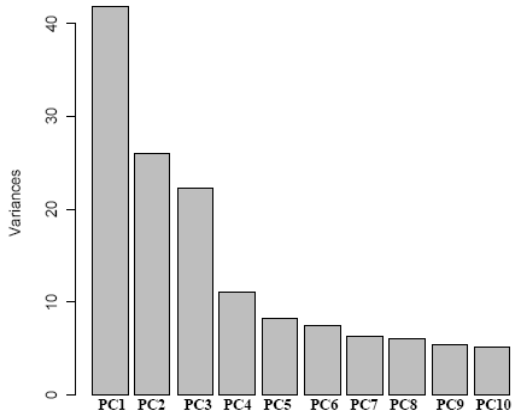

Fig. 2 Screeplot of Top 10 principal components (PCs).

From above figure, we determined 4 as the number of principal components (PCs). Then we got PCs for 80 documents using SVD-PCA. The following figure shows the result of SVD-PCA.

Doc #	PC1	PC2	PC3	PC4
1	2.591274	-3.0827508	0.4813304	-0.7171706
2	2.514716	-4.3660141	-0.5926086	1.6769411
3	2.651611	-6.2916608	-2.0429489	2.4285974
4	2.651611	-6.2916608	-2.0429489	2.4285974
5	2.943499	-3.6309257	2.4725174	3.3832317
…				
76	3.159518	-8.4218044	-4.7578218	1.4009504
77	2.481203	-4.8516302	-0.8211147	-2.7172378
78	2.200017	-3.8377048	-3.6012114	-11.7621848
79	-8.986990	2.1634336	-0.8544428	0.3446253
80	1.634265	-4.1265055	-1.3808656	-2.8129552

Fig. 3 Result of SVD-PCA.

Doc # is the I.D. number of each document in the above figure. We are able to do the segmentation approach using above top 4 PCs. For segmentation of STP strategy modeling, we use an unsupervised neural networks model called SOM. We can look at the SOM result in the next figure.

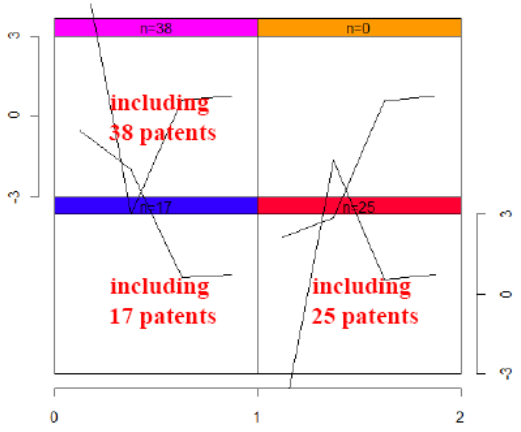

Fig. 4 SOM result.

From the SOM result, we find 3 segments for STP strategy modeling. The following table shows the segmentation result by SOM.

Table 2: Segmentation result by SOM

Segment	Patent no.	# of patents
1	8, 18, 19, 20, 21, 22, 23, 24, 36, 40, 41, 46, 50, 55, 58, 61, 69	17
2	1, 2, 3, 4, 5, 6, 7, 12, 13, 16, 25, 26, 28, 29, 30, 32, 33, 35, 37, 38, 42, 43, 44, 45, 49, 52, 53, 54, 57, 59, 60, 62, 74, 75, 76, 77, 78, 80	38
3	9, 10, 11, 14, 15, 17, 27, 31, 34, 39, 47, 48, 51, 56, 63, 64, 65, 66, 67, 68, 70, 71, 72, 73, 79	25

From segmentation to targeting, to define each segment is needed. So we define 3 segments by extracting top 5 keywords as the following table.

Table 3: Top 5 keywords in 3 segments

Segments	Keywords
Segment 1	Software, Configuration, Hardware, Service, Computer
Segment 2	Diagnostic, bungle, Claim, Business, Simulation
Segment 3	Communication, Connection, Electronic, Internet, Mail

We did not consider the following keywords, because they showed all segments or were not meaningful. They were ‘and’, ‘the’, ‘marketing’, ‘engineering’, ‘for’, ‘which’, ‘this’, ‘that’, and so forth.

Table 4: Defining segments by top 5 keywords

Segments	Defining segment
Segment 1	Technology field for constructing software/hardware systems of marketing engineering
Segment 2	Technology field for developing evaluation system of management performance and simulation tool of trial and error
Segment 3	Technology field for building on-line marketing system based on Internet.

To select targeted segment, we propose recency–frequency (RF) chart as a perceptual map.

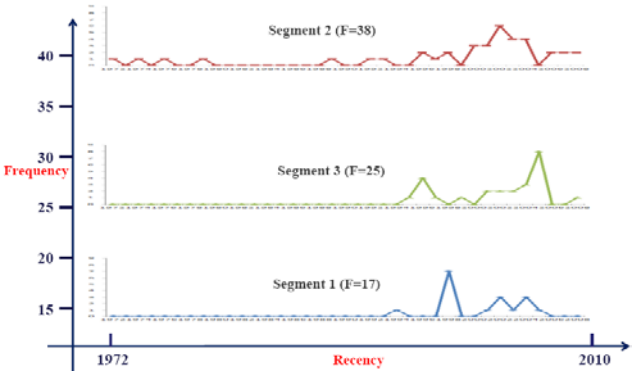

Fig. 5 RF chart.

According to vertical axis of FR chart, we knew segment 2 is higher than others. Also we can find segment 2 and segment 3 have more recency weights than segment 1.
By the frequency of RF chart, we can select segment 2 as a target technology market. Also we determined segment 3 for selecting target technology market based on recency of RF chart. But we want to determine only segment for targeting. So we need a solution of the problem. To settle this problem, we proposed an recency – frequency (RF) score function as the following.

$$\text{RFscore}(i) = \alpha \times \text{Tf}_i + (1-\alpha) \times \left(\frac{\sum_{k=fy}^{ly} (k - (fy-1))}{\text{Tf}_i} \right) \quad (3)$$

Where *i* is segment number and α is a weight. Tf_i is defined as total frequency of segment *i*. *fy* and *ly* represent first year and last year of the patents. The segment with the largest score becomes a target market. We selected segment 2 as a target market of the technology of marketing engineering by the following table.

Table 5: RF scores of 3 segments

Segments	Parameters	
	α=0.5	α=0.4
Segment 1	23.21	24.45
Segment 2	32.78	31.73
Segment 3	27.94	28.53

Though the values of α, the RF scores of segment 2 were as large as ever. Therefore in segmentation step of STP strategy modeling, we determined segment 2 to the target market.
Next we analyzed the selected target market. The following figure shows the number of patents in segment 2 by year.

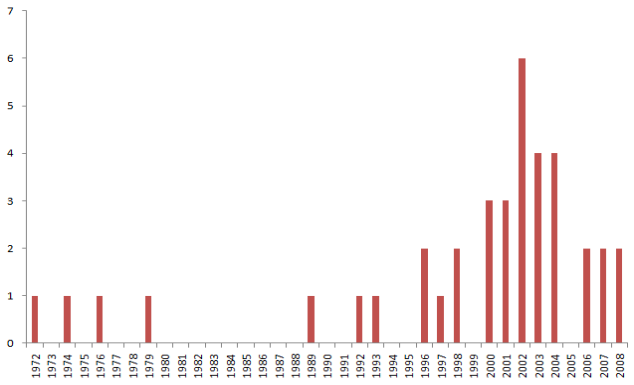

Fig. 6 # of patents in segment 2 by year.

In targeting step, we can define the technologies of segment 2 for marketing engineering as evaluation systems of management performance and simulation tools of trial and error in diverse business environments using top 5 keywords.
In the next positioning step, we consider some positioning strategies. To begin with the needs of marketing engineering technology for segment 2 have been increased. Also we can advice companies related to marketing engineering have to do R & D for segment 2 based technologies. Therefore, the companies will be strong enterprises with good intellectual properties and business models of marketing engineering. In addition, their benefits and social contributions will be increased.
In our experiment, we used R-project statistical computing package as computing software [13].

5. Conclusions and Future works

A company's R & D strategy of marketing engineering technology is at an early state, because there is a few patent up to now in the world. So we have many chances in marketing engineering as the following. There are patents, research publications, developing new technologies, developing vacant technologies, and so forth.
We think our research has some limitation of this study. We have not any publication related to this case study, because there are few research papers about our study. So we had so difficult to work our research. In addition, we thought to need more complete gate for STP strategy modeling through technology marketing. We can expect better performance using advanced RFM score function of the segment. In this study, we used RF score except monetary (M). We have to find out a measure for M of a patent.

References

[1] X. Chen, W. Yin, P. Tu, and H. Zhang, "Weighted k-Means Algorithm Based Text Clustering," Proceedings of International Symposium on Information Engineering and Electronic Commerce, 2009, pp. 51-55.
[2] M. Fattori, G. Pedrazzi, and R. Turra, "Text mining applied to patent mapping: a practical business case," World Patent Information, Vol. 25, 2003, pp. 335–342.
[3] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurasamy, Advances in knowledge discovery and data mining, AAAI Press/The MIT Press, 1996.
[4] I. Feinerer, K. Hornik, and D. Meyer, "Text Mining Infrastructure in R", Journal of Statistical Software, Vol. 25, Iss. 5, 2008, pp. 1-54.
[5] G. H. Golub, and C. Reinsch, "Singular value decomposition and least squares solutions," Numerische Mathematik Vol. 14, No. 5, 1970, pp. 403-420.
[6] J. F. Hair, B. Black, B. Babin, and R. E. Anderson, Multivariate Data Analysis, Prentice Hall, 1992.

[7] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning – Data Mining, Inference, and Prediction, Springer, 2001.
[8] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. S., Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, 2002, pp. 881–892.
[9] P. Kotler, and K. L. Keller, Marketing Management, Prentice Hall, 2009.
[10] T. Kohonen, Self-Organizing Maps, Springer, 2000.
[11] S. Lee, B. Yoon, and Y. Park, "An approach to discovering new technology opportunities: Keyword-based patent map approach", Technovation, Vol. 29, 2009, pp. 481-497.
[12] G. L. Lilien, and A. Rangaswamy, Marketing Engineering, Prentice Hall, 2003.
[13] R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org, 2010.
[14] Y. H. Tseng, C. J. Lin, and Y. I. Lin, "Text mining techniques for patent analysis", Information Processing and Management, Vol. 43, 2007, pp. 1216-1247.
[15] United States Patent and Trademark Office (USPTO), www.uspto.gov
[16] P. Wang, I. M. Cockburn, and M. L. Puterman, " Analysis of Patent Data-A Mixed Poisson Regression Model Approach," Journal of Business & Economic Statistics, Vol. 16, No. 1, 1998, pp. 27-41.
[17] B. Yoon, and Y. Park, "Development of New Technology Fore casting Algorithm: Hybrid Approach for Morphology Analysis and Conjoint Analysis of Patent Information," IEEE Transactions on Engineering Management, Vol. 54, No. 3, 2007, pp. 588-599.
[18] B. Yoon, and S. Lee, "Patent analysis for technology forecasting: Sector-specific applications," Proceeding of IEEE International Conference on Engineering Management, 2008, pp. 1-5.

Sunghae Jun is an associate professor in the department of Statistics, Cheongju University, Korea. He received the B.S., M.S., and Ph.D. degrees in department of Statistics from Inha University, Korea, in 1993, 1996, and 2001. Also he got his doctorate in computer science and engineering from Sogang University, Korea, 2007. He worked in NCR as a data mining consultant from 2000 to 2001. His research fields are machine learning, evolutionary computing, and management of technology(MOT).