This article was downloaded by: [National Science Library]

On: 26 April 2012, At: 17:03

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Technology Analysis & Strategic Management

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/ctas20>

Patent text mining and informetric-based patent technology morphological analysis: an empirical study

Xu Feng^a & Leng Fuhai^b

^a Strategy Research Center, Institute of Scientific and Technical Information of China, Fuxing Road No.15, Beijing, 100038, China

^b National Science Library, Chinese Academy of Science, Beijing, 100080, China

Available online: 26 Apr 2012

To cite this article: Xu Feng & Leng Fuhai (2012): Patent text mining and informetric-based patent technology morphological analysis: an empirical study, Technology Analysis & Strategic Management, 24:5, 467-479

To link to this article: <http://dx.doi.org/10.1080/09537325.2012.674669>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Patent text mining and informetric-based patent technology morphological analysis: an empirical study

Xu Feng^{a*} and Leng Fuhai^b

^aStrategy Research Center, Institute of Scientific and Technical Information of China, Fuxing Road No.15, Beijing, 100038, China; ^bNational Science Library, Chinese Academy of Science, Beijing, 100080, China

Patent technology morphological analysis utilises the advantages of both patent information analysis and morphological analysis to provide more detailed information on current and future patent technology. Current patent technology morphological analysis methods are largely reliant on manual expertise in the construction of morphological boxes with few approaches to the evaluation of future morphological configurations. We developed a patent text mining and informetric-based patent technology morphological analysis technique. We defined the basic parameters of the morphological box as the factors in factor analysis, with a patent keywords matrix, and the clusters in clustering analysis with factor scores. Patent citation, year of patent registration, keyword frequency, and contributing factors were used to evaluate future morphological configurations. We used an empirical study of liquid crystal display wide viewing angle patents to validate our method. The Thomson Reuters Derwent Innovation Index was used to collect patent text datasets. Our study indicates that this method is feasible for the implementation of patent technology morphological analysis. Our method provides advantages in terms of cost and time reductions during morphological box construction and more flexible methods for evaluating morphological configurations. We discuss problems and possible development of our method in the last section.

Keywords: patent text mining; morphological analysis; factor analysis; clustering analysis; patent citation; conjoint analysis

1. Introduction

Future-oriented technology analyses such as technology foresight, technology forecasting, and technology opportunity analysis, have been widely studied areas since the RAND Corporation developed the Delphi method in the 1950s, to analyse new military systems (Cahnin et al. 2008). Various methods were developed to conduct these studies. Previous researchers have categorised these methods into specific areas, including technology forecasting (Slocum 2001; Martino 2003), technology future analysis (Technology Future Analysis Working Group 2004), and technology intelligence (Lichtenthaler 2005). Experimental studies using these research methodologies were

*Corresponding author. Email: xufeng@istic.ac.cn

also reported. Based on an analysis of previous research, Cahnin et al. (2008) concluded that a combination of two or more methods was commonly used to study possible profiles and future technology trends.

Expert opinion methods such as Delphi and scenario analysis are widely used, but patent analysis is increasingly being adopted in studies owing to the natural relationship between patent text and technology. Patents are the most important information resource used for future-oriented technology analysis. Patent analysis methods such as patent mapping and patent index can be used in many technology-related studies (Lee et al. 2008). Porter and Cunningham (2005) provided a comprehensive overview and guidelines for patent mining-based tech analysis. The main application of patent analysis is in the identification of trends and profiles for existing technologies, but patent analysis cannot provide insights about the future of technology. Combining patent analysis with other methods is the most effective approach for future-oriented studies.

Text mining and informetric analysis are widely used methods for conducting future-oriented researches. Co-word analysis and literature-based knowledge discovery are both useful text mining methods for conducting such research, with many valuable outcomes (Law et al. 1988; Callon, Courtial, and Laville 1991; Leydesdorff and Zhou 2008). These methods facilitate deep exploration of textual content enabling the discovery of relationships or structures among the topics or themes of the text.

2. Morphological analysis of patent technology

Morphological analysis (MA) was introduced by Zwicky in the 1940s for structuring and investigating the total set of relationships contained in multi-dimensional, non-quantifiable problem complexes (Ritchey 2005). Ritchey (2003) developed a systematic MA tool and conducted numerous important studies using MA. MA applications are found in the fields of policy analysis and futures studies, including strategic decision support combined with Bayesian networks (Waal and Ritchey 2007) and future studies combined with scenario modeling (Ritchey 2009). The principle of MA is to construct a morphological box of the problem, which contains the appropriate parameters (or dimensions) and the relevant values or conditions assigned to them. Analysis of possible future configurations for the problem is achieved by testing combinations of parameters for the morphological box.

The two key elements of MA are the morphological box and the evaluation of morphology configurations (MC). We define the morphological box as a morphological structure (MS) and refer to the MS parameters, and their values, as attributives and levels, respectively. Thus, the MS is a structure consisting of attributives and levels. Levels are the alternatives for each attributive. Evaluation of MC methods generally uses a cross-consistency assessment, which requires that the decision-maker produces a priority list of future MCs (Ritchey 1998). Numerous articles on this subject are available on the website of the Swedish Morphological Society (<http://www.swemorph.com>), which provides more detailed information on the method and applications.

Yoon and Park (2004) introduced a patent text keyword-based morphological analysis approach for technology forecasting. Yoon and Park (2005, 2007, 2008) provided a detailed introduction to their hybrid method and experimental studies of thin-film technology, light-emitting diode, wide-viewing angle (TFT-LED WVA) patent technology. Their method used text-mining to extract keywords from patent texts in the United States Patent and Trademark Office (USPTO), followed by MS construction using the keywords assisted by domain experts. The authors used the MS to combine patent citation data with conjunction analysis (CA) to evaluate future patent technology

MCs and deliver a priority list of possible combinations of morphological attributives and levels. We refer to Yoon and Park's method as patent technology morphological analysis (PTMA).

The PTMA method combines the advantages of MA, text mining and patent analysis, to provide a profile of current technology and possibilities for future technology. PTMA has been applied in other studies (Wang et al. 2008; Gu and Huang 2009; Huang and Li 2009). Yoon, Phaal, and Probert (2008) also extended PTMA to technology roadmapping.

There are several problems with the PTMA method developed by Yoon and Park. First, MS construction during the initial stage of the PTMA process, is still highly reliant on expert opinions. Keywords are extracted from the patent text as parameters for the structure, followed by a factor analysis to help experts with MS construction, but the major component of the MS is still defined by experts, making this procedure expensive and time-consuming. Expert matching of keywords to the MSs might lead to further problems, because of the possibility of misunderstandings of keywords by experts. Second, the evaluation of future MCs is critical to the practical application of PTMA. Yoon and Park used patent citation analysis and conjunction analysis to prepare priority lists for MCs. Furthermore, some patent technology might impose limitations during patent citation analysis, because of an absence of citations or citation delay. Thus we investigated whether there might be better methods that provide more alternatives to decision-makers.

We developed a text-mining and informetric-based PTMA (IPTMA) to address the problems of PTMA. First, we used co-word analysis based on keywords extracted from the patent database to construct a MS, which to some extent reduces the expense and time intensive aspects of expert-based methods. We assumed that keyword analysis methods, such as factor analysis and clustering analysis, which are used in co-word analysis of patent texts, reflect the actual structure of keywords in the patent text, thereby avoiding possible misunderstandings by experts. Second, we assumed that the structures constructed by co-words analysis would provide more statistical information about keywords and structures, which may be useful for future MC evaluation. We used all available information to conduct our evaluation and compare the results with the method of Yoon and Park.

We used the liquid crystal display wide viewing angle (LCD-WVA) patent resource for empirical validation of our IPTMA, which is a similar technology area to that tested in the studies of Yoon and Park.

3. Data and methods

Data

Thomson Reuters Derwent Innovation Index (DII) was used as a data source in our study. DII is one of the largest commercial patent databases, which is widely used as a research information resource by patent technology researchers.

Using retrieval experiments and background study of LCD-WVA technology, we defined the following retrieval tactic.

SUBJECT = ('wide viewing angle') AND SUBJECT = (LCD or 'liquid crystal display')
AND PN = (US*)

We retrieved two patent datasets for our experimental studies. The first dataset (SET1) was a collection of patents prior to and including 2005. The second (SET2) covered the years 2006–2008. SET1 was used for the construction of MS and SET2 was used as the validation dataset.

Figure 1. LCD-WVA patent growth during the years 1976–2008.

Figure 2. List of patentees during the years 1976–2008.

After priority list produced by the MC analysis were excavated by the SET1, we examined whether they had been exploited in the SET2 for validation.

SET1 contained 312 patents, whereas SET2 contained 184 patents. Figure 1 shows the growth in patent numbers during the years 1976–2008. LCD-WVA patents show a rapid growth trend, especially after 2000. Figure 2 shows patentees in SET1. LG Philips LCD Co. Ltd, Samsung Electronics Co. Ltd and Sharp KK were the leading patent owners in the LCD-WVA.

Keyword extraction and pretreatment

Text document analysis (TDA) software was used for keyword extraction and pretreatment of the datasets. We use TDA to extract keywords (words and phrases found in TDA) from the DII datasets and merged the keywords extracted from the title and abstract into one keyword field, before cleaning the field with TDA tools. We subsequently constructed two matrices, a patent–keyword matrix and a keyword–keyword matrix. Both matrices were used in the experimental study.

MS construction method

It is widely accepted that informetric methods can be used to outline, or construct, profiles or structures for research topics, or themes. This method differs from expert opinion methods, because informetric approaches require greater care. An MS used in MA is generally meaningful semantic descriptions of the technology or problem.

We constructed an MS based on informetric analysis of keywords, where keywords were carefully selected. To achieve this we designed a systematic keyword selection method combining high frequency keywords, amendatory mutual information measurement (AMIM), and artificial selection. AMIM is a method introduced by Wang (2007), which selects keywords that

are semantically related to the main theme of interest. The keywords matrix was used to identify co-occurrences of keywords and calculate the AMIM. We implemented a minor modification to Wang's AMIM, where we used co-occurrence instance frequency to represent the co-occurrence of keyword records, which emphasised the overall frequency of keywords in the patent text.

$$I(X, Y) = \ln(X'Y') \log_2 \frac{p(X, Y)}{p(X)p(Y)}$$

where $X'Y'$ is the co-occurrence instance frequency of DII, $p(X)$ and $p(Y)$ are the records of keywords and 'wide viewing angle', respectively, and $p(X, Y)$ is the co-occurrence records of keywords and 'wide viewing angle'.

According to the TDA definition, 'instances' are the overall frequency of keywords, while 'records' are the appearances of keywords in the patent.

The complete process diagram for our method is shown in Figure 3. Clustering analysis and factor analysis are the most widely used informetric methods, where co-words analysis is used to study topic structures and hotspots (Whittaker 1989; Stegmann 2003; Leydesdorff and Vaughan 2006; Yang 2007). According to Kostoff et al. (2007), factor analysis tends to identify quantitative relationships among theme components, whereas clustering analysis tends to identify structural relationships among themes. Leydesdorff and Welbers (2011) proposed that it was preferable to factor analyse 2-mode word-document matrix to exploit the semantic structures. We considered the requirements for MS and developed the following informetric-based MS construction method.

(1) Factor analysis for MS level definition

It is a requirement of MS that the MS levels should be semantically independent. A patent-keyword matrix was used for factor analysis with SPSS15.0. We use principal component analysis (PCA) to extract factors and varimax to rotate factors, in our experiment. Factors extracted by PCA are linearly independent which satisfies the requirements for MS level definition. Factors scores and contributions (loading) resulted from factor analysis are also kept for the following analysis.

Figure 3. Process diagram for MS construction.

(2) *Clustering analysis for definition of MS attributives*

Clustering analysis was used to cluster factors identified by factor analysis, so as to construct the MS attributives. Factor clusters were used as MS attributives. A hierarchical clustering method was used for conducting clustering analysis in SPSS 15.0.

MC evaluation

A major obstacle for MA-related studies, which is still not widely reported, is that MC evaluation is often difficult because of the huge number of MCs. For example, a typical MS with 6–8 attributives would produce 50,000–5,000,000 MCs. Cross-consistency assessment was used to eliminate undesirable MCs, but the method was still limited without the assistance of expert opinion. Yoon and Park (2007) used patent citation analysis and conjoint analysis in combination with cross-consistency assessment. MS was simpler in our study, because the DII patents are limited to approximately 120 words of text. This meant it was not necessary to conduct cross-consistency assessment.

As previously mentioned, we used informetric methods to construct the MS for LCD-WVA. There is an abundance of MS statistical information that can be used for MC evaluation, other than the patent citations used by Yoon and Park. We used the frequency of keywords, factor contributions, and the year of main patent access number for evaluation, and we also made a comparison with the patent citation method. Conjoint analysis was used to evaluate MS attributives and level, which we combined with patent year and patent citation analysis. Our evaluation method was as follows.

- (1) Patent citation combined with conjoint analysis: considering MC with more patent citations has higher importance;
- (2) Patent date combined with conjoint analysis: considering MC with newer patent has higher importance;
- (3) Frequency of keywords: considering MC with higher keywords frequency has higher importance;
- (4) Factor contribution in factor analysis: considering MC with higher factor contribution has higher importance.

Framework diagram for our MC evaluation method is showed in Figure 4. We compared all methods by performing an evaluation with both datasets. We analysed SET1 and defined the identified MCs and unidentified MCs, where the later can be regarded as future patent technology

Figure 4. Framework diagram of MC evaluation.

MCs that need to be evaluated. MC analysis of SET2 enabled us to determine whether any newly identified MCs appearing after 2005. Newly identified MCs allowed us to validate the value of conducting PTMA for LCD-WVA, but they also allowed us to compare our four future MC evaluation methods.

Patent citation cannot be used for direct evaluation, because of citation delays, so we used revised citation data. We used the oldest year (1990) of the patent number main registration to revise the patent year, which enhanced the importance of newer patents. The two revision methods are as follows.

$$C_r = \frac{1}{n} \sum_{i=1}^n \left(\frac{C_i}{2009 - Y_i} \right), \quad i = 1, 2, \dots, n$$

C_r is the revised citation number. C_i is the original citation number. Y_i is the year of the main registration number.

$$Y_{mt} = \frac{1}{n} \sum_{i=1}^n (Y_i - 1990 + 1), \quad i = 1, 2, \dots, n$$

Y_{mt} represents the revised year. Y_i represents the year of the main registration number.

4. Results

MS construction based on informetric methods

Table 1 shows a list of keywords selected for MS construction. Keywords were selected from the top one hundred high frequency keywords detected with AMIM, which scores above zero. Our selection process eliminated numerous meaningless, or useless, keywords from the raw keywords list, such as the internal words of DII.

Table 2 shows the results of factor analysis of the patent-keyword matrix, where seven factors with eigenvalues above 1 were extracted from the raw matrix. Factor analysis was also performed with the keyword-keyword matrix for words widely used in co-word studies. The factor analysis results for the two matrices showed that the results of the keyword-keyword matrix tend to

Table 1. Keywords selected for MS construction.

No.	Records	Keywords	AMIM
1	56	polarise	0.092256082
2	26	parallel	0.212495729
3	25	plate	1.108054498
4	14	protrusions	0.059161627
5	11	retardation	0.67438093
6	9	perpendicular	0.45272042
7	9	black matrix	0.010747986
8	8	alignment film	1.079251665
9	8	compensation film	0.40040551
10	7	driving voltage	0.910815153
11	7	reflective	0.380411036
12	7	optical axis	0.345648355
13	6	alignment direction	0.201410604

Table 2. Results of factor analysis (Rotated with Varimax).

Keywords	Factors						
	1	2	3	4	5	6	7
Optical axis	0.848	0.025	0.189	0.048	−0.011	−0.004	0.082
Polarise	0.807	0.090	−0.131	0.066	−0.066	0.016	−0.112
Perpendicular	0.037	0.763	0.207	−0.097	0.004	0.044	−0.022
Parallel	0.019	0.699	0.043	0.083	−0.056	−0.241	0.001
Alignment direction	0.062	0.638	−0.219	−0.007	−0.010	0.201	0.067
Compensation film	0.322	−0.061	0.743	−0.278	−0.014	0.053	0.091
Plate	−0.217	0.133	0.709	0.341	−0.104	0.019	−0.034
Retardation	0.118	−0.038	0.011	0.890	−0.059	0.054	0.071
Protrusions	−0.036	−0.040	−0.038	−0.042	0.655	0.068	0.115
Black matrix	−0.027	−0.009	−0.035	−0.023	0.612	0.032	−0.007
Driving voltage	−0.037	−0.016	−0.103	−0.117	−0.217	−0.800	0.190
Alignment film	−0.072	−0.040	−0.140	−0.176	−0.401	0.523	0.460
Reflective	0.011	−0.049	−0.066	−0.101	−0.165	0.137	−0.849

Figure 5. Clustering factor analysis results.

represent different technology configurations. In comparison, the result factors for the patent-keyword matrix can represent the main aspects of LCD-WVA technology, and provide a structural description of the technology with factors and keywords rotated within them. We used these factors as MS levels during further MS construction. We also retained the factor scores, and their contribution, which were used for clustering analysis and MC evaluation, respectively.

Figure 5 shows the results of clustering analysis with factor scores gained from the factor analysis. It can be seen that three clusters grouped seven main factors. These clusters represent the relationships of factors and were used as MS attributives contained in the levels (factors).

Using the levels derived from factor analysis and attributives derived by clustering analysis with factor scores, we completed MS construction, as shown in Table 3. The MS contains three attributives ‘polarise’, ‘alignment’ and ‘compensation’. Each attributive has two or three levels, which are represented by several keywords. We designated levels and attributives based on the frequency of keywords in the structure, aided by background information regarding LCD-WVA.

MC analysis

Permutation and combination of MS provided 12 possible MCs for use in the configuration of patent technology. We found patents with a complete MS using the patent-keyword matrix. Many

Table 3. MS of LCD-WVA derived by the informetric methods.

Attributives	Levels	Keywords
A (polarise)	A1 (ipolarise)	polarise, optical axis
	A2 (protrusions)	protrusions, black matrix
B (alignment)	B1 (alignment direction)	parallel, perpendicular, alignment direction
	B2 (reflective)	reflective
C (compensation)	C1 (retardation)	retardation
	C2 (alignment film)	alignment film, driving voltage
	C3 (compensation film)	plate, compensation film

Table 4. Result of MC analysis for SET1.

No.	MCs	Number	PN
1	A1-B1-C3	7	2003-844413 1999-493006 1995-346230 1995-148875 1990-202503 1990-009823 1990-009822
2	A1-B2-C3	6	2005-628569 2004-033422 2004-010472 2002-681357 2002-583059 2001-079434
3	A1-B1-C1	4	2002-610225 2002-530537 1998-528933 1990-031310
4	A1-B2-C1	3	2001-529257 2001-254346 1999-001559
5	A2-B1-C3	1	2005-280058

Note that we used the main registration number given by DII as the unique PN of each patent.

patents have an incomplete MS, but this is attributable to the limitations of patent text or missing attributives. These levels are not considered key points of patent technology and are not necessarily described in the patent text.

We found five MCs in 21 patents for SET1, as shown in Table 4. Thus, seven MCs (see Table 5) had never been identified prior to 2005, and these seven MCs were regarded as the future MCs for evaluation.

We compared different MC evaluation methods and also conducted MC analysis with SET2. As shown in Table 6, six MCs were identified in SET2. Four of these MCs had already appeared in SET1, but two of them were new MCs that appeared after 2005. This result shows that two new MCs appeared in the patent database after 2005. This validated our MA method for identifying new configurations of LCD-WVA technology.

Table 5. Seven MCs not identified prior to 2005.

No.	MCs
1	A1-B1-C2
2	A2-B1-C3
3	A1-B2-C2
4	A2-B2-C1
5	A2-B1-C2
6	A2-B2-C3
7	A2-B2-C2

Table 6. MCs analysis results for SET2.

No.	MCs	Number	PN
<i>Identified MCs</i>			
1	A1-B1-C3	6	2008-L39688 2008-H31228 2008-F28201 2006-066127 2006-057434 2006-048398
2	A1-B1-C1	4	2008-G49745 2006-765554 2006-492137 2006-447455
3	A1-B2-C3	3	2008-D81523 2008-C46675 2006-707419
4	A1-B2-C1	2	2007-737485 2006-646892
<i>Unidentified MCs in SET1</i>			
5	A1-B1-C2	1	2008-H00241
6	A2-B2-C3	1	2008-G33494

Using the two MC analysis results, we compared all the different methods we previously discussed.

Table 7 shows evaluation results for all 12 MCs using the four methods. We found that all five MCs identified in SET1 were ranked relatively highly by the four evaluation methods. This means that these MCs successfully represented the main technology structures or characteristics. This shows that it was very useful to conduct a detailed analysis of technology trends with the MCs, and that this method will be very beneficial for enterprises making decisions about a patent technology development strategy.

Table 8 shows the evaluation results for the seven MCs that were not identified in SET1, using the four methods. Two MCs were newly identified in SET2 and we found that the methods could all rank the new MCs with a relative high ranking, especially 'A1-B1-C2'. All methods were useful for MC evaluation. A combination of evaluation methods allowed us to list the priority of future MCs, but also provided more comprehensive information about possible MCs that might be valuable for setting up a patent technology development strategy. We found different alternatives

Table 7. MC evaluation results for all 12 MCs using four evaluation methods.

MCs		Methods			
		Citation analysis + CA	Patent date + CA	Frequency of keywords	Factor contribution
MCs identified in SET1	A1-B1-C1	1	4	2	2
	A1-B1-C3	2	2	1	1
	A1-B2-C1	3	3	5	7
	A1-B2-C3	4	1	3	5
	A2-B1-C3	8	6	7	4
MCs identified in SET2	A1-B1-C2	5	10	4	3
	A2-B2-C3	10	5	9	10
MCs unidentified in SET2	A2-B1-C1	6	8	8	6
	A1-B2-C2	7	9	6	9
	A2-B2-C1	9	7	11	11
	A2-B1-C2	11	12	10	8
	A2-B2-C2	12	11	12	12

Table 8. MC evaluation results for seven future MCs using four evaluation methods.

Evaluation method	Identified in 2006–2008				Unidentified in 2006–2008		
	A1-B1-C2	A2-B2-C3	A2-B1-C1	A1-B2-C2	A2-B2-C1	A2-B1-C2	A2-B2-C2
Citation analysis + CA	1	5	2	3	4	6	7
Patent date + CA	5	1	3	4	2	7	6
Frequency of keywords	1	4	3	2	6	5	7
Factors contribution	1	5	2	4	6	3	7
Total rank of unidentified MCs			1	2	3	4	5

based on the evaluation results. For example, MCs that list high in the citation analysis and CA evaluation method might be recommended when an enterprise regards patent value as more important (more citations imply higher patent value).

5. Discussion and conclusions

The development of text mining, patent analysis and future-oriented technology analysis, using a combination of multiple methods is increasing in popularity as a method of providing more detailed information about future technology. We used text mining and an informetric method, based on the PTMA method introduced by Yoon and Park (2005), for MS construction and the development of an alternative method to traditional expert opinion methods. Our experimental results show that our method is useful for MS construction in PTMA and the method also reduces time consumption and cost. However, several problems require resolution, such as the selection of keywords used for construction of MS and the designation of attributives and levels for the MS.

Our MC analysis also indicates the practical feasibility of our methods. We found that new combinations were identified in the SET2 texts which we did not find in SET1, which validates the utility of our MS construction method. We use information resulting from our informetric analysis to evaluate the MCs and compared them with the method of Yoon and Park (2007). This

comparison indicates that evaluation results for future MCs provided by methods such as keyword frequency, or factor score-based MC evaluation, are very similar to patent citation and conjoint analysis. The differences among results means it is useful to compare evaluation results and this provides more comprehensive information for decision-makers to aid the development of future patent technology development strategy.

The MS may change with the development of technology, particularly because attributives or levels may be added or eliminated (Yoon and Park 2005). The method we developed makes it simple to modify the MS for patent technology and perform the PTMA.

More studies are required to validate the method introduced in this paper. Further experiments in other technology areas would improve the applicability of our method. Additional studies might resolve some problems found in our method, such as keyword selection and the designation of the attributives and levels. We only tested the patent text of DII, where patent text size is limited to mostly less than 120 words, which means that further tests of patent technology text with higher information limits are required. Other patent texts such as the patent database of USPTO could be used to extend this analysis.

Acknowledgements

This project was supported by a grant (No. 70873123/G030701) from the National Science Foundation of China (NSFC).

Notes on contributor

Xu Feng is an Associate Researcher in the Research Center for Strategic Science and Technology Issues of the Institute of Scientific and Technical Information of China. He received his PhD in Management Science with a specialisation in information research in 2010 from Graduate University of Chinese Academy of Sciences. His main research interests include future-oriented technology analysis, decision-making in R&D management and governmental innovation strategy.

Leng Fuhai is a Researcher in National Science Library of Chinese Academy of Sciences and a Professor in Graduate University of Chinese Academy of Sciences. His dominant fields of research are information research and competitive intelligence.

References

- Cahnin, C., M. Keenan, R. Johnston, F. Scapolo, and R. Barré, eds. 2008. *Future-oriented technology analysis*. Heidelberg: Springer.
- Callon, M., J.P. Courtial, and F. Laville. 1991. Co-word analysis as a tool for describing the network of interactions between basis and technological research: the case of polymer chemistry. *Scientometrics* 22: 155–205.
- Gu, C.J., and L.C. Huang. 2009. A systematic study of the work that web mining does in the technology opportunities analysis. Paper presented at 'First international workshop on database technology and applications', 25–26 April 2009, 299–302. London: IEEE.
- Kostoff, R.N., J.A. del Río, H.D. Cortés, C. Smith, A. Smith, C. Wagner, L. Leydesdorff, G. Karypis, G. Malpohl, and R. Tshiteya. 2007. Clustering methodologies for identifying country core competencies. *Journal of Information Science* 33: 21–40.
- Law, J., S. Bauin, J., P. Courtial, and J. Whittaker. 1988. Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. *Scientometrics* 14: 251–64.
- Lee, S.J., S.H. Lee, H.J. Seol, and Y. Park. 2008. Using patent information for designing new product and technology: keyword based technology roadmapping. *R&D Management* 38: 169–88.
- Leydesdorff, L., and L. Vaughan. 2006. Co-occurrence matrices and their applications in information science: Extending ACA to the web environment. *Journal of the American Society for Information Science and Technology* 57: 1616–28.
- Leydesdorff, L., and K. Welbers. 2011. The semantic mapping of words and co-words in contexts. <http://www.leydesdorff.net/words/index.htm>

- Leydesdorff, L., and P. Zhou. 2008. Co-word analysis using the Chinese character set. *Journal of the American Society for Information Science and Technology* 59: 1528–30.
- Lichtenthaler, E. 2005. The choice of technology intelligence methods in multinationals: Towards a contingency approach. *International Journal of Technology Management* 32: 388–407.
- Martino, J.P. 2003. A review of selected recent advances in technological forecasting. *Technological forecasting & Social Change* 70: 719–33.
- Porter, A.L., and S.W. Cunningham. 2005. *Tech mining: Exploiting new technologies for competitive advantage*, 215–48. Hoboken, NJ: John Wiley.
- Ritchey, T. 1998. Morphological analysis – A general method for non-quantified modeling. <http://www.swemorph.com/pdf/gma.pdf>
- Ritchey, T. 2003. MA/Casper – Advanced computer support for morphological analysis. <http://www.swemorph.com/macasper.html>
- Ritchey, T. 2005. Futures studies using morphological analysis. <http://www.swemorph.com/pdf/futures.pdf>
- Ritchey, T. 2009. Developing scenario laboratories with computer-aided morphological analysis. <http://www.swemorph.com/pdf/cornwallis3.pdf>
- Slocum, M.S. 2001. Technology forecasting: From emotional to empirical. *Creativity and Innovation Management* 10: 139–52.
- Stegmann, J. 2003. Hypothesis generation guided by co-word clustering. *Scientometrics* 56: 111–35.
- Technology Futures Analysis Methods Working Group. 2004. Technology futures analysis: Toward integration of the field and new methods. *Technological Forecasting and Social Change* 71: 287–303.
- Waal, A.D., and T. Ritchey. 2007. Combining morphological analysis and bayesian networks for strategic decision support. <http://www.swemorph.com/pdf/mabn.pdf>
- Wang, J.F. 2007. *Scientometric based methodology for science and technology knowledge evolution relationship analysis*. Beijing: Chinese Academy of Sciences.
- Wang, J.W., L.C. Huang, J. Li, and J. Li. 2008. Empirical research on the technology opportunities analysis based on morphology analysis and conjoint analysis. Paper presented at PICMET 2008 Proceedings: Technology Management for a Sustainable Economy, 27–31 July 2008, Cape Town, South Africa, 2041–7. Portland, OR: PICMET.
- Whittaker, J. 1989. Creativity and conformity in science: Titles, keywords and co-word analysis. *Social Studies of Science* 19: 473–96.
- Yang, L.Y. 2007. *The theoretical and applied study of occurrence and co-occurrence phenomena*. Beijing: Chinese Academy of Sciences.
- Yoon, B.G. 2008. On the development of a technology intelligence tool for identifying technology opportunity. *Expert Systems with Applications* 35: 124–35.
- Yoon, B.G., and Y.T. Park. 2004. Morphology analysis approach for technology forecasting. In *IEEE 2004 International Engineering Management Conference*, Vol. 2, 566–70. London: IEEE.
- Yoon, B.G., and Y.T. Park. 2005. A systematic approach for identifying technology opportunities: Keyword-based morphology analysis. *Technological Forecasting & Social Change* 72: 145–60.
- Yoon, B.G., and Y.T. Park. 2007. Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information. *IEEE Transactions on Engineering Management* 54: 588–99.
- Yoon, B.G., R. Phaai, and D. Probert. 2008. Morphology analysis for technology roadmapping: Application of text mining. *R&D Management* 38: 51–68.