A hybrid approach to identify and forecast technological opportunities based on topic modeling and sentiment analysis

Tingting, Ma

Logistics School, Beijing Wuzi University, matingtingmay@163.com

Ruiping, Cheng

Logistics School, Beijing Wuzi University, bjann99019@163.com

Hongshu, Chen

Management and Economic School, Beijing Institute of Technology, hongshu.chen@bit.edu.cn

Xiao, Zhou

Economic and Management School, Xidian University, belinda1214@126.com

The spring up of new & emerging technologies brings a lot of innovation opportunities for society, which enables technology opportunities analysis attracts increasing attention by both industry and academia recently. This study proposes a hybrid approach which integrates topic modeling, sentiment analysis, patent mining and expert judgment to identify technological topics and the potential development opportunities. In order to illustrate how the approach is validated and optimized, and to present its potential to contribute technical intelligence for research and development management, we apply the hybrid approach to analyze a set of 9883 DII records that involved dye sensitized solar cell research. The main contributions of this study include three-fold. First, we distinguished the terms in the different parts of DII patent documents when utilizing them to recognize technical topics. Second, we utilized the terms extracted from the Advantage and the Use part to identify topics on technical problems and applications, and proposed a probability-based topic relation measurement method to identify the relationships of the technical problems and applications with the core sub-technologies. Third, we introduced both topic modeling and sentiment analysis to support technical topic analysis.

\begin{CCSXML}
<ccs2012>
<concept>
<concept_id>10002944.10011123.10011124</concept_id>
<concept_desc>General and reference~Metrics</concept_desc>
<concept_significance>500</concept_significance>
</concept>
</ccs2012>
\end{CCSXML}
\ccsdesc[500]{General and reference~Metrics}

Additional Keywords and Phrases: Technological Opportunities Analysis, Topic Modeling, Sentiment Analysis, Topic Relation Recognition

1 INTRODUCTION

In the past, technological opportunities analysis (TOA) is highly dependent on expert judgment[1]. However, the ambiguous and uncertainness of new & emerging technologies (NETs) makes it not easy for experts to reach a consensus for technological forecasting[1]. In order to make up the drawback, Dr. Porter and his team propose a semi-automatic approach that combines tech mining and bibliometrics to generate intelligence on technologies by mining the wealth of information available in public science, technology and innovation (ST&I) database[2]. The approach is well applied to facilitate technological opportunity identification of NETs so as to provide objective basis for judgment. Since patent data has more wealthy and high-quality technical information than other ST&I data[3], many efforts have been made to advance TOA method based on patent mining. However, current studies mainly focus on exploring technical innovative opportunities by extracting topic information on technologies from patent content, few studies pays attention to mining technology application opportunities. In fact, patent document not only records technical novelties, details and advantages but also describes uses of patented technology, which are valuable for exploring technical application opportunities. As Derwent patents describes technical detail, novelty, advantage and use respectively, we attempts to recognize both technological topics and application topics by mining intelligence separately from the different parts of Derwent patent document. Topic modeling method is introduced to recognize topics from patents and we employed sentiment analysis, combined with the traditional bibliometric indicators, to judge the value of topics from multi-aspect. The hybrid method can not only quantitatively explain whether the topics are hot or not, but also show the attitudes of inventors towards the topics. Finally, we propose a probability-based topic relation measurement method to link technological topics with application topics, which is better for us to understand application scenarios of identified technologies.

In sum, our study proposes a hybrid approach that combines topic modeling and sentiment analysis to identify and forecast technological opportunities. The study also presents a case analysis for DSSC, which serves to illustrate how the approach is validated and improved, and its potential to contribute technical intelligence for research and development (R&D) management. The main contributions include three folds: 1) exploring both technical innovative opportunities and application opportunities by mining different parts of Derwent patent document; 2) integrating sentiment analysis and bibliometric indicators to judge the value of topics from multi-aspect; 3) proposing a probability-based topic relation measurement method to identify the relationships of the applications with the core sub-technologies.

2 DATA AND METHODOLOGY

In this paper, we use Derwent patent data to support the TOA. By reading and analyzing the different parts of the DII abstracts, we find that comparing to the Novelty, Detailed description and Description of drawing(s) parts focusing on describing technical novelties and details, the contents in the Use parts are more emphases on describing where this technology can be applied. In these regards, we firstly employ topic modeling method to recognize both technological topics and application topics by mining different parts of Derwent patent. Then, we integrate sentiment analysis with the quantitative bibliometric assessment methods to identify hot, valuable and potential technologies. The sentiment analysis is employed to gauge the attitudes of domain experts towards these topics. Finally, we propose a probability-based topic relation measurement method to link the identified technologies with their highly related applications, in order to figure out promising applications for specific technologies.

2.1 Stage 1 - Data gathering and preprocessing

Our approach first involves collecting Derwent patents related to a target NETs. We retrieve and download raw patent data from Derwent Innovation Index (DII) database based on its batch export function, and then import it into Vantagepoint software for extracting terms from the specific features of Derwent patents, including Abstract Use, Abstract Novelty, Abstract Detailed description, Abstract Description of drawing(s) and Tech Focus. Following that, we perform the term clumping process using the ClusterSuite [program developed by J.J. O'Brien, with Stephen J. Carley, at Georgia Tech –available at www.VPInstitute.org] to clean data and remove meaningless terms[4]. And then, the terms appearing only once are removed to improve operation efficiency. Finally, the cleaned terms are prepared to support the following analysis.

2.2 Stage 2 - Recognizing both technological topics and application topics

First, we merge the keywords extracted from abstract novelty, abstract detailed description, abstract description of drawing (s) and tech focus, and then construct a "Document-Terms" matrices as the base for generating technological topics. Meanwhile, another "Document-Terms" matrices for generating application topics is also constructed based on the keywords of Abstract Use. Based on these matrices, Latent Dirichlet Allocation (LDA) method is employed to generate latent topics and the topic distributions on patents because it has the best performance among several topic modeling algorithms when dealing with large-scale documents and interpreting latent topics[5][6]. we determine the appropriate number of topics for LDA-based topic modeling by calculating the perplexity[7], which is a popular indicator to measure the quality of probability model. Finally, according to the top keywords of each topics, engaging with experts' opinion, technological topics and application topics are recognized respectively.

2.3 Stage 3 - Identifying hot and potential technologies

The purpose of this stage is to identify hot and potential technologies by extracting technological topics from patents using the optimized model above and then assessing them from two aspects: "quantity" and "quality". For judge their "quantity", we mainly relied on the patent quantity indicators, including the quantity basic indicator and the quantity growth indicator. For judge their "quality", we propose a sentiment-based indicator - positive score, by utilizing sentiment analysis to gauge the attitude of domain experts toward a technology. Based on the two aspects, we set up a multi-dimensional evaluation system to comprehensive assess technologies.

2.3.1 The "quantity" assessment

Patent quantity is a common indicator used to measure the attention received by a technology because it reflects the R&D activity[1]. The patent quantity basic indicator, representing by the total patent number, reflects the overall attention received by a technology. And the patent quantity growth indicator, usually measuring by the change rate of patent number over time, presents the recent and even near future concern degree of a technology.

Since the LDA model is a soft clustering method that assign each document a topic distribution rather than a specific topic, we cannot directly use the patent number to measure technical topics. Instead, we took the total distribution weight (TDW), which was measured by summing the probabilities of patents distribution on a technological topic (Eq.(1)) (Jeong, et al.[5] to assess the overall attention received by the technology.

$$\text{Total distribution weight (TDW) of Topic k} = \sum_{i=1}^{M} w_{ik}, i = \{1, 2, \ldots, M\} \quad (1)$$

Where w_{ik} denotes the distribution probability of the i^{th} document on the topic k，M denotes the total number of documents in the corpus.

Similarly, we use the change rate of distribution weight (CRDW) over time to measure the recent and even near future concern degree of a technology(Chen, et al.[8]. The “Document-Topic” distribution Matrix is firstly ranked by application years of patent documents, as shown in Figure 2. Then we sum a group of elements in a column that are associated with patents published in the same year, and use the summation to present the annual distribution weight (ADW) of a topic. Based on the ADW, we calculate the CRDW indicator (Eq.(2)).

$$
\begin{array}{r|cccccc|l}
 & Topic1 & Topic2 & Topic3 & \cdots & Topic\ \mathrm{K} & & \\
Document1 & 0.0027 & 0.0382 & 0.0398 & \cdots & 0.0938 & \rbrace & \\
Document2 & 0.1903 & 0.0943 & 0 & \cdots & 0.0483 & \rbrace & Year1 \\
\vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \rbrace & \\
Document401 & 0.0982 & 0.0763 & 0.0387 & \cdots & 0.0273 & \rbrace & \\
Document402 & 0.0058 & 0.1983 & 0.2934 & \cdots & 0 & \rbrace & Year2 \\
\vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \rbrace & \\
Document851 & 0 & 0.0395 & 0.0964 & \cdots & 0.0498 & \rbrace & \\
\vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \rbrace & Year\ T \\
Document\ \mathrm{M} & 0.0049 & 0 & 0.0984 & \cdots & 0.1928 & \rbrace &
\end{array}
$$

Figure 1: An example of a topic distribution matrix in chronological order

$$\text{Change rate of distribution weight(CRDW) of Topic K} = \frac{ADW_k^{2018} + ADW_k^{2017} + ADW_k^{2016}}{ADW_k^{2015} + ADW_k^{2014} + ADW_k^{2013}} \quad (2)$$

Where ADW_k^T stands for the annual distribution weight of the topic k in the year T. Our study sets 2018 as the latest year for calculation because the DII data of 2019 is incomplete due to the time-lag of collection.

2.3.2 The "quality" assessment

Evaluating technologies with only the aid of patent quantity indicators may have some risks. For example, there is one technology which has a large number of patents with less novelty, while another has a few high-novelty patents. In this situation, only using patent quantity to compare the two technologies may lead to misjudgment. For this, we need the judgments of domain experts to evaluate a technology’s real value. Hence, we conduct a deep learning-based sentiment analysis over the short sentence of patents to gauge the attitudes of domain experts towards the technological topics[9][10][11]. First, we divide the full text of patents into sentences, and randomly select 10% of the sentences as the training set. Then, we assign them with a sentiment polarity (positive or neutral) and label them manually since there are seldom negative sentiment expressed in texts of patents. Second, the short sentences in the training set are segmented into words to form a corpus, and the word2vec of Python was used to train it to construct the word vectors, using the Continuous Bag-of-Word Model (CBOW). Third, we employ the LSTM neural network model[12-14] to build the classifier using the labelled short sentences in forms of the word vectors. When the accuracy reach to a specific threshold (above 80%), we think this classifier meet the requirements. Fifth, we apply the classifier to judge the sentiment polarities of all the sentences, and then calculate the positive score of a patent by counting the number of sentences with positive polarity in the patent. Final, we use the patent distribution on a topic to weight the positive scores of its associated patents and calculate the average positive score (APS) to reflect the common judgment of domain experts toward the topic (Eq.(3)).

$$\text{Average positive score (APS) of Topic k}=\frac{\sum_{i=1}^{M} w_{ik} p_i}{\sum_{i=1}^{M} w_{ik}},\ i=(1,2,\ldots,M) \quad (3)$$

Where w_{ik} denotes the distribution probability of the i^{th} patent document on the topic k, p_i denotes the positive scores of the i^{th} patent document, and M denotes the total number of patent documents in the corpus.

2.3.3 The multi-dimensional evaluation system

Based on the "quantity" and "quality" assessment, particular technologies with significant potential firstly rise to the surface with high values in most or all indicators. Then, by comprehensive analyzing the total distribution weight and the positive score, we can find the technologies which is valuable but underestimated in the past. Moreover, combing the change rate of distribution weight and the positive score, the technologies which is hot recently but the value still needs to be improved also can be identified.

2.4 Stage 4 – Exploring linkage between technologies and applications

We propose a probability-based topic relation measurement method to identify key relationships between technological topics and application topics. The topic relation measurement method is designed based on the assumption of LDA model that a document is related to a small number of topics rather than only one. For example, one document may associate a technological topic as well as an application topic with two different possibilities, here we present them as P(d=T) and P(d=A). Since our study separately implement the LDA processes to generate technological topics and application topics, it determines that the P(d=T) and P(d=A) are independent to each other. Hence, the possibility that the document is related to both the technological topic and the application topic can be calculated as:

$$P(d=T\cap A)=P(d=T)P(d=A) \quad (4)$$

We believe that the higher the possibility that two topics associate the same documents, the closer the relationships between these two topics. Based on the assumption, we propose the following probability-based method to measure relationship between topics (Eq.(5)).

$$\text{Relation weight between Topic k and Topic n}=\sum_{i=1}^{M} w_{ki} v_{ni},\ i=(1,2,\ldots,M) \quad (5)$$

Where w_{ki} denotes the probability that the i^{th} patent document associates the topic k, v_{ni} denotes the probability that the i^{th} patent document associates the topic n, $w_{ki}v_{ni}$ denotes the probability of the i^{th} patent document associates both the topic k and the topic n, and M denotes the total number of patent documents in the corpus.

Based on the method, we explore the connections between technologies and the technical problems and applications.

3 CASE STUDY

3.1 Raw Data Retrieval & Feature Extraction

We select dye sensitized solar cells (DSSCs) technique to implement the case study since our group has developed familiarities of this technology through a series of "tech mining" analyses before. We introduce the search terms from our previous research[15][16] and firstly retrieve 9,883 Derwent patents from 1991 to 2019. Then, we apply the

VantagePoint software to extract terms from the Title and the different parts of Abstract. Through implementing the term clumping and cleaning[4], we obtain 4802 terms in Title, 3892 terms in the Abstract Advantage, 3120 terms in the Abstract Use, and 5285 terms in the rest parts of the Abstract, including the Abstract Novelty, Detailed description and Description of drawing(s). Besides, we extract 4576 term from the Tech Focus of Derwent patents. To sum up, we obtain a total of 8587 terms.

3.2 Identifying hot and potential technologies of DSSCs

We recognize technological topics from the retrieved 9,883 Derwent patents of DSSCs. The optimized LDA model is firstly applied to generate 50 latent topics. We remove two topics that are hard to define, and then, with the experts' judgment, we combine the rest 48 topics and generate 27 technological topics. Based on the identified technological topics, we measure the extent of getting attentions of these topics from the two "quantity" indicators: TDW and CRDW, and evaluate their potential value using the "quality" indicator: APS. The methods of calculating the three indicators are introduced in the methodology part. Table 1 lists the indicator value of these 27 topics and marks the indicators greater than the average score.

Table 1: The Indicator Value Of The 27 Topics With The Marks

Topic	**Indicator value**			**Mark**		
	TDW	**CRDW**	**APS**	**TDW**	**CRDW**	**APS**
Organic dye	706	67.0%	2.69	√	√	√
Graphene & Carbon material	354	57.8%	2.56	√	√	√
Organic polymeric material	346	55.9%	2.95	√	√	√
TiO2	340	60.0%	2.48	√	√	√
Apparatus or power supply system containing DSSC	711	88.7%	2.44	√	√	
Photoanode modified method	520	88.4%	2.36	√	√	
Photoelectric conversion layer	677	34.6%	2.61	√		√
Light absorption layer	421	50.9%	2.64	√		√
Metal oxide semiconductor layer	392	33.3%	2.47	√		√
Glass type sealing material	386	44.2%	2.52	√		√
Metal catalyst	385	46.6%	2.85	√		√
Polymer electrolyte	497	55.4%	2.46	√		
Structure of solar cell	439	46.9%	2.23	√		
Metal substrate	410	35.9%	2.30	√		
Organic hole transport materials	179	98.2%	2.75		√	√
P-type/n-type semiconductor	154	74.7%	2.51		√	√
Preparation of nano materials	155	68.7%	2.35		√	
Sulfide for counter electrode	141	88.9%	2.44		√	
Electrode active metal material	135	50.7%	2.52			√
Conductive polymer	319	42.7%	2.26			
Laminated solar-cell module	284	42.3%	2.42			
Transparent conductive film	250	37.8%	2.45			
Organic solvent electrolyte	215	46.7%	2.35			
Metal complex dye	214	39.2%	2.21			
Lonic liquid electrolyte	204	51.8%	2.32			
ZnO	187	45.4%	2.28			
Metal oxide semiconductor particles	169	47.1%	2.38			
Average Scores of Indicator	340	55.5%	2.47			

From Table 1, we can see that there are four technological topics, including “Organic dye”, “Graphene & Carbon material”, “Organic polymeric material” and “TiO2”, have the relative higher scores of all the three indicators, which indicates that the four technologies of DSSCs have always been concerned from the past to the present, and their value have been recognized by most DSSC researchers. Specifically, from the results we can clearly see that the organic dye (706, 67.0%,2.69) has attracted more attention than its substitute - metal complex dye (204, 51.8%, 2.32). It has been considered to be the most promising dye sensitizer of DSSCs because of its characters of high molar extinction coefficient, low cost, easy modification of molecular and convenient synthesis[17]. Besides, “Graphene & Carbon material” is a popular material of DSSC electrode, which has advantage in low cost[18]. Moreover, graphene and carbon are also very valuable materials to produce photoelectrode with high surface area, which is beneficial for dye absorption[19]. “Organic polymeric material” and “TiO2” are another two important and promising basic materials for manufacturing DSSCs[20,21], and particularly, the organic polymeric material has the highest APS value.

Beyond that, we plot a TDW-APS two-dimensional scatter diagram (as shown in Figure 2), where the upper area distributes the technological topics with relative higher value and the right area spread the topics drawing more attention. From Figure 2, we discover that “Organic hole transport materials”, “P-type/N-type semiconductor” and “Electrode active metal material” are three valuable sub-technologies for advancing DSSCs, which were underestimated in the past. It indicates that these technologies have the potential to become new hotspots in DSSCs research in the future and provide innovation opportunities for the evolution of DSSC technology.

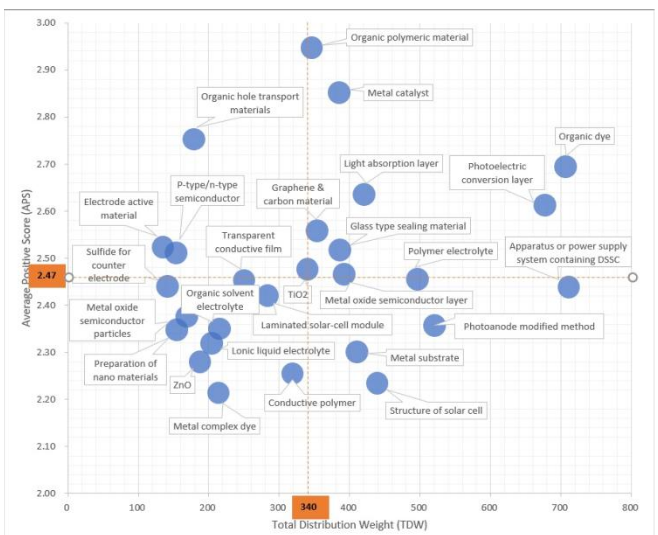

Figure 2: The TDW-APS two-dimensional scatter diagram for the technical topics of DSSCs

Besides, we also draw a CRTW-APS scatter diagram (as shown in Figure 3) to help discover technical topics that draw increasing attention and need to break through. As we can see from the figure 3, four technologies, including “Apparatus or power supply system containing DSSC”, “Sulfide for counter electrode”, “Photoanode modified method”, and “Preparation of nano materials” have been found in the lower right area. “Apparatus or power supply system containing DSSC” is an important application research of DSSCs. The rise of this research topic indicates that DSSC technology is becoming mature and move towards to commercialization. However, the relatively low positive score implies that there are still some developmental hurdles to overcome in the way of productization and commercialization.

Besides, “Sulfide for counter electrode” have attracted much attention recently due to the unique physical and chemical properties of sulfide such as high energy density and strong optical properties[22]. Comparing to traditional Pt counter electrode, sulfide counter electrode has advantages in low cost and easy to prepare, however, its catalytic mechanism is not clear by now and need further theoretical research and experiments to discover more novel properties[23]. In addition, “Photoanode modified method” and “Preparation of nano materials” are two important sub-technologies of DSSC photoanode advancing photoanode of DSSCs. The photoanode modified method can reduce the interface impedance between anode and dye[24], and improved nanostructures can obtain larger surface area and higher porosity[24,25], which are all beneficial to advancing DSSC photoanode technique and improving photoelectric conversion efficiency of DSSCs. However, advancing DSSC anode technique is the most complicated problem of DSSCs and still faces many technical challenges, such as the contradiction between carrier diffusion length and optical absorption efficiency that exists when preparing nano materials[26].

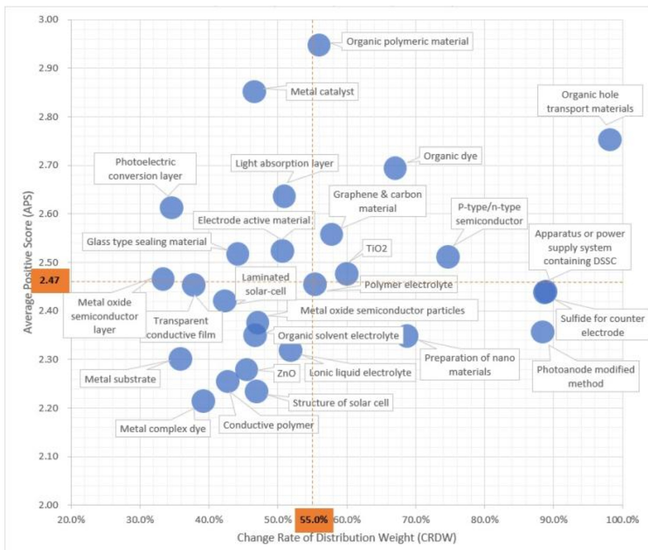

Figure 3: The CRDW-APS two-dimensional scatter diagram for the technical topics of DSSCs

3.3 Discovering key applications of DSSCs linking with sub-technologies

We separately extract 3892 terms and 3120 terms from the Advantage parts and the Use parts of the DSSC patent abstracts, based on which we finally recognize 25 topics on applications1 of DSSCs (shown in Table 2).

From the Table 2 and Figure 4, we can see that some sub-technologies of DSSCs can be applied to the other kinds of batteries, including fuel cell, lithium ion battery, silicon solar cell, organic thin-film solar cell and CdTe/CIGS solar cell. Except these, “Photosensor”, “Organic Light Emitting Diode (OLED)” and “Building” have the highest weights2, which indicate that these are likely to be the most potential applications of DSSCs in the future. Moreover, the Figure 6 shows that the sub-technologies of “Photoelectric conversion layer” and “Organic Dye” are both have great application

1 We firstly generate 28 topics on applications and then remove three too general topics which includes “Photovoltaic cells”, “Photo conversion element”, and “Dye sensitized solar cell modules”.

2 In the Table 2, “Photocatalyst film”, “Sensitizing dye”, “Carbon counter electrode” and “Polymer electrolyte” are the topics on the component technologies of DSSCs.

opportunities in the field of photosensor and OLED, and “Organic hole transport materials” also can be applied to OLED. Besides, “Apparatus or power supply system containing DSSC” can be used to assist the construction of energy-saving building[27], and it also could be applied to vehicle and portable electronic devices, such as mobile telephone, computer and calculator et al., as auxiliary power supply.

Table 2: The Topics On The Applications Of Dssc Sub-Technologies

No.	Topics	TDW	No.	Topics	TDW
1	Fuel cell/lithium ion batter	385	14	Display device	149
2	Photosensor	343	15	Solid-state DSSC	133
3	Photocatalyst film	332	16	Polymer electrolyte	130
4	OLED	298	17	Organic photoelectric conversion element	113
5	Building	295	18	Drug delivery system	107
6	Silicon solar cell	244	19	Biosensor/electrochemical sensor	105
7	Sensitizing dye	239	20	CdTe/CIGS solar cell	105
8	Photoelectric transducer	205	21	Vehicle	103
9	Sealing application	198	22	Portable electronic product	102
10	Carbon counter electrode	192	23	Gas/humidity/molecule sensor	98
11	Environment purification	187	24	Drugs or cosmetics	85
12	Organic thin-film solar cell	174	25	Image sensors	84
13	Adhensive/coating/packaging material	159			

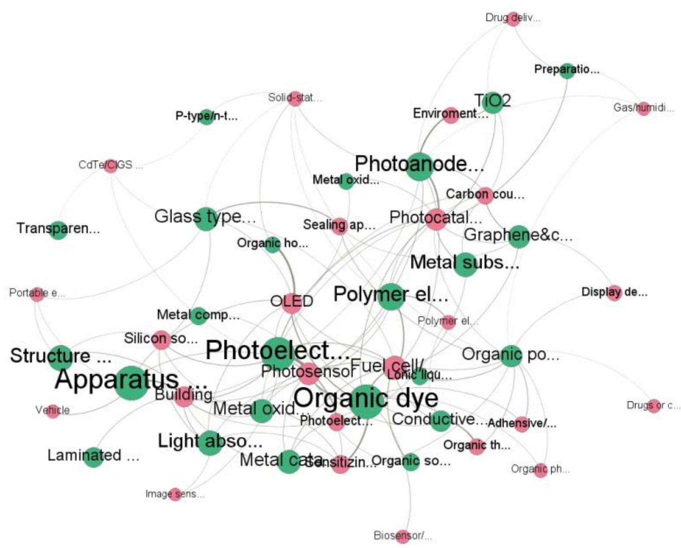

Figure 4: The applications linking to the sub-technologies of DSSCs

Beyond that, we discover some unexpected interesting possible applications, including “Environment purification”, “Drug delivery system” and “Drugs or cosmetics”. Specifically, “Photoanode modified method” and “TiO2” technologies have application opportunities in the field of “Environment purification” because that the “Photoanode modified method” focus on producing porous semiconductor nanomaterials which have advantages in absorbing harmful substances from water and air, and similarly, the porous TiO2 also can be used to purifying environment[28]. Besides, the “Photoanode modified method” and “TiO2”, as well as “Preparation of nano materials” have another application possibility - “Drug delivery system”. TiO2 and some other nano materials can be used as a drug carrier for drug delivery, and the TiO2 is considered to be a potential photodynamic therapy material recently[29,30]. In addition, “Organic polymeric material”, which is an important composition of polymer electrolyte of DSSCs, is also a potential technology to be applied in the field of “Drug or cosmetics”, where it is useful for the therapeutic and/or cosmetic use of skin diseases.

4 CONCLUSION

There are three main contributions in our study. First, we identified both technological topics and application topics by mining different parts of Derwent patent. Second, we propose a probability-based topic relation measurement method to link the technical applications with the core technologies of NETs. By doing so, more detailed information can be exploited to assist in discovery of important and potential application opportunities. Third, we introduce sentiment analysis to support topic assessment and construct a multi-dimensional evaluation system to identify potential innovation opportunities from both the "quantity" and "quality" aspects. It enables more scientific and comprehensive evaluation, reducing the probability of misjudgment[31].

There are also some parts should be improved in the further. First, the performance of the LDA model on recognizing topics from short texts should be validated. Thus, in next step we can employ several candidate topic modeling methods to select the optimal model. Second, the proposed probability-based topic relation measurement method can only identify links between topics from a quantitative perspective, but cannot reveal how these topics correlated with each other. The “Subject-Action-Object” (SAO) semantic analysis has advantage in exploring semantic relationship between topics[32], so how to integrate semantic analysis with our method to profile in-depth relationships between topics is another future work of our study.

ACKNOWLEDGMENTS

This work was supported by the Science Foundation of Ministry of Education of China [grant numbers 18YJC870014]; the Chinese National Science Foundation for Young Scholars [grant number 71804016, 71704139]; the Beijing Municipal Commission of Education [grant number SM201910037005]; and Youth Scientific Research Fund Project of Beijing Wuzi University [grant number, 2022XJQN11].

REFERENCES

[1] Xiao Zhou, Lu Huang, Alan Porter & Jose M. Vicente-Gomila. 2019 .Tracing the system transformations and innovation pathways of an emerging technology: Solid lipid nanoparticles. Technological Forecasting & Social Change 146(SEP 2019), 785-794. https://doi.org/10.1016/j.techfore.2018.04.026.

[2] Alan L. Porter & Michael J. Detampel. 1995 .Technology opportunities analysis. Technological Forecasting & Social Change 49, 3(JUL 1995), 237-255. https://doi.org/10.1016/0040-1625(95)00022-3.

[3] Tingting Ma, Yi Zhang, Lu Huang... & Donghua Zhu. 2016 .Text mining to gain technical intelligence for acquired target selection: A case study for China's computer numerical control machine tools industry. Technological Forecasting & Social Change 116(MAR 2017), 162-180. https://doi.org/10.1016/j.techfore.2016.10.061.

[4] Yi Zhang, Alan L. Porter,Zhengyin Hu... & Nils C. Newman. 2014 .“Term clumping” for technical intelligence: A case study on dye-sensitized solar cells. Technological Forecasting & Social Change 85(JUN 2014), 26-39. https://doi.org/10.1016/j.techfore.2013.12.019.

[5] Byeongki Jeong, Janghyeok Yoon & Jae-Min Lee. 2019 .Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. International Journal of Information Management 48(OCT 2019), 280-290. https://doi.org/10.1016/j.ijinfomgt.2017.09.009.

[6] Chiru C, Rebedea T & Ciotec S. 2014 .Comparison between LSA-LDA-Lexical Chains. In the Proceedings of the 10th International Conference on Web Information Systems and Technologies, 3-5 April, 2014, Barcelona, Spain, 255-262.

[7] David M. Blei,Andrew Y. Ng & Michael I. Jordan. 2003 .Latent Dirichlet Allocation. Journal of machine learning research 3, 4-5(MAY 15 2003), 993-1022.

[8] Hongshu Chen, Guangquan Zhang, Donghua Zhu & Jie Lu. 2017 .Topic-based Technological Forecasting Based on Patent Data: A Case Study of Australian Patents from 2000 to 2014. Technological Forecasting & Social Change 119(JUN 2017), 39-52. https://doi.org/10.1016/j.techfore.2017.03.009.

[9] Hassan A & Mahmood A. 2017 .Deep Learning Approach for Sentiment Analysis of Short Texts. In the Proceedings of the 3rd IEEE International Conference on Control, Automation and Robotics (ICCAR), APR 22-24, 2017, Nagoya, JAPAN, 705-710.

[10] Jurgovsky J & Granitzer M. 2015 . Comparing Recursive Autoencoder and Convolutional Network for Phrase-Level Sentiment Polarity Classification. In the Proceedings of the 20th International Conference on Applications of Natural Language to Information Systems, JUN 17-19, 2015, Univ Passau, Passau, GERMANY, 160-166.

[11] Daekook Kang & Yongtae Park. 2014 .Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. Expert Systems With Applications 41, 4(MAR 2014), 1041-1050. https://doi.org/10.1016/j.eswa.2013.07.101.

[12] Yuan Luo. 2017 .Recurrent neural networks for classifying relations in clinical notes. Journal of Biomedical Informatics 72(AUG 2017), 85-95. https://doi.org/10.1016/j.jbi.2017.07.006.

[13] Martin Woellmer, Florian Eyben, Alex Graves, Bjoern Schuller & Gerhard Rigoll. 2010 .Bidirectional LSTM Networks for Context-Sensitive Keyword Detection in a Cognitive Virtual Agent Framework. Cognitive computation 2, 3(SEP 2010), 180-190. https://doi.org/10.1007/s12559-010-9041-8.

[14] Ruben Zazo, Alicia Lozano-Diez, Javier Gonzalez-Dominguez... & Joaquin Gonzalez-Rodriguez. 2017 .Language Identification in Short Utterances Using Long Short-Term Memory (LSTM) Recurrent Neural Networks. PLoS ONE 11, 1(JAN 2016). https://doi.org/10.1371/journal.pone.0146917.

[15] Tingting Ma, Alan L. Porter, Ying Guo, Jud Ready, Chen Xu & Lidan Gao. 2013 .A technology opportunities analysis model: applied to dye-sensitised solar cells for China. Technology Analysis & Strategic Management 26, 1(JAN 2 2014), 87-104. https://doi.org/10.1080/09537325.2013.850155.

[16] Allen H. Huang, Reuven Lehavy, Amy Y. Zang & Rong Zheng. 2017 .Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach. Management Science 64, 6(JUN 2018), 2833-2855. https://doi.org/10.1287/mnsc.2017.2751.

[17] S. Krishnan & K. Senthilkumar. 2018 .Theoretical probe on modified organic dyes for high-performance dye-sensitised solar cell. Current Applied Physics 18, 9(SEP 2018), 1071-1079. https://doi.org/10.1016/j.cap.2018.05.025.

[18] Chenjing Gao, Hongrui Wang, Qianji Han... & Mingxing Wu. 2018 .High-efficiency magnetic carbon spheres counter electrode for dye-sensitized solar cell. Electrochimica Acta 264(FEB 2018), 312-318. https://doi.org/10.1016/j.electacta.2018.01.134.

[19] Safia A. Kazmi, Salman Hameed, Arham S. Ahmed... & Ameer Azam. 2017 .Electrical and optical properties of graphene-TiO 2 nanocomposite and its applications in dye sensitized solar cells (DSSC). Journal of Alloys and Compounds 691(JAN 2017), 659-665. https://doi.org/10.1016/j.jallcom.2016.08.319.

[20] Zhang Xiaobo. 2018 .Enhancing Natural BChl a Adsorption Capacity and Photoelectric Performance of BChl a-based DSSC by Improving TiO2 Photoanode. International Journal of Electrochemical Science 13, 7(JUL 2018), 6598-6607. https://doi.org/10.20964/2018.06.63.

[21] I. Kataoka, S. Yamada, H. Shiotsuka, and H. Zenko. 2011 ."Solar cell module with polymeric resin sealing layers," ed: EP.

[22] Wu Mingxing, Wang Yudi, Lin Xiao... & Ma Tingli. 2011 .Economical and effective sulfide catalysts for dye-sensitized solar cells as counter electrodes. Physical chemistry chemical physics : PCCP 13, 43(2011), 19298-19301. https://doi.org/10.1039/c1cp22819f.

[23] Yang Jie, Bao Chunxiong, Zhu Kai... & Zou Zhigang. 2014 .High catalytic activity and stability of nickel sulfide and cobalt sulfide hierarchical nanospheres on the counter electrodes for dye-sensitized solar cells. Chemical communications (Cambridge, England) 50, 37(2014), 4824-4826. https://doi.org/10.1039/c4cc00001c.

[24] Hai Ying Shi,Jun Qing Tian & Wei Zheng. 2014 .Dye-Sensitized Solar Cells Assembled with Modified Photoanode and Carbon Nanotubes as Counter Electrode. Advanced Materials Research 977(2014), 55-58. https://doi.org/10.4028/www.scientific.net/AMR.977.55.

[25] Md Zaved H Khan & Xiuhua Liu. 2019 .Role of Nanostructured Photoanode and Counter Electrode on Efficiency Enhancement of DSSCs. Journal of Electronic Materials 48, 7(JUL 2019), 4148-4165. https://doi.org/10.1007/s11664-019-07212-8

[26] Qiuping Liu, Yang Zhou, Yandong Duan... & Yuan Lin. 2013 .Improved photovoltaic performance of dye-sensitized solar cells (DSSCs) by Zn+Mg co-doped TiO 2 electrode. Electrochimica Acta 95(APR 2013), 48-53. https://doi.org/10.1016/j.electacta.2013.02.008.

[27] Dr. Andrea Reale, Dr. Lucio Cinà, Dr. Ambra Malatesta... & Prof. Aldo Di Carlo. 2014 .Estimation of Energy Production of Dye-Sensitized Solar Cell Modules for Building - Integrated Photovoltaic Applications. Energy Technology 2, 6(JUN 2014), 531-541. https://doi.org/10.1002/ente.201402005.

[28] C.H. Ao & S.C. Lee. 2004 .Indoor air purification by photocatalyst TiO2 immobilized on an activated carbon filter installed in an air cleaner. Chemical Engineering Science 60, 1(JAN 2005), 103-109. https://doi.org/10.1016/j.ces.2004.01.073.

[29] Lai Shuting, Zhang Wei, Liu Fang... & Zhou Wuyi. 2013 .TiO2 nanotubes as animal drug delivery system and in vitro controlled release. Journal of nanoscience and nanotechnology 13, 1(JAN 2013), 91-97. https://doi.org/10.1166/jnn.2013.7086

[30] Guilong Zhang, Lukui Chen, Xiaoyuan Guo... & Ning Gu. 2017 .Nanoparticle-mediated Drug Delivery Systems (DDS) in the Central Nervous System. Current Organic Chemistry 21, 3(2017), 272-283. https://doi.org/10.2174/1385272820666161017170325

[31] Xiao Zhou, Ying Guo, Fangshun Li, Jin Wang, Huanan Wei, Miaomiao Yu & Siliang Chen. 2020 .Identifying and Assessing Innovation Pathways for Emerging Technologies: A Hybrid Approach Based on Text Mining and Altmetrics. IEEE Transactions on Engineering Management 68, 5(OCT 2021), 1360-1371. https://doi.org/10.1109/tem.2020.2994049.

[32] Xuefeng Wang, Pingping Ma, Ying Huang... & Zhinan Wang. 2017 .Combining SAO semantic analysis and morphology analysis to identify technology opportunities. Scientometrics 111, 1(APR 2017), 3-24. https://doi.org/10.1007/s11192-017-2260-y.