The Hotspot Technologies and Cutting-edge Technologies of Organic Solar Cells

Q. Chen*‡, F. Chen*

* National Science Library, Chinese Academy of Sciences, 100190, Beijing, China

** University of Chinese Academy of Sciences, 100049, Beijing, China

(chenqm@mail.las.ac.cn, chenfang@mail.las.ac.cn)

‡ Corresponding Author; Qimei Chen, National Science Library, Chinese Academy of Sciences, 100190, Beijing, Tel: +86 10 82626611, Fax: +86 10 82626611, chenqm@mail.las.ac.cn

Received: 30.11.2019 Accepted:28.06.2020

Abstract- The objective of this paper is to identify and forecast the technological development of organic solar cells by the analysis of patents. This paper researches the unsupervised machine learning methods and complex network theory. Based on the Python language, the K-means++ algorithm is used to cluster the data for discovering hotspot technologies, the theory of structure holes is used to identify potential cutting-edge technologies. The results indicate the hotspot technologies of organic solar cells are mainly involved in the five fields: device architecture and morphology control, electrode and electron transport layer design, active layer design, interface materials and fabrication, battery assembly and packaging process. And the potential cutting-edge technologies include: fabrication and morphology of active layer, device structure and interface engineering, transparent electrodes and substrates, donor and acceptor materials, and so on. The proposed analysis framework in this paper is applicable to different science and technology domains.
Keywords- Organic solar cells; technological forecasting; patent text-mining; machine learning; complex networks

Özet- Bu yazının amacı, patentlerin analizi ile organik güneş pillerinin teknolojik gelişimini tanımlamak ve tahmin etmektir. Bu makale, denetimsiz makine öğrenme yöntemlerini ve karmaşık ağ teorisini araştırmaktadır. Python diline dayanarak, sıcak nokta teknolojilerini keşfetmek için verileri kümelemek için K-means ++ algoritması kullanılır, potansiyel delik teknolojilerini tanımlamak için yapı delikleri teorisi kullanılır. Sonuçlar organik güneş pillerinin sıcak nokta teknolojilerinin temel olarak beş alanda yer aldığını göstermektedir: cihaz mimarisi ve morfoloji kontrolü, elektrot ve elektron taşıma katmanı tasarımı, aktif katman tasarımı, arayüz malzemeleri ve imalatı, pil montajı ve paketleme işlemi. Ve potansiyel en son teknolojiler şunları içerir: aktif katmanın imalatı ve morfolojisi, cihaz yapısı ve arayüz mühendisliği, şeffaf elektrotlar ve substratlar, verici ve alıcı malzemeler, vb. Bu makalede önerilen analiz çerçevesi farklı bilim ve teknoloji alanlarına uygulanabilir.
Anahtar Kelimeler- Organik güneş pilleri; teknolojik tahmin; patent metni madenciliği; makine öğrenme; karmaşık ağlar

1. Introduction

The global use of solar photovoltaic system is accelerating rapidly due to the ever-decreasing cost and improvement on cell efficiency [1]. Solar cell, which generates the electricity directly from sunlight, is expected to play a major role in solving the global energy crisis in an environmental-friendly and sustainable way. Among them, organic solar cells have attracted a great deal of attentions owing to several advantages, including wide range of applicable materials, low-cost, and compatible with flexible substrates [2]. Organic solar cell a third generation photovoltaic cell, consists of an organic photovoltaic active layer placed between a transparent electrode and a metal electrode [3].

With the continuous progress of society, unprecedented achievements have been made in scientific and technological innovation in various fields, and a large number of patent documents have also been produced and accumulated. Patent is the largest technical information source in the world and the carrier of scientific and technological knowledge. Patent applications and registrations in the domain of organic solar cells have been increasing rapidly over the past two decades.

Patent text-mining is a kind of effective method for decision-making of technological development. With the explosive growth of the number of patents, the massive patents have been time-consuming and labor-intensive, and even beyond the processing limit of human beings. How to quickly filter out useful information and the internal patterns of data from this information has become a difficult problem in front of us. Therefore, it is necessary to develop intelligent data analysis methods.

Python is an interpretable, high-level, general-purpose programming language. One of the areas where Python excels at is analysis of data and visualization. Based on the Python language, this paper uses the NLTK package to preprocess the patent text data, uses the TF-IDF method to vectorize the text. The K-means++ algorithm is used to cluster the data. The NetworkX library is used to create, manipulate graphs and analyze networks. The theory of structure holes is used to identify key nodes.

2. Methodology

In order to identify the hotspot technologies and cutting-edge technologies in the field of organic solar cells, unsupervised machine learning methods and complex network theory were used to analyze the patent data. Figure 1 demonstrates the flow work chart throughout the process from the beginning of data retrieval, to the end of results and conclusions, and steps in between. Several common and important methods during the main current were employed, such as VSM and TF-IDF used to represent the patent text, K-means++ in SKlearn for clustering, algorithms in NetworkX for structural holes. Then the open-source software, "Gephi", is employed to visualize the clustering results.

2.1. Data retrieve

This paper based on the keywords retrieval to build patent retrieval strategy. The first step is to clear target field and select the appropriate patent database (Derwent, Patentics, Incopat, etc.). Then, the International Patent Classification (IPC), Derwent Manual Code (MC) and terms of the patent documents are combined to build patent retrieval strategy.

2.2. Text representation model

Text representation is the most important part during the text mining mission, also for the technology topic detection. First and foremost, the raw text materials from Derwent patent data have advantages because of its deep tagged text content, which includes the fields like “Abstract”, “Use”, ”Novelty”, “Technology Focus”, etc. In order to express the patent technologies precisely, discarding redundant patent text, it’s prior to select the fields of “Use” and ”Novelty”, and “Abstract” as a remedy if they are empty. As for tokenization, NLTK is used and Snow-ball-Stemmer algorithm is imported to implement the NLP mission of raw text. Although lemmatization methods are also popular, they showed some shortcomings in this project, probably because of many polysemy phenomena.

VSM is currently one of the most widely used and better performing topics representation methods. In VSM, the feature representation method of topics is TF-IDF algorithm, and the feature selection method of topics is weight of evidence for text. In VSM, texts are seen as vector space composed of a set of orthogonal term vectors. After words tokenization, stop words are removed, and then word frequency is calculated and transformed into TF-IDF. Vector matrix is created by algorithms from the python package of SKlearn. If the total number of features in the data set is n, it constitutes an n-dimensional vector space [4].

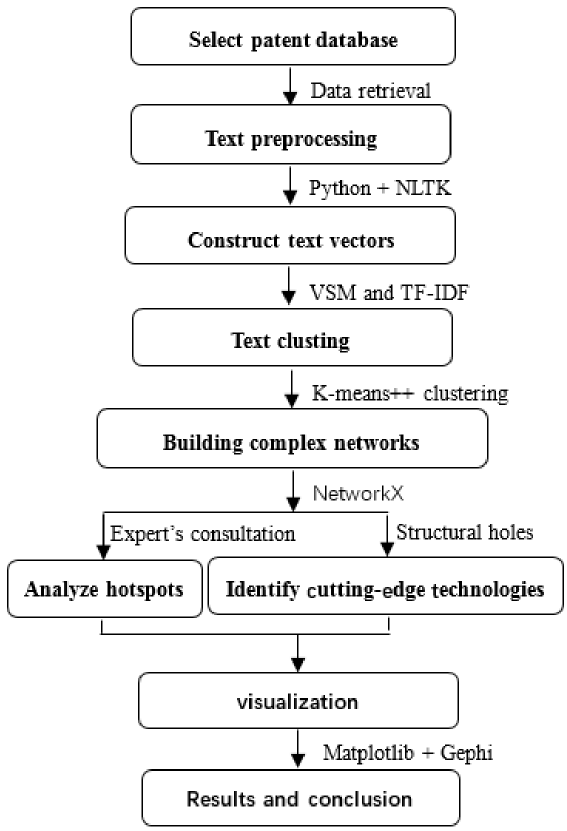

Figure 1. Schematic of the Proposed Text Mining and Visualization Approach

2.3. K-means++ algorithm for text clustering

The K-Means algorithm is a commonly used clustering algorithm. The algorithm works using definite data set. Predefined clustering centroids are used to perform clustering operation. Centroids should be selected as far from one another as possible to obtain a valid clustering. In the other step, each data set point is correlated with the nearest centroid. The preceding two steps are reduplicated until centroids are stable. Due to its simplicity and versatility, k -means remains popular since it was firstly used. But the algorithm itself has certain problems. For example, the long calculation time under large data volume or undesired clustering results. The k-means++ algorithm eliminates the above-mentioned drawbacks by using a better algorithm for choosing the initial center guesses [5]. So, in this paper, K-means++ algorithm is adopted.

2.4. Complex network analysis in python

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. NetworkX is written in Python and can be used to generate graphs, compute centrality measures and path length-based properties as well as graph components and graph motifs. The global topics including the hotspot and potential cutting-edge technologies can be connected through the complex network, concretely demonstrated by the graphs. Further analysis including the degree analysis, modularity clustering, structural hole and link prediction can be implemented.

The network graphs can be visualized by matplotlib, Graphviz etc, Gephi is used in this paper to show the final graph though. The nodes and edges are showed through the more exquisite visualization solution.

2.5. Recognition and analysis of structural hole

A structural hole (SH) can be understood as a gap between two unconnected nodes. When the two nodes are connected by a third one, the gap is filled and creates important advantages for the bridging node. Therefore, the hole (bridging node) has access to different flows of information and get more benefits from its neighbors. The major indices for structural hole, such as efficiency, constraint, degree, are calculated through the algorithms in NetworkX. As the theory explains, the higher the network constraint index, the smaller the structural holes have. Conversely, the smaller the network constraint index is, the more important position the node has [6]. Based on this, the potential cutting-edge technologies were identified.

3. Data Sources and Technical Analysis

3.1. Data Sources

In this paper, the patent data comes from Derwent Innovation Index (DII) patent database, which contains more than 40 patent office authorized the basic invention. The search strategy was based on the association of keywords related to organic solar cells in the topical field, considering the International Patent Classification of the WIPO and Derwent Manual Code. Patent families with application years (referring to priority years) of 2016 to 2019 are retrieved. The cut-off date of data retrieval is May 23, 2019. A total of 3879 patent families were retrieved.

3.2. Technical Analysis

Through the above data analysis processing, we get 40 clusters and their descriptions. With manual cleaning, the same topics were merged. Finally, 34 hotspot technologies were show in Table 1.

Table 1 . The hotspot technologies of organic solar cells patents

Number	**Hotspot technologies**	**Quantity of patents**
1.	Fabrication and morphology of active layer	286
2.	Device architecture	233
3.	Organic photovoltaics donor and acceptors materials	195
4.	Transparent electrodes and substrates	178
5.	interface engineering	159
6.	Application and fabrication of organic photovoltaics	158
7.	Power conversion efficiency	157
8.	Optoelectronic properties in organic solar cells	125
9.	Composition in active layers	123
10.	Substrates	122
11.	Preparation technology of inorganic transport layers	118
12.	Active layer preparation and solvent engineering	115
13.	Solution process and fabrication of active layer	115
14.	Electrode materials	113
15.	Application of film organic photovoltaics	95
16.	electron transport layer	94
17.	encapsulating material	92
18.	Synthesis of donor and acceptors materials	90
19.	Solar cell models and fabrication	89

20.	Organic transistor and light-emitting element	83
21.	Active layer materials and fabrication	81
22.	Preparation process of conductive material	81
23.	Display technology	81
24.	morphology additives	80
25.	Energy storage	79
26.	Organic optoelectronic devices and architecture	76
27.	Other conductive materials	64
28.	Interfacial transport layers	60
29.	Optoelectronic devices	56
30.	Synthesis of donor and acceptors materials in active layer	54
31.	large-area preparation	51
32.	Organic optoelectronic devices and display	50
33.	Polymer Synthesis	40
34.	metal oxide transport layers	27

These hotspot technologies are mainly involved in the five fields: device architecture and morphology control (#A), electrode and electron transport layer design (#B), active layer design (#C), interface materials and fabrication (#D), battery assembly and packaging process (#E). The main clusters were visualized in Figure 2.

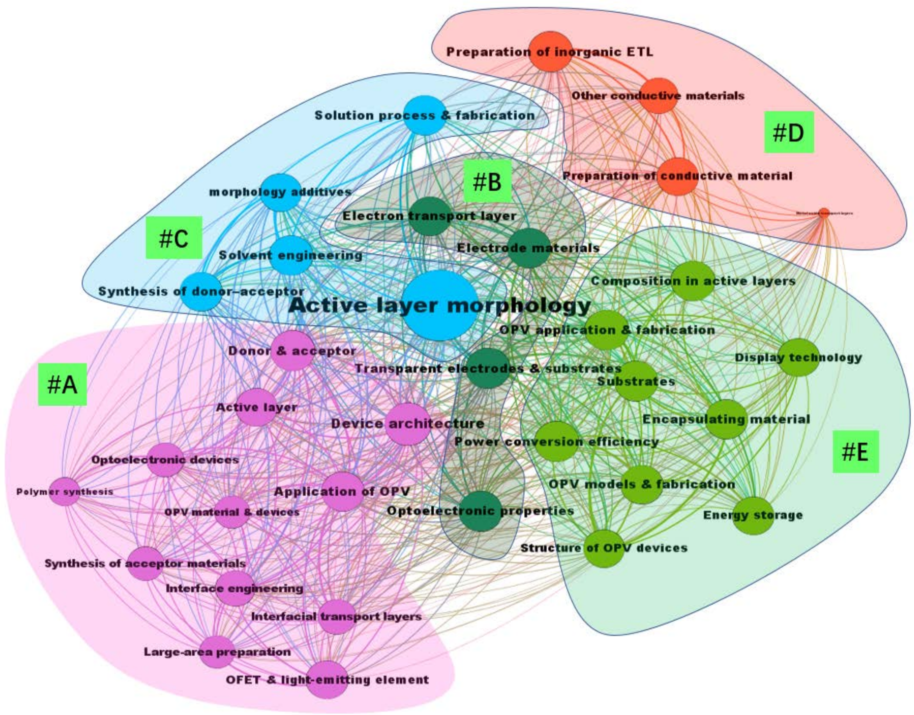

Figure 2. The Cluster Analysis of Organic Solar Cells Patents.

Table 2. presents the details, for each technology concentration area, of the main classifications identified.

Table 2. The topical cluster analysis of organic solar cells patents

Classification	**Hotspot technologies**	**Quantity of patents**
#A Device	A1: Device architecture;	1157

architecture and morphology control	A2: Donor and acceptor materials; A3: Application of film organic photovoltaics; A4: Organic transistor and light-emitting element; A5: Active layer materials and fabrication; A6: interface engineering; A7: Interfacial transport layers; A8: Optoelectronic devices; A9: Synthesis of acceptor materials in active layer; A10: large-area preparation; A11: Organic optoelectronic material devices and display; A12: Polymer Synthesis.	
#B Electrode and electron transport layer design	B1: Transparent electrodes and substrates; B2: Optoelectronic properties in organic solar cells; B3: Electrode materials; B4: electron transport layer.	510
#C Active layer design	C1: Active layer morphology; C2: active layer preparation and solvent engineering; C3: Solution process and fabrication of active layer; C4: Synthesis process of donor and acceptor; C5: morphology additives.	686
#D Interface materials and fabrication	D1: Preparation technology of inorganic transport layers; D2: Preparation process of conductive material; D3: Other conductive materials; D4: metal oxide transport layers.	290
#E	E1: Application and fabrication of organic	977

Battery assembly and packaging process	photovoltaics; E2: Power conversion efficiency; E3: Composition in active layers; E4: Substrates; E5: encapsulating material; E6: Solar cell models and fabrication; E7: Display technology; E8: Energy storage; E9: Structure of organic optoelectronic devices.	

Structural holes are "an important extension of social network theory". This theory argues that the stronger one's connections are to the central players, especially in large networks, the more certainty that one is within the stream where relevant information flows [7]. According to the theory of structural holes, the potential cutting-edge technologies include: fabrication and morphology of active layer, device structure and interface engineering, transparent electrodes and substrates, donor and acceptor materials, and so on. Table 3 presents the Top 10 potential cutting-edge technologies.

Table 3. The topical cluster analysis of organic solar cells patents (Structural holes index equals ln(constraint))

Number	**cutting-edge technologies**	**Average degree**	**Structural holes**	**Quantity of patents**
1.	Fabrication and morphology of active layer	126.254	-2.119	286
2.	Device architecture	58.491	-2.045	233
3.	Organic photovoltaics donor and acceptors materials	64.43	-2.014	178
4.	Transparent electrodes and substrates	45.207	-1.973	195
5.	interface engineering	30.991	-1.858	81
6.	Application and fabrication of organic photovoltaics	31.511	-1.85	113
7.	Power conversion efficiency	26.84	-1.836	158
8.	Optoelectronic properties in organic solar cells	21.79	-1.819	125
9.	Composition in active layers	16.74	-1.817	157
10.	Substrates	23.593	-1.805	94

4. Conclusion

For a long time, Solar cells are mostly prepared on the basis of inorganic materials such as crystalline silicon. Inorganic solar cells have the disadvantages of complicated process, high cost, large energy consumption and heavy pollution. Now, people contribute to develop the new type of solar cell with low cost, high efficiency, flexibility and environmental friendliness.

Organic solar cells have the advantages of structure diversity, large-area and low-cost printing preparation, flexibility, translucency or even transparency material. In addition to being a normal power generation device, it has great application potential in other fields such as energy-efficient buildings, wearable devices and mobile electronic charging equipment, which has aroused great interest in industrial and academic circles [8].

The bottleneck restricting the development of organic solar cells is the photoelectric conversion efficiency. The preparation of a solution-processable active material with high efficiency, low cost, and good reproducibility is the basis for improving the photoelectric conversion efficiency.

Photovoltaic technology applications take into account many indicators such as efficiency, cost and longevity. Efficiency is always the first consideration. In the past few years, although organic solar cells technology has developed rapidly, the photoelectric conversion efficiency has exceeded 14%, but the efficiency is still lower than other types of solar cells [9] [10]. Therefore, how to take advantage of organic materials, by optimizing material design and improving battery structure and preparation process, to obtain higher photoelectric conversion efficiency has attracted much attention [11]. Technology products based on good foldable and flexible carbon materials will be a foreseeable development direction for the discipline of materials science.

Acknowledgements

We would like to give our sincere gratitude to Prof. Guo Yunlong and Prof. Meng Lei. They gave our great help by providing necessary materials, advice of great value and inspiration of new ideas.

References

[1] S. Gautam, D. B. Raut, P. Neupane, D. P. Ghale and R. Dhakal, "Maximum power point tracker with solar prioritizer in photovoltaic application", 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA), Birmingham, pp. 1051-1054, November 2016.

[2] C. Choe, D. Lee, I. Seo, H. Kim, “Patent citation network analysis for the domain of organic photovoltaic cells: Country, institution, and technology field”, Renewable and Sustainable Energy Reviews, vol. 26, pp. 492–505, October 2013.

[3] O. Eyecioglu, M. Beken, O. Yagcı and O. Icelli, "Effect of Boric Acid Doped Pedot:PSS Layer on the Photovoltaic Parameters of P3HT:PCBM Based PV Cell," 2018 7th International Conference on Renewable Energy Research and Applications (ICRERA), Paris, pp. 973-975, October 2018.

[4] L. Shengdong, L. Xueqiang, W. Tao, S. Shuicai, “The Key Technology of Topic Detection Based on K-means”, 2010 International Conference on Future Information Technology and Management Engineering, pp. 387-390, October 2010.

[5] M. M. Öztürk, U. Cavusoglu, A. Zengin, “A novel defect prediction method for web pages using k-means++”, Expert Systems with Applications, vol. 42, pp. 6496-6506, 2015.

[6] Y. Hui, C. Xi, L. Zun, L. Yongjun, “Identifying key nodes based on improved structural holes in complex networks”, Physica A: Statistical Mechanics and its Applications, vol. 486, 318-327, 2017.

[7] R. S. Burt, “Structural Holes: The Social Structure of Competition”, University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship, 1992.

[8] M. M. G. Lawan, J. Raharijaona, M. B. Camara and B. Dakyo, "Power control for decentralized energy production system based on the renewable energies — using battery to compensate the wind/load/PV power fluctuations," 2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA), San Diego, CA, USA, pp. 1132-1138, November 2017.

[9] Y. Mochizuki and T. Yachi, "Effective Series-Parallel Cell Configuration in Solar Panels for FPM Power Generation Forest," 2018 7th International Conference on Renewable Energy Research and Applications (ICRERA), 2018, pp. 294-300, October Paris.

[10] H. Sun, R. Zhang, S. Lin, K. Yang, "Perovskite solar cells employing Al2O3 scaffold layers," 2014 International Conference on Renewable Energy Research and Application (ICRERA), Milwaukee, WI, pp. 442-444, October 2014.

[11] A. Dilan, "Quantum photovoltaic effect: Two photon process in solar cell," 2015 International Conference on Renewable Energy Research and Applications (ICRERA), Palermo, pp. 1084-1088, November 2015.