

# Combining SAO semantic analysis and morphology analysis to identify technology opportunities

Xuefeng  $Wang<sup>1</sup> \cdot$  Pingping  $Ma<sup>1</sup> \cdot$  Ying Huang<sup>1</sup> • Junfang Guo<sup>1</sup> · Donghua Zhu<sup>1</sup> · Alan L. Porter<sup>2,3</sup> · Zhinan Wang<sup>1</sup>

Received: 7 January 2016 / Published online: 4 February 2017 © Akadémiai Kiadó, Budapest, Hungary 2017

Abstract Increasingly complex competitive environments drive corporations in almost all industries to conduct omnibearing innovation activities to enhance their technological innovation capability and international competitiveness. Against this background, we propose subject–action–object (SAO) based morphological analysis to identify technology opportunities by detecting prioritized combinations within the morphology matrix. SAO structures emphasize the key concepts with provision of diverse technology information based on semantic relationships. The combination of SAO semantic structures can support the establishment of matrix, which consists of two dimensions: compositions and properties of technology. Later, novel indicators are used to evaluate the subsequent technological feasibility of each new configuration under a customized analysis and prior combinations aided by a high score can be identified. We apply this method to the case of dye-sensitized solar cells (DSSCs) in patents documents. The approach holds promise to strengthen information support systems for commercial enterprises in technical innovation and market innovation activities. We believe the analysis can be adapted well to fit other technologies, especially in their emerging stage.

Keywords Technology opportunities analysis - SAO semantic analysis - Morphological analysis - Technology mining - Dye-sensitized solar cells (DSSCs)

**JEL Classification O32 · C80** 

 $\boxtimes$  Xuefeng Wang wxf5122@gmail.com

<sup>&</sup>lt;sup>1</sup> School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China

<sup>&</sup>lt;sup>2</sup> School of Public Policy, Georgia Institute of Technology, Atlanta, GA 30332, USA

<sup>3</sup> Search Technology, Inc., Atlanta, GA 30092, USA

## Introduction

In the knowledge-oriented world economy, acquiring and maintaining competitive advantages, a key element for survival of commercial enterprises, is largely driven through information technology, which is a paramount venue for technological innovation. Mining of patent documents, scientific and technological journals publications, other resource analyses, etc., could help assess development paths that could synergize the support for such technical innovation. These technological innovation activities have potential to exploit the enterprises with a better competitive edge than ever before and it is vital to have a keen observation to spot the drift and development trends in the commercial sector (along with smart decisions for the exploitation of competitive advantage). Moreover, such innovational activities could be utilized to secure technology opportunity analysis (TOA), which could strengthen critical decision making to synergize the information support systems for commercial enterprises, non-profit scientific research institutions, and government departments of scientific and technological management.

Technology opportunity is a set of prospective technological advancements, which hint at some movement toward research and development aimed to achieve better performances (Klevorick et al. [1995](#page-20-0)). It brings innovations in the given field of knowledge and also reflects the cost of achieving some normalized unit of technical advancement (Yoon et al. [2014\)](#page-20-0). According to most studies, technology opportunities can be an exploration of a new vacuum area (Lee et al. [2009](#page-20-0); Geum et al. [2013](#page-19-0)) or indication of existing promising technology based gaps with inherent potential (Shibata et al. [2008;](#page-20-0) Yoon [2008\)](#page-20-0). Likewise, there are also attempts to use bibliometric indicators including publications counts (Albert [1991\)](#page-19-0), citations (Kim et al. [2014\)](#page-20-0) and patent claims (Lee et al. [2015\)](#page-20-0) from technological documents. These indicators are useful even though they mainly focus on basic information of patents or papers; there is also probably oversight on some heuristic and specific points that can only be identified through text. In order to overcome these limitations, some researchers have combined bibliometrics with other methods including text mining (Porter and Cunningham [2005](#page-20-0)), network analysis (Furukawa et al. [2015\)](#page-19-0), and semantic analysis (Sitarz and Kraslawski [2012\)](#page-20-0). These past data extrapolation mainly rely on judgment and expertise levels of humans to monitor the direction and focus of given technological strategy for development (Yoon et al. [2013\)](#page-20-0). In addition to their resemblance with exploration inside the black box, these techniques lack explicit detection orientation to some extent and instead lean more toward experts. Nevertheless, these tools have potential to provide sufficient background information to shadow the latest developing technologies such as the latest functions or new materials. Still, their projections lack pertinence and generative capabilities to be utilized in the applied areas market and in product designs.

To deal with this, we propose a technique that combines subject–action–object (SAO) semantic analysis and morphology analysis (MA) designed to identify technology opportunities. For a given system, Morphology Analysis is used to detect a full feasible range of combinations of selective variables. Among those combinations, utilized and unutilized combinations of technology can be estimated so that the analysis can give the direction of detecting novelties in aspects of promising technologies and vacuums. However, in conventional Morphology Analysis, construction of the kernel morphology matrix is largely dependent on the experts; the matrix describes compositions of technologies, but remains limited to its properties. As a remedy, our study prescribed morphology matrix<sup>1</sup> building that is based upon a combination with SAO semantic analysis. It offers inclusion of SAOs that provide synergy through two types of structures: ''Partitive SAO structures'', used for identifying the composition of technologies and ''Attribute SAO structures'' that attempt to identify properties of technologies. The combination with SAO is considered to be quantitative as well as pave the way for mining potential and detailed knowledge of technology. Moreover, from the professional or comprehensive point of view, the resultant functional matrix of technology analysis can be supposed to distinguish the probability of prospective practical alternatives to fulfill market demands. Resultant high priority combinations, appraised by specific indicators, are the nominated opportunities to fulfill the niche of a market over one or multiple problems. In accordance, the case of Dye-Sensitized Solar Cells (DSSCs) illustrates the detailed procedure and technique of the proposed method, combining SAO semantic analysis and Morphological Analysis.

The rest of the paper is organized as follows: second section, the theoretical background of research framework is presented; third section , a detailed description of the SAO-based MA methodology; fourth section, an illustration is proposed to approach the development of SAO-based MA (keeping DSSCs in view as a case study); and finally fifth section, the concluding remarks with a summary, discussion, and further research ideas.

## Theoretical background

#### Morphological analysis (MA)

Morphological Analysis is a problem-solving and quantification oriented scheme that recognizes a system's known parameters by reducing the solution space of integrated consequences of all feasible combinations of a variable that define a given system (Zwicky [1969;](#page-21-0) Sharif and Irani [2006](#page-20-0)). It is a systematic approach that combines distinctive advantages with prime informative content and an incredible visualization of solutions to follow techniques to model intricate problems rather than solutions (Pidd [2009](#page-20-0)). Hence, the morphological box allows a broad range of probable and prospective solutions to a problem.

Such techniques have captured attention recently for having practical usefulness in fields such as forecasting, scenario planning, engineering, product design, and strategy formulation. The MA technique has demonstrated utilization of a functional-technology matrix for color television set circuitry, supported by a systematic procedure and techniques using quantitative data in technology forecasting (O'Neal [1970\)](#page-20-0). Usually, the application of scenario analysis is combined with MA to exploit risk management in developing or assisting in choosing and deciding both quantitatively and qualitatively regarding selection and the magnitude of technological configurations to be utilized in scenario modeling (Rhyne [1995\)](#page-20-0). A pioneering engineering firm, the Norris Brothers in Sussex, England, in a historical milestone of an engineering design, applied MA in the development of the Bluebird hydroplane and cars (Norris [1963\)](#page-20-0). This highly contributed to product optimization, ensuring minimum design cost, and time constraints in Computer-Aided Design (Belaziz et al. [2000\)](#page-19-0). A systematic approach aided by MA and an

<sup>&</sup>lt;sup>1</sup> Our matrix consists of rows and columns. Later on, rows represent shapes or problems, while columns would represent components/dimensions/subsystems. Later, each column is further divided, called ''level of respective columns''.

application-oriented New Business Model Development (NBMD) helped business practitioners to develop, evaluate, and choose superior business models to meet their business objectives (Im and Cho [2013\)](#page-19-0).

Two noteworthy elements in MA are systematic analysis and the inventive selection of alternatives (Glenn and Gordon [2003\)](#page-19-0). However, MA's dependence on experts' intuition is a limitation, as there is no systematic approach to differentiate among dimensions and shapes—a core step in MA. Efforts are utilized to establish a morphological matrix by identifying the relationship between two noun phrases in SAO structures. Further, qualitative alternatives play crucial roles in the identification of technology opportunities in MA, which is the last step in the selection of the best and most attractive solutions (Zwicky [1969\)](#page-21-0). Conventional MA has some limitations, such as the hindrances to prioritizing alternatives, which can no longer be overcome (Yoon et al. [2014\)](#page-20-0). For a forward step towards a solution, it is necessary to evaluate every combination and choose the most suitable one.

#### Subject–action–object (SAO) semantic analysis

SAO semantic analysis is a fact-oriented modeling technique based upon the theory of Inventive Problem Solving (Russian acronym: TRIZ). S and O denote the components, and A denotes the effect or relationship between components in the invention. That is, S denotes the ''means'' and A–O denotes the ''end.'' For example, in the simple sentence, "The computer has a keyboard," "computer" is the subject, "keyboard" is the object, and ''has'' means that the keyboard is part of the computer (i.e., indicating the relationship between the subject and object).

SAO structures can be visualized to depict a variety of technology information. Finding similar SAO structures to establish a similarity matrix is a creative technique to solve problems, such as by forming an inventor competence map (Moehrle et al. [2005](#page-20-0)), exploring technological competition trends (Wang et al. [2015](#page-20-0); Yoon et al. [2013](#page-20-0)), and identifying patent infringement (Park and Yoon [2014\)](#page-20-0).

Analyzing the relationship between subject and object to explore solutions to certain technological problems is another trend in SAO methodology. Building the cause-andeffect function network utilizing the expression ''AO'' as a function to find inventive principles and discover innovative functions aided in generating a system leading to technology innovations (Kim et al. [2010](#page-20-0)). The Function-Oriented Database is an SAObased patent retrieval system to support Function-Oriented Search as a tool for capitalizing in searching patent databases to capture existing solutions to new problems (Choi et al. [2012a\)](#page-19-0). Accelerating open literature based discovery from An to Bn and from Bn to Cn with the help of identifying meaningful links between two SAO structures results in more accurate assessments for logic evidences (Vicente-Gomila [2014\)](#page-20-0).

To computationally extract SAO structures from patent textual descriptions, the use of natural language processing (NLP) is essential in part-of-speech tagging (Sekimizu et al. [1998\)](#page-20-0). SAO structure shows Partitive (Cascini et al. [2001](#page-19-0)) and Attribute relationships among products or technologies (Choi et al. [2012b](#page-19-0)). If the action word is a Partitive verb such as "have," "composed of," "include," or "be made of," the subject may include the object as a component. If the action word is an Attribute verb such as ''reduce,'' ''improve,'' ''increase,'' or ''protect,'' then the subject could affect the object as a problem. Therefore, we propose establishing a morphological matrix as paramount in Morphology Analysis aided by identifying specific SAO relationships.

## Methodology for SAO-based morphology analysis

In this paper, we offer a systematic approach to identify new technology opportunities. The framework is illustrated in Fig. 1. The process involves five distinctive steps: (1) selection of patent documents; (2) Identification of key technical problems and core technology components; (3) Extraction of SAO semantic structures; (4) Generation of a technological morphology matrix; and (5) Technology opportunity analysis.

## Selection of patent documents

The first step is the selection of patent documents. The main objective of this step is to select relevant patent documents with minimum irrelevant ones. The range of disciplines and continuous rapid development makes the selection process for patents even more ambiguous due to many hidden pitfalls and choices in the selection. One major concern in bio-research with respect to basic research is making the criteria too broad. To collect quality data, a patent retrieval query is used. Search strategy is obtaining, analyzing, classifying, and refining the related terms through mutual involvement of data mining scientists, field experts, and databases to better recall and precision (Porter et al. [2008;](#page-20-0) Huang et al. [2015\)](#page-19-0). The approach is split into two major steps: quantitative and qualitative based upon customized knowledge (Zhang et al. [2013](#page-20-0)). The quantitative method involves the extraction of high frequency words and their classification into subsystems through the clustering method using natural language processing techniques. Qualitative methods base term evaluation against predefined cluster criteria and the extension of term scopes utilizing field related experts and their experiences. Later, technology analysts again refine the terms while the recall and the precision potential of terms for a respective dataset remain the priority and finally downloading the required data. It is important to note here that there is tradeoff between the extension of accurate focus and the chances of errors in precision with respect to recall. For the final selection of terms to be used in the query, the retrieval of record measurements and their assessment play a key role in different methods



Fig. 1 Framework for identifying new technological opportunity

regarding matching with the target technology scope and the comparison of their applications.

## Identification of core technology components and key technical problems

In this phase, core technology components and key technical problems of a particular field through tech mining and comprehensive document surveys are required for the construction of a morphological matrix. That abstract and title are to extract the terms (single words or multiple-word phrases) using automated semantic analysis tools like NLP, which has gained significance due to terms diversity and varying incidence frequencies that offer serious challenges to human natural expertise over certain ranges. Moreover, it is valuable to evaluate every item with higher accuracy, so as to produce the desired results with less ambiguity. Therefore, we identify the core technology components and key technical problems based on term clumping and principal components analysis (PCA) using the professional desktop text mining software, VantagePoint (see References). Term clumping is a step used to manipulate these terms and topical phrases in the corpus by cleaning, combining, and grouping them. There are various statistical methods that can gain synergy when used in combination with term clumping, like PCA—a basic mode of turning terms into various relatively synonymous term groups to better represent the data and reduce the dimensions.

## Extraction of SAO semantic structures

An SAO semantic structure can be extracted from any description in the textual format including the title, abstract, claims, etc. In general, titles and abstracts are precise and have been regarded as the most meaningful parts; considering this, we extracted SAO semantic structures from abstracts. For this process, the software tool, GoldFire from Invention Machine (see References) is used. GoldFire is a problem–solution oriented technique and it relies on semantic search technology to extract scientific and technical information. Ensuring this, we exploited core components and key problems that have been identified in the last section to extract the SAOs structures. Isolation of two types of SAO structures include: "Partitive SAOs," the core components; and "Attribute SAOs," the key problems. On one hand, Partitive type indicates an inclusion linkage among its component for respective technology. For example, in a given technology, the action word ''contains'' means there is an inclusion relationship where the subject phrase includes an object phrase. Technology components are occasionally subject-phrase, while object-phrase are constant. Besides, in extraction of Partitive SAOs, the technology or core components of technology, are used as Subject and Partitive verbs are used as Action. On the other hand, Attribute type is a demonstration of AO (Action–Object), showing how the subject word reforms the technology attribute, referring to problems. For example, the action word ''increase'' means that the technology component can improve the linking technology problem. This gives rise to extract Attribute SAOs, where technical problems present as AOs. Thus, for GoldFire, two kinds of queries were prepared: (1) Partitive SAOs under the input, ''What technology/component does it contain?'' and (2) Attribute SAOs under the input, ''How do we solve the problem?''

#### Generation of the technological morphology matrix

The main objective of this step is to build the framework for the technological morphology matrix consisting of key technology components (dimensions/column) and key technical problems (shapes/rows). To achieve quantitative features, a morphological matrix was constructed, as shown in Table 1. We clustered valid subjects and objects from Partitive SAOs to obtain morphological information (composition) of the technology. In clustering, co-occurrence of subjects and objects in the same SAO structure depicts closeness. In each component, subjects and objects are clustered into different levels for respective components. Clustering of the terms in Partitive SAOs can be observed from the network picture although help from experts guided the way for identifying each cluster to obtain an accurate result. Accordingly, fuzzy matching to compare the subjects in Attribute SAOs with terms in the clustering is the next activity. The matching terms with respective levels (in the components) in clustering are marked with various problems to obtain the basic constitute of functional morphological matrix. In addition, in order to deal with impure data, cleaning is needed for these terms before clustering and matching. Hence, the effective data filtration to efficiently describe components and subcomponents of technology is preconditioned for superior results. For instance, common and basic words, which do not instruct a particular technology, such as "layer," "material," etc. are normally needed to clean; a similar act is needed as a merging of some terms with the same meaning in a field of technology, such as ''organic dye'' and ''organic pigment.''

#### Technology opportunity analysis

In the functional morphology matrix, each level can be multiplied with other levels for each component in a given problem to make a combination, which describes the structures of the given technology. For selection of prospective combinations among all combinations, the value of respective configurations can be decisive to find vacuums and highly feasible technology opportunities; it can also be calculated as sums of values of each level in respective combinations (Yoon and Park [2007\)](#page-20-0). Three indicators have been described to evaluate the value of each level (shown in Table [2](#page-7-0)). In these indicators, MI proposes the popularity of the level during the respective time period by using numbers of patents, while GTI shows the developmental future potential of the level. In mathematics, the slope or gradient of a line is a number that describes both the direction and the steepness of the line. Thus we use the slope of the last part of the curve to show the growth trend of the next period of the curve. To remove the deflection of these two quantitative indices, II, a qualitative analysis, can balance the value of a level under specific demands. Scores in 10-point-scale can be allocated by experts to each level of a component to evaluate their contribution and importance level against specific problems in the form of higher scores than others; then we can define means of all scores from experts as II.



Indicators	Evaluation index	Type	Explanation
MI GTI	Magnitude index Growth trend index	Quantitative Quantitative	Patent productivity for respective level Slope of last year in curve-fitting
П	Importance index	Qualitative	Mean scores from experts' grading method

<span id="page-7-0"></span>Table 2 Indicators for evaluating the values of levels

For final values of the level, we use the entropy weight method, which accounts for the quantity and quality of the data (Hung and Chen [2009\)](#page-19-0), to calculate the weight for each indicator.

In the decision matrix  $X =$  $x_{11} \quad \cdots \quad x_{1n}$  $: \mathbb{R}^+ \to \mathbb{R}^+$  $x_{m1} \quad \cdots \quad x_{mn}$  $\overline{1}$  $\frac{1}{2}$  $\overline{1}$ 7 5, the entropy weight of the j-th index is

$$
w_j^* = \frac{1 - e_j}{n - \sum_{j=1}^m (1 - e_j)}, \quad j = 1, 2, \dots, m
$$

where  $f_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}$ ,  $e_j = -\frac{1}{\ln m} \sum_{i=1}^{m}$  $\sum\limits_{j=1}f_{ij}\ln f_{ij}$ 

The  $e_j$  represents entropy of the *j*-th index and  $f_{ij}$  indicates probabilities of a random variable computed from a probability mass function. For  $w_j^*$ ,  $\sum_{j=1}^m w_j = 1$ ,  $w_j \in [0, 1]$ . The more diversity as well as the less entropy  $X_{ij}$  shows, the more influence and quantity of information the criterion has, and thus more weight it has. In this way, the ultimate value of each level equals the sum of the index value multiplied by the corresponding weight. The value of each combination can be calculated by the sum of the ultimate value of the respective levels.

In real life scenarios of marketplaces, consumer preferences in decision making about products and services are a function of the number of features per product with customization. There are always continuous effects to trace unique solutions that have the inherent capability to effectively solve a broader range of problems. Keeping this in mind, developers and decision makers prefer to generate technology opportunities with relatively broader target areas of solution. Building on this platform and keeping track of all the component technologies with the respective weight of problem solving, prior combinations were classified based on their high value in solving different kinds of problems. These prior combinations, including unused combinations and utilized combinations, are taken in our analysis to segregate prospective technology opportunities. On one hand, utilized configurations can be considered as promising technology, in which there is a technical popularity and extreme competition for short or long periods of time. On the other hand, unused combinations can be considered as technology vacuums, in which there could be a big breakthrough that could lead to a significant opportunity. Above all, technology opportunities can be identified and better applied in different scenarios from the viewpoint of multifarious requests.

#### Generation of the technological morphology matrix based on SAO structures

In this current study, patents were collected from the Derwent Innovation Index (DII), $\frac{2}{3}$  a database in the Web of Science based on a multi-step Boolean search algorithm. The selected field is DSSCs, a third-generation photovoltaic technology (Wang et al. [2014](#page-20-0)). The timespan from 1991 to 2014 was selected with consideration of the time lag behind the application date by an average period of 18 months. The retrieval query for DSSCs has been researched by Guo et al. [\(2012](#page-19-0)), and our study retrieves literature shown in Table [3](#page-9-0) that is referred to in their studies. Also, a total of 7003 patent documents including titles and abstracts were downloaded from DII.

A morphological matrix, consisting of dimensions as columns and shapes as rows, is the basis for determining the maximum possible combinations for technology components. Empirically, in 2014, the field of DSSC was classified into four subsystems: ''Photoan-ode," "Sensitizer," "Electrolyte," and "Counter-Electrode" (Zhang et al. [2014\)](#page-21-0). In consideration of cost-efficiency and accuracy to build the matrix, we directly utilized these four major components (subsystems) as dimensions of the matrix and later put them into the query for extracting Partitive SAOs to construct core levels of components.

For finding problems in the DSSC field, the data was taken from the ''ADVANTAGE'' part of the abstract, which mostly describes the purpose of the invention and which problems it can solve. We put all patents into VantagePoint and used term clumping to clean and group the data. Table [4](#page-9-0) shows the top 15 terms we got through the NLP section of VantagePoint. Accordingly, three problems—efficiency, low cost, and durability—are presented as core technical problems.

After defining core components and problems in DSSC, SAOs are extract for building a morphological matrix. In the current article, we used GoldFire software as the main tool to extract the SAO structure from the abstracts of the patent documents. By using components and problems as the search query in GoldFire, we got 12,002 Partitive SAOs and 4014 Attribute SAOs to gather 2717 "efficiency" data, 759 "low cost" data, and 538 "durability'' data. Search query and Partial SAO structures of DSSC technology patents utilizing GoldFire software is shown in Table [5](#page-10-0). Noticeably, GoldFire can automatically identify most synonyms and other forms of words. i.e., word ''contain'' in search strategy can find patents with synonyms like "comprise," "compose," and other forms of words like "containing," "contains," "contained," and etc. Hence, human expertise is only needed for relevant knowledge information, which is limited to the NLP, to control the input of the software. For example, in the main search strategy ''What can photoanode contain?'', NLP can find relevant information about composition of photoanode. However, other terms like ''optical electrode,'' ''first electrode,'' ''anode'' which have same meaning as ''photoanode'' in field of DSSCs will be ignored by machine. These professional terms only can be identified by field experts. Thus main search strategy should be accompanied with several relevant search strategy aided by experts to complete the searching.

 $2$  Derwent Innovations Index $^{\circledast}$  is capable of advance, accurate, and detailed patent databases, especially in the field of chemical, electrical, and mechanical engineering. It is enhanced by Derwent World Patents Index<sup>®</sup> and Derwent Patent Citation Index®. The abstract of patents includes USE, ADVANTAGE, NOVELTY, DESCRIPTION OF DRAWINGS, and DETAILED DESCRIPTION. [http://thomsonreuters.com/](http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/derwent-innovations-index.html) [en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/derwent-innovations-index.html.](http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/derwent-innovations-index.html)



<span id="page-9-0"></span>Table 3 Retrieval query for DSSCs technology

TS topic of patents

Final #4 OR #3 OR #2 OR #1

Rank	No. of patents	NLP/phrases	Rank	No. of patents	NLP/phrases
	2209	Dye-sensitized solar cell	6	390	Electrolyte
2	1134	Photoelectric conversion efficiency		372	Durability
3	620	Method	8	352	Efficiency
$\overline{4}$	558	Improved	9	320	Electrode
5	439	Cell	10	294	Low cost

Table 4 Top 10 terms of 7003 patents according to VantagePoint

From Partitive SAOs, morphological structure can be detected for each components through a network of subjects and objects that co-occurred in the same SAO structure. Clusters of the subjects and objects in each component is shown in Fig. [2](#page-11-0) and the data filtration and clustering were completed under the expert assistance. In the clustering, component ''Potoanode'' has four clusters: TiO2 film(P1), ZnO film(P2), SnO2 film(P3), and other metal oxide film(P4); component ''Sensitizer'' has two clusters: metal complex dye(S1) and organic dye(S2); component "Electrolyte" has three clusters: solid electrolyte $(E1)$ , liquid electrolyte $(E2)$  and gel electrolyte $(E3)$ ; component "Counter Electrode" has four components: metal counter electrode(C1), carbon counter electrode(C2), polymer counter electrode $(C3)$ , and inorganic compound counter electrode $(C4)$ . By fuzzy matching the subjects of Attribute SAOs with the terms in the clustering (matched more than 80%), we can detect the obvious relationship i.e., which cluster the subjects belong to and what kind of properties the components own.

After these initial preparations of the DSSC technology, the final morphological matrix was built (see Table [6\)](#page-13-0). The matrix columns reveal the composition of technology, while the rows are an indication of properties of technology. The relationship of the morphological structure to understand the technological configuration is interesting: patents can be described as combinations of levels in each component. For instance, the feature of photovoltaic devices or cells (U.S. Patent 2011005589) can be described by the combination: titanium oxide film (P1), organic dye (S2), liquid electrolyte (E2), and metal (platinum)



<span id="page-10-0"></span>Table 5 Search query and partial SAO structures of DSSC technology patents utilizing GoldFire software (relevant search information are used to substitute italicized words of main search query)

counter electrode (C1). From the matrix we can know that the cells have the ability to solve problems like improving efficiency and reducing costs (except for durability because the counter electrode is weak for maintaining durability of cells). The composition and attribution of patents in the field of technology can both be observed from this functional matrix. Moreover, for improving efficiency, reducing costs, and providing durability, the permutation knowledge can be used to compute the total combinations: 96 (i.e.,  $4 \times 2 \times 3 \times 4$ ); 24 (i.e.,  $3 \times 2 \times 2 \times 2$ ) and 18 (i.e.,  $3 \times 2 \times 3 \times 1$ ), respectively.

<span id="page-11-0"></span>

Fig. 2 Morphology of four components: a photoanode, b sensitizer/dye, c electrolyte, and d counter electrode

Obviously, the problem of efficiency has a wider range of solutions than other problems, and proportionally a greater amount of patent documents have been dedicated to improve the efficiency of DSSCs. As a result, it can be assumed that efficiency is the main problem to be solved in the development of DSSCs.





### Identification of technology opportunities

The analysis of technology opportunities depends on identifying prioritized and superior combinations. Thus, before analyzing all possible combinations, there is a need to evaluate the levels (the elements of combinations) with three indicators—magnitude index (MI), importance index (II), and growth trend index (GTI)— in order to judge the priority of

	Photoanode	Sensitizer	Electrolyte	Counter- electrode
Improve efficiency	TiO2 film(P1) ZnO film(P2) SnO2 film(P3) other metal oxide film(P4)	Metal complex dye(S1) organic $dye(S2)$	Solid electrolyte (E1) liquid electrolyte (E2) gel electrolyte (E3)	Metal counter electrode $(C1)$ carbon counter electrode $(C2)$ polymer counter electrode $(C3)$ inorganic compound counter electrode $(C4)$
Reduce cost	TiO2 film(P1) ZnO film(P2) other metal oxide film(P4)	Metal complex dye(S1) organic $dye(S2)$	Liquid electrolyte (E2) gel electrolyte (E3)	Metal counter electrode $(C1)$ carbon counter electrode $(C2)$
Provide durability	TiO2 film(P1) SnO2 film(P3) other metal oxide film(P4)	Metal complex dye(S1) organic $dye(S2)$	Solid electrolyte (E1) liquid electrolyte (E2) gel electrolyte (E3)	Carbon counter electrode $(C2)$

<span id="page-13-0"></span>Table 6 Morphology matrix of DSSCs technology

each combination over others. Table [7](#page-14-0) shows the tabulation of the normalized index of each level in each problem in DSSCs.

MI is a judgment factor to estimate the popularity of levels while GTI tries to depict the growth trend of levels. In GTI, data for each year is used to make cubic polynomial fitting for the graph. Besides, in the growth cycle, trajectories of technologies depict relatively varied patterns of maturity along the graph. Consequently, the developing trend could vary from the start to follow the classical s-curve life cycle with four stages: introduction, rapid growth, maturity, and decline (Park et al. [2013\)](#page-20-0). In cubic polynomial fitting, GTI can be estimated through measuring the slope of the last part of the curve in a final dependent variable and a determination coefficient (R2) of curves that are all above 0.5 and have a highest value achieved at 0.94512. II, which is estimated by experts, can act as a buffer to minimize the expected deviation raised by the previous two quantitative indicators. In this intention, experts are invited to gauge the importance through allocating scores (10-pointscale) to each level of the component, and then the levels with more contribution in solving specific problems can get higher scores than others. The mean of all scores from experts is defined as II. In the last, the Entropy Weight Method can calculate weights of the three indicators for respective problems as sums of the indicators multiplied by the weight. It is noteworthy that all the indices here added 2 because the index should be greater than 0 in the Entropy Weight Method.

For specific problem solving, the value of combinations can be directly calculated as the sum of values of respective levels in each component. In the current case, our intention is to identify multi-function combinations with the inherent capability to solve diverse problems at a commercial level. The provision of weight to each problem, as well as the value of level in each problem, can calculate the comprehensive score of levels, and thus, the synthetic value of a configuration is the sum of the comprehensive value of levels (CVL) selected in each component. To find the novelty in accordance with increasing



<span id="page-14-0"></span>Scientometrics (2017) 111:3–24 17

Table 7 Zero-mean normalization of index for each level in certain problem

Indexes  $=$  MI, GTI, II

efficiency and performance in estimating the appropriate solution for other problems, we consulted with experts to allocate weight to different problems: weight  $1 = 0.6$  for improving efficiency, weight  $2 = 0.2$  for reducing cost, and weight  $3 = 0.2$  for providing durability. Notably, the weight can be changed along with the requirements of the marketing environment. The value of levels in both professional analysis (specific problems) and comprehensive analysis (different problems) are shown in Table [8](#page-16-0).

The four dimensions in the field of DSSCs—"Photoanode," "Sensitizer," "Electrolyte,'' and ''Counter electrode'' (Zhang et al. [2014](#page-21-0)) equally affect the selection of unique combinations. As in morphological analysis, the combination of target objects can reveal interesting facts (Zwicky [1969](#page-21-0)) hence, we selected levels in each component as renewed combinations; the value of the combinations is the sum of the value (CVL) of these levels. These combinations could provide the full range of customization of product partitioning involve in product design (Belaziz et al. [2000](#page-19-0)) although in our case, these combinations are backed by the synergy of SAO. Table [9](#page-17-0) shows the top 20 combinations of value for the three problems, the 17 combinations utilized, and the three combinations unused. These combinations are considered as prior combinations after the analysis of proposed indicators.

Among prior utilized combinations, it was found that for the period of 1991 to 2014, patents with 17 combinations occupy 68% while the patents, which cite these 17 combinations, occupy 55%. Thus, it can be deduced that these 17 combinations occupy greater influence than other combinations to demonstrate heat in the form of research activity on these configurations. To verify this trend for the prior combinations, a total of 576 documents were downloaded from DII for the year 2015 using the same search query. It was found that 7 of 17 combinations were available in 2015 and captured 89% of all combinations. Specially, the combination TiO2 film-organic dye-liquid electrolyte-metal counter electrode (Rank 3) has excellent performance in both occupation (23% of patents) and influence (11% of citations), and it also performs well (29% of patents) in 2015. Conclusively, prior utilized combinations can be regarded as promising research; however, for short and middle term, these are still considered as technology opportunities. Especially in the view of enterprises that have intentions to utilize their (first time or diversification) resources to determine direction or alternative for choosing business model (Im and Cho [2013\)](#page-19-0), selecting existing promising technology (in our case is DSSCs) is always a good choice, as it not only reduces the cost in the exploring period of a product or service design, but it also attains high effectiveness in the product itself. Besides consideration of these combinations as opportunities, they also impose some challenges due to intense competitions and reallocation of resources (e.g., seven combinations survived from 17 combinations in 2015), although which can also be buffered (Yoon et al. [2013\)](#page-20-0). Thus, for a big company that has already invested huge capital, the sunk cost efficiency can be maximized by obtaining advantages through exploring these promising technologies when they start to come into DSSCs. The distribution of these 17 combinations as Numbers of Patents (1991–2014), Numbers of Citations (1991–2014), and Numbers of Patents (2015) are shown in Fig. [3.](#page-17-0)

Numbers of Patents (1991–2014) Numbers of Citations (1991–2014) Numbers of Patents (2015).

For the prior 3 unused combinations, interestingly the actual technical vacuum exists. To assure that whether these vacuums are opportunities or not, surveys and consultations with experts are necessary. Accordingly, it was found that the TiO2 film, as the most widely used form of photoanode, has the best performance compared to other materials including ZnO, Nb2O5, SnO2, etc. Coming towards dye sensitizers, metal complex dye

<span id="page-16-0"></span>

VL<sub>i</sub> value of level in the i-th problem, CVL comprehensive value of levels  $VI_i$  value of level in the i-th problem,  $CVI$  comprehensive value of levels

Rank	Combination	Value	Occupied?
1	TiO2 film-organic dye-liquid electrolyte-carbon counter electrode	10.871	Yes
2	TiO2 film-metal complex dye-liquid electrolyte-carbon counter electrode	10.625	Yes
3	TiO2 film-organic dye-liquid electrolyte-metal counter electrode	10.062	Yes
4	TiO2 film-organic dye-gel electrolyte-carbon counter electrode	9.989	Yes
5	TiO2 film-metal complex dye-liquid electrolyte-metal counter electrode	9.816	Yes
6	TiO2 film-metal complex dye-gel electrolyte-carbon counter electrode	9.743	Yes
7	TiO2 film-organic dye-liquid electrolyte-inorganic compound counter electrode	9.678	Yes
8	TiO2 film-metal complex dye-liquid electrolyte-inorganic compound counter electrode	9.432	Yes
9	TiO2 film-organic dye-solid electrolyte-carbon counter electrode	9.318	Yes
10	TiO2 film-organic dye-gel electrolyte-metal counter electrode	9.18	Yes
11	TiO2 film-metal complex dye-solid electrolyte-carbon counter electrode	9.072	No
12	TiO2 film-organic dye-liquid electrolyte-polymer counter electrode	9.027	Yes
13	TiO2 film-metal complex dye-gel electrolyte-metal counter electrode	8.934	Yes
14	TiO2 film-organic dye-gel electrolyte-inorganic compound counter electrode	8.796	No
15	TiO2 film-metal complex dye-liquid electrolyte-polymer counter electrode	8.781	Yes
16	SnO2 film-organic dye-liquid electrolyte-carbon counter electrode	8.671	Yes
17	other metal oxide film-organic dye-liquid electrolyte-carbon counter electrode	8.664	Yes
18	ZnO film-organic dye-liquid electrolyte-carbon counter electrode	8.623	Yes
19	TiO2 film-metal complex dye-gel electrolyte-inorganic compound counter electrode	8.55	No
20	TiO2 film-organic dye-solid electrolyte-metal counter electrode	8.509	Yes

<span id="page-17-0"></span>Table 9 Top 20 technology combinations in the DSSC technology and each level is separated by "-"



Fig. 3 Distribution of 17 utilized combinations. a Numbers of patents (1991–2014), **b** numbers of citations (1991–2014) and c numbers of patents (2015)

and organic dye have competitive development in the field of DSSCs. For electrolytes, one thing we need to consider is that in the past, conventional liquid electrolytes have been the main focus for the high-energy conversion efficiency in DSSCs. However, liquid spilling is always the biggest problem for liquid electrolytes. In the wake of the technical breakthrough of gel electrolytes and solid electrolytes, the gel and solid forms of electrolytes are

going to be in the dominant position in sensitizers. Coming towards the counter-electrodes, as previous platinum electrodes are expensive and have strict conditions for manufacturing, other new electrodes, especially carbon counter electrode, are increasingly developed to replace platinum to achieve mass production. Thus, the combination TiO2 film –metal complex dye –solid electrolyte –carbon counter electrode (Rank 11) could be linked with higher probabilities of technology opportunities. Other unoccupied configurations also have possibilities to become opportunities, although they need more refined research activities with comprehensive industrial survey (Klevorick et al. [1995\)](#page-20-0). For the long term, these unoccupied combinations were thought to be technology opportunities and could bring the desired benefits (Belaziz et al. [2000\)](#page-19-0) once there is a big technical breakthrough in them. Thus, in the DSSC field, professional enterprises that work for years and want to expand into existing and new markets shared through exploring radical novelty, vacuum technology opportunity is the attractive choice.

### Concluding remarks and future work

In this study, we proposed an SAO-based systematic Morphology Analysis (MA)—relatively a systematic, customized, quantitative, and results-oriented approach that is capable of unearthing the technology morphology to extract novel technology opportunities. On one side, semantic SAO structures clearly depict structural relationships against the technological components and fully reflect specific key-findings in the patent dataset. Particularly, an added emphasis is given to the relation between Subjects and Objects in SAOs. We present Partitive and Attribute relationships of SAO semantic structures despite conventional SAO research of using all SAO combination possibilities with much redundancy (Yoon and Kim [2011](#page-20-0)). Furthermore, these kinds of SAOs have potential to build a morphological relationship matrix that manifests the composition of technologies from Partitive SAOs and the property of technologies from Attribute SAOs. Consequently, prescribed functional morphology matrix is conceivable to be rated over conventional methods of MA, which leans toward the reliance on qualitative and expert-dependence. On the other side, each dimension (and their respective levels) against problems/shapes examine the prospective, possible, and practical alternatives that a system could utilize to frame competitive advantage and respective adoption, instead of checking out all configurations (Yoon and Park  $2007$ ). At the same time, for technology opportunities, two types of opportunities (existing promising technology and technical vacuum) have been identified under different requirements of product design that can significantly alter the orthodox complexion of a singular project development (Alvarez and Ritchey [2015](#page-19-0)). Hence, suggested contribution present a more flexible and diversified analysis in identifying the technology opportunities and aids decision makers in formulating R&D strategies in the process of technical innovation and market innovation activities.

However, there are some limitations that need to be considered for future developments. The first constraint is the technique for extracting the SAO structure utilizing abstracts. Although the abstracts, as prime content that is rated and utilized by the research community to conduct SAO analysis, have still exerted restrictions on some technological information. The second constraint is regarding the full automation of the method, including establishing the morphological matrix; utilizing the processing software; analyzing the extensive amount of data; and supervizing instructions of domain experts on the modeling matrix. Extensive provision of information to experts with focused objectivity

<span id="page-19-0"></span>may decrease the limitation of expert knowledge to upgrade the efficacy. The third constraint is the judgment through indicators on the basis of which score is given to each level against the respective problem. This requires more vigilance, background practical knowledge, and expertise in the specified area; otherwise, complications may arise regarding evaluation and precision.

In the future, the exploitation of NLP techniques to extract SAO structures could encompass nearly all the information of a single patent. Achieving automation in experiments for establishing a morphological matrix is another meaningful research direction. Our experimental technique could be applied at a commercial level and have a competitive capability to attain more information on technology applications with improvements in technology opportunity identification. Moreover, detecting novelty by using prior combinations as the indicators can also be a good direction. A future effort could be made which would integrate this text based method with the IPC or others.

Acknowledgements We acknowledge support from the General Program of the National Natural Science Foundation of China (Grant No. 71373019) and the National High Technology Research and Development Program of China (Grant No. 2014AA015105). The findings and observations contained in this paper are those of the authors and do not necessarily reflect the views of the supporters. The authors would like to thank Omer Hanif, Chao Yang, Pengjun Qiu, Yun Fu, and other colleagues in the ''Co-lab of technology innovation'' of Beijing Institute of Technology, Georgia Institute of Technology, and Manchester University, for their advice and feedback.

## **References**

- Albert, M. B. (1991). Direct validation of citation counts as indicators of industrially important patents. Research Policy, 20, 251–259.
- Álvarez, A., & Ritchey, T. (2015). Applications of general morphological analysis: From engineering design to policy analysis. Acta Morphologica Generalis, 4(1), 1–40.
- Belaziz, M., Bouras, A., & Brun, J.-M. (2000). Morphological analysis for product design. Computer-Aided Design, 32(5–6), 377–388.
- Cascini, G., Lucchesi, D., & Rissone, P. (2001). Automatic patents functional analysis through semantic processing. In 12th ADM international conference. Rimini, Italy.
- Choi, S., Kang, D., Lim, J., & Kim, K. (2012a). A fact-oriented ontological approach to SAO-based function modeling of patents for implementing Function-based Technology Database. Expert Systems with Applications, 39(10), 9129–9140.
- Choi, S., Park, H., Kang, D., Lee, J. Y., & Kim, K. (2012b). An SAO-based text mining approach to building a technology tree for technology planning. Expert Systems with Applications, 39(13), 11443–11455.
- Furukawa, T., Mori, K., Arino, K., Hayashi, K., & Shirakawa, N. (2015). Identifying the evolutionary process of emerging technologies: A chronological network analysis of World Wide Web conference sessions. Technological Forecasting and Social Change, 91, 280–294.
- Geum, Y., Jeon, J., & Seol, H. (2013). Identifying technological opportunities using the novelty detection technique: A case of laser technology in semiconductor manufacturing. Technology Analysis & Strategic Management, 25(1), 1–22.
- Glenn, J. C., & Gordon, T. J. (2003). Futures research methodology version. Washington, DC: American Council for the UNU.
- Guo, Y., Ma, T. T., Porter, A. L., & Huang, L. (2012). Text mining of information resources to inform forecasting innovation pathways. Technology Analysis & Strategic Management, 24(8), 843–861.
- Huang, Y., Schuehle, J., Porter, A. L., & Youtie, J. (2015). A systematic method to create search strategies for emerging technologies based on the Web of Science: Illustrated for 'Big Data'. Scientometrics, 105(3), 2005–2022.
- Hung, C. C., & Chen, L. H. (2009). A fuzzy TOPSIS decision making model with entropy weight under intuitionistic fuzzy environment. In Proceedings of the international multiconference of engineers and computer scientists 2009 (Vol. 1, pp. 13–16). Hong Kong, March 18–20.
- Im, K., & Cho, H. (2013). A systematic approach for developing a new business model using morphological analysis and integrated fuzzy approach. Expert Systems with Applications, 40(11), 4463–4477.
- <span id="page-20-0"></span>Kim, H., Choi, S., Jeong, C., & Kim, K. (2010). Cause-and-effect function analysis. In IEEE international conference on management of innovation & technology (pp. 518–523). doi: [10.1109/ICMIT.2010.](http://dx.doi.org/10.1109/ICMIT.2010.5492765) [5492765](http://dx.doi.org/10.1109/ICMIT.2010.5492765)
- Kim, B., Gazzola, G., Lee, J.-M., Kim, D., Kim, K., & Jeong, M. K. (2014). Inter-cluster connectivity analysis for technology opportunity discovery. Scientometrics, 98(3), 1811–1825.
- Klevorick, A. K., Levin, R. C., Nelson, R. R., & Winter, S. G. (1995). On the sources and significance of interindustry differences in technological opportunities. Research Policy, 24(2), 185–205.
- Lee, C., Kang, B., & Shin, J. (2015). Novelty-focused patent mapping for technology opportunity analysis. Technological forecasting and social change, part B, 90, 355–365.
- Lee, S., Yoon, B., & Park, Y. (2009). An approach to discovering new technology opportunities: Keywordbased patent map approach. Technovation, 29(6–7), 481–497.
- Moehrle, M. G., Walter, L., Geritz, A., & Müller, S. (2005). Patent-based inventor profiles as a basis for human resource decisions in research and development. R&D Management, 35(5), 513–524.
- Norris, K. W. (1963). The morphological approach to engineering design. In Conference on design method. New York: Macmillan.
- O'Neal, C. (1970). New approaches to technological forecasting—morphological analysis: An integrative approach. Business Horizons, 13(6), 47–58.
- Park, H., Ree, J. J., & Kim, K. (2013). Identification of promising patents for technology transfers using TRIZ evolution trends. Expert Systems with Applications, 40(2), 736–743.
- Park, I., & Yoon, B. (2014). A semantic analysis approach for identifying patent infringement based on a product–patent map. Technology Analysis & Strategic Management, 26(8), 855–874.
- Pidd, M. (2009). Tools for thinking: Modelling in management science. London: Wiley.
- Porter, A. L., & Cunningham, S. W. (2005). Tech mining: exploiting new technologies for competitive advantage (Wiley series in systems engineering). New York: Wiley.
- Porter, A. L., Youtie, J., Shapira, P., & Schoeneck, D. J. (2008). Refining search terms for nanotechnology. Journal of Nanoparticle Research, 10(5), 715–728.
- Rhyne, R. (1995). Field anomaly relaxation. Futures, 27(6), 657–674.
- Sekimizu, T., Park, H. S., & Tsujii, J. (1998). Identifying the interaction between genes and gene products based on frequently seen verbs in medline abstracts. Genome Inform Ser Workshop Genome Inform, 9, 62–71.
- Sharif, A. M., & Irani, Z. (2006). Applying a fuzzy-morphological approach to complexity within management decision making. Management Decision, 44(7), 930–961.
- Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. Technovation, 28(11), 758–775.
- Sitarz, R., & Kraslawski, A. (2012). Application of semantic and lexical analysis to technology forecasting by trend analysis—thematic clusters in separation processes. In B. Ian David Lockhart & F. Michael (Eds.), Computer aided chemical engineering (Vol. 30, pp. 437–441). Amsterdam: Elsevier.
- Vicente-Gomila, J. M. (2014). The contribution of syntactic–semantic approach to the search for complementary literatures for scientific or technical discovery. Scientometrics, 100(3), 659–673.
- Wang, X., Li, R., Ren, S., Zhu, D., Huang, M., & Qiu, P. (2014). Collaboration network and pattern analysis: Case study of dye-sensitized solar cells. Scientometrics, 98(3), 1745–1762.
- Wang, X., Qiu, P., Zhu, D., Mitkova, L., Lei, M., & Porter, A. L. (2015). Identification of technology development trends based on subject–action–object analysis: The case of dye-sensitized solar cells. Technological Forecasting and Social Change, 98, 24–46.
- Yoon, B. (2008). On the development of a technology intelligence tool for identifying technology opportunity. Expert Systems with Applications, 35(1–2), 124–135.
- Yoon, J., & Kim, K. (2011). Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. Scientometrics, 88(1), 213–228.
- Yoon, B., & Park, Y. (2007). Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information. IEEE Transactions on Engineering Management, 54(3), 588–599.
- Yoon, B., Park, I., & Coh, B.-Y. (2014). Exploring technological opportunities by linking technology and products: Application of morphology analysis and text mining. Technological Forecasting and Social Change, 86, 287–303.
- Yoon, J., Park, H., & Kim, K. (2013). Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. Scientometrics, 94(1), 313–331.
- Zhang, Y., Guo, Y., Wang, X., Zhu, D., & Porter, A. L. (2013). A hybrid visualisation model for technology roadmapping: Bibliometrics, qualitative methodology and empirical study. Technology Analysis & Strategic Management, 25(6), 707–724.

<span id="page-21-0"></span>Zhang, Y., Zhou, X., Porter, A. L., & Vicente Gomila, J. M. (2014). How to combine term clumping and technology roadmapping for newly emerging science & technology competitive intelligence: ''Problem & solution'' pattern based semantic TRIZ tool and case study. Scientometrics, 101(2), 1375–1389.

Zwicky, F. (1969). Discovery, invention, research: Through the morphological approach. New York: Macmillan-Science.