A Machine Learning Approach for Solar Power Technology Review and Patent Evolution Analysis

Amy J.C. Trappey [1], Paul P.J. Chen [1], Charles V. Trappey [2],* and Lin Ma [3]

[1] Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu 300, Taiwan; trappey@ie.nthu.edu.tw (A.J.C.T.); s106034539@m106.nthu.edu.tw (P.P.J.C.)
[2] Department of Management Science, National Chiao Tung University, Hsinchu 300, Taiwan
[3] Science and Engineering Faculty, Queensland University of Technology, Brisbane, QLD 4000, Australia; l.ma@qut.edu.au
* Correspondence: trappey@faculty.nctu.edu.tw; Tel.: +886-35727686

check for updates

Abstract: Solar power systems and their related technologies have developed into a globally utilized green energy source. Given the relatively high installation costs, low conversion rates and battery capacity issues, solar energy is still not a widely applied energy source when compared to traditional energy sources. Despite the challenges, there are many innovative studies of new materials and new methods for improving solar energy transformation efficiency to improve the competitiveness of solar energy in the marketplace. This research searches for promising solar power technologies by text mining 2280 global patents and 5610 literature papers of the past decade (January 2008 to June 2018). First, a solar power knowledge ontology schema (or a key term relationship map) is constructed from the comprehensive literature and patent review. Non-supervised machine learning techniques for clustering patents and literature combined with the Latent Dirichlet Allocation (LDA) topic modeling algorithm identify sub-technology clusters and their main topics. A word-embedding algorithm is applied to identify the patent documents of the specified technologies. Cross-validation of the results is used to model the technology progress with a patent evolution map. Initial analysis show that many patents focus on solar hydropower storage systems, transferring light generated power to waterpower gravity systems. Batteries are also used but have several limitations. The objectives of this research are to review solar technology development progress and describe the innovation path that has evolved for the solar power domain. By adopting unsupervised learning approaches for literature and patent mining, this research develops a novel technology e-discovery methodology and presents the detailed reviews and analyses of the solar power technology using the proposed e-discovery workflow. The insights of global solar technology development, based on both comprehensive literature and patent reviews and cross-analyses, helps energy companies select advanced technologies related to their key technical R&D strengths and business interests. The structured solar-related technology mining can be extended to the analysis of other forms of renewable energy development.

Keywords: solar power; energy generation; patent portfolio; clustering; LDA; word2vec; technology mining; text mining

1. Introduction

Climate change and quickly depleting nonrenewable energy sources are a driving force behind sustainable energy research and development that is impacting all countries and enterprises. The development of green energy, or renewable energy, is now a critically active and growing research topic. Of the many kinds of renewable energy, solar power is the most common and well-known source of energy which can be obtained easily, and has fewer limitations to purchase and install. Solar energy

generation is still comparatively expensive when compared to fossil fuels, and the methods for storing the energy is often insufficient for power supply through the night, prolonged storms and overcast cloudy weather. The purpose of this paper is to define the technology development of solar power and forecast the solutions which have the greatest chance for market adaptation as well as providing a source of energy during the day which may also be stored and used at night.

Sunlight is a major source of inexhaustible free energy on the earth. Several renewable energy sources (i.e., hydraulic, biomass, geothermal, and solar) can be utilized to yield sufficient energy for power generation. Of these, solar energy has significant global potential since geothermal and hydraulic (e.g., damns) are limited by geographic locations and biomass (e.g., wood and agricultural products, solid waste, landfill gas, biogas, ethanol, biodiesel) requires combustion that actually worsens the severity of carbon emissions [1]. Technologies are being developed to generate electricity from harvested solar energy. Several solar energy systems are economically viable and have been applied throughout the world as renewable alternatives (but not completely replacing conventional energy sources) [2]. Countries, such as the United States, Germany and China, have significant technological R&D and manufacturing capabilities that can be used to promote domestic low-carbon policies and develop an internationally competitive green industry [3]. Solar research is associated with the current drive toward reducing global carbon emissions, a major global environmental, social, and economic issue that energy manufacturing competes against the ubiquitous use of fossil fuels. Studies show that the latest development of solar materials has created a new research frontier to combine solar cells with Internet of Things (IoT) devices to build smart grids with night time capabilities [4].

This solar technological review research is thoroughly and uniquely conducted and cross-referenced based on both collections of academic literature and global patents. The systematic investigative process flow is illustrated in Figure 1. Patent documents are retrieved from Derwent Innovation (DI), which includes online patent datasets from more than 90 national and regional patent corpuses and is widely used for global patent-based analyses and case studies [5]. Academic papers are retrieved from the Web of Science (WoS), an online scientific citation indexing service providing access to papers in more than 20,000 journal databases as crucial references to cross-disciplinary research. Both collected literature and patent documents are reviewed and analyzed using text mining and machine learning techniques for natural language processing and knowledge extractions. Then, the domain knowledge ontology is constructed, consisting of the main categories, sub-categories and their relationships. In any given domain, the ontology model can be iteratively and periodically retrained and modified while more relevant literature and patents are updated from both WoS and DI corpuses. Afterward, based on the ontology schema, the research further discovers the major technological evolution trends in major categories, using a modified formal concept analysis (MFCA) approach. Worth noting about the proposed technology mining methodology, the patent evolutions in major clusters identified using the non-supervised clustering and LDA algorithms, are further cross-referenced to literature to strengthen the validity of the discovered R&D development trends. The detailed solar power background and its technology mining steps are depicted in the following sections. The case study of patent evolutions for three sub-technical clusters is also described in the section before the conclusion section. The purpose of this research is to review solar technology development and describe the current development path for the domain. By constructing a machine learning program system, the readers can better understand the detailed technologies under each category. The proposed methodology framework can be used to further explore additional technical aspects of solar technology.
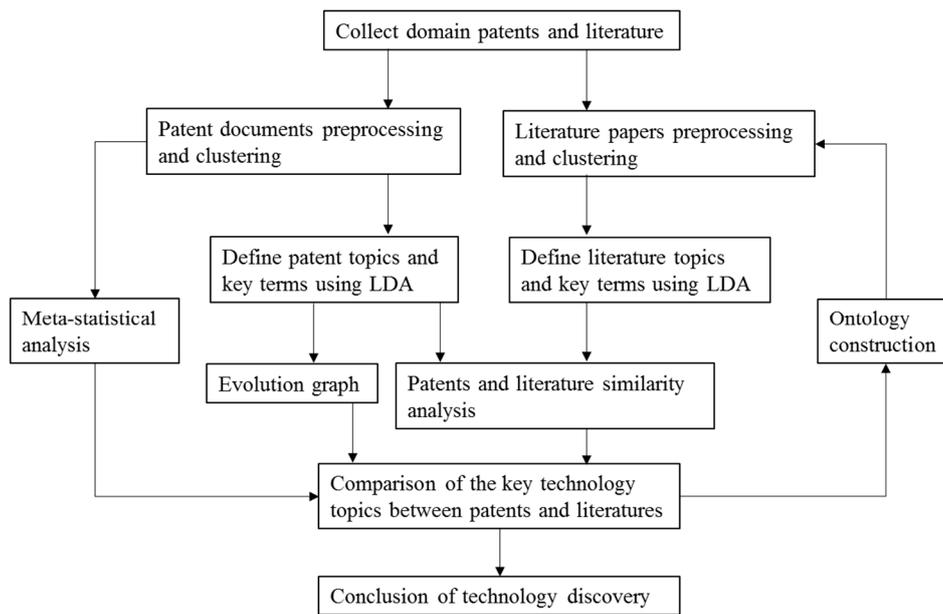
Figure 1. The research flowchart of the technology mining and analysis.

Section 2 presents a literature review of solar power technology and similar research using text mining to explore the current development path. Section 3 introduces the methodology framework and the approaches used in this research. Section 4 demonstrates the patent analysis and program results based on the approaches in Section 3. Section 5 summarizes and describes the technology path discovered and organized in this research, and illustrates the contributions as well as future work for both researchers and companies.

2. Literature Review

This section provides a brief overview of the technology domain, namely solar power and solar power energy storage. Literature relevant to the technology are reviewed to create basic ontology graphs of the domain to construct the search strings used to query online patent and literature databases. An iterative approach is used to search the references and prior arts to improve the search for additional patents and research papers.

2.1. Solar Power Cells and Energy Storage

As technology matures and the product life cycle enters the growth stage, there is a fast, increasing demand for equipment and services. Renewable energy sources are a viable but expensive alternative with ongoing concerns about the efficiency, cost and implementation across widespread electrical grid infrastructures. Most renewable energy technologies are in the late introductory stage of the product life cycle and yet demand has not seen fast, growing demand. This type of market response is often called the Gompertz effect since significant capital investments have been made in non-renewable energy facilities that are not fully depreciated and can function for many more decades [6]. Renewable energy such as wind and solar power cannot produce power reliably with current technology since power production rates change with seasons, months, days, or even within a day. The marketplace requires large scale and affordable solutions to alleviate fluctuating output and provide methods to store excess production for later consumption [7]. Solar energy is one of the most common and popular sources of clean energy, and the requirement to have access to sunlight is a very simple requirement compared to other solutions. Direct solar radiation may have the greatest potential for large-scale utilization once viable energy storage technology is developed. Kabir [1] reviewed and discussed both the merits and limitations of solar energy technologies. A number of technical problems affecting renewable energy

research are also highlighted, along with beneficial interactions between regulation policy frameworks and future prospects.

Concentrating solar power (CSP) plants generate solar thermal electricity without greenhouse gas emissions and is a key energy technology with a negative impact on climate change. A thermoelectric solar plant uses a set of units arranged in the following order [8]. The first unit in the sequence is a mirror designed to collect solar radiation and concentrate it at a focal point. The second unit, linked to the solar concentrator, is the receiver and the heat exchanger which circulates heat transfer fluid (such as molten salt or synthetic oil) to absorb the concentrated heat. The final unit consists of a second heat exchanger that transfers the accumulated thermal energy to another fluid (usually steam) which drives a turbine electric generator.

To reduce the cost per area required by photovoltaic (PV) cells, solar concentrators rely on a set of mirrors or moving mechanical structures to direct the light to the concentrator as the sun moves. Solar concentrators have disadvantages since they need to track the sun's position and may be affected by overheating from the concentration of light and heat on the solar cells [9]. The advantages of using volume holographic optical elements [10] are appealing for lightweight and cheap solar concentrator applications and are expected to become an important advancement when integrated into solar panels. Ferrara et al. [9] presented a review of holographic-based solar concentrators using different materials. The physical principles and main advantages and disadvantages, such as cool light concentration, selective wavelength concentration and the possibility to implement passive solar tracking are discussed. Different configurations and application strategies are also discussed in this study.

Unlike solar PV technologies, CSP plants use steam turbines that match conventional electrical generating services. CSP plants can be equipped with fossil fuel systems to deliver additional energy or to produce electricity during the night or when clouds block the sun [11]. There are four types of CSP reflection mirrors: solar power towers, Fresnel reflectors, Sterling dishes and parabolic troughs. CSP can use molten salt to store heat, enabling the generation of electricity for several hours even without sunshine. During off-peak hours, the CSP's power generation can be adjusted according to electricity demand. The power generation can be shut down quickly and the accumulated heat can be stored by the molten salt [12]. Today's most advanced CSP systems are towers integrated with two-tank, molten-salt thermal energy storage, delivering thermal energy at 565 °C for integration with conventional steam as Rankine power cycles. The power towers trace their lineage to the 10-MWe pilot demonstration of Solar Two in the 1990s. The design lowered the cost of CSP electricity by approximately 50% over the prior generation of parabolic trough systems. However, the decrease in cost of CSP technologies has not kept pace with the falling cost of PV systems [13]. Ma et al. [14] examined and compared two energy storage technologies, i.e., batteries and pumped hydro storage (PHS), for the renewable energy powered micro-grid power supply system on a remote island. It was found that the employment of conventional battery had higher life-cycle costs (LCC) than the advanced deep cycle battery, indicating that using deep cycle batteries is more suitable for a standalone renewable power supply system. The pumped storage combined with battery bank had almost half LCC as a conventional battery, making this combined option more cost-competitive than the sole battery option.

Solar photovoltaic (PV) technologies may also be used to convert solar energy into long term storable forms by using electricity to cause chemical reactions, such as the conversion of water to hydrogen and oxygen. Solar PV systems produce no greenhouse gas emissions during operation, do not produce other pollutants such as oxides of sulfur and nitrogen, and limit the use of water for cooling [15]. Knowledge of solar radiation is important for the integration of energy systems using solar panels on buildings, greenhouses, or with grid networks. For the optimal management of energy, the development of forecasting tools is needed to anticipate the rates of energy consumption. Since global horizontal irradiation data are rarely measured, Notton et al. [16] built an artificial neural network model to estimate the values. As solar collectors are often tilted to face the sun, a second ANN model was further developed to transform horizontal irradiation data into global tilted irradiation data.

The most widely adapted solar cell is constructed with silicon wafers and accounts for about 90% of the total global output [17]. Due to the shortage of raw materials, the traditional silicon wafer solar cells are not meeting the demand and cost requirements of the fast-growing global market. Thin film conductors have become the technology focus of new generation solar cells since they do not require much silicon. There are many types of thin film solar cells, including germanium films (amorphous germanium a-Si, microcrystalline germanium c-Si, stacked a-Si/c-Si), compound semiconductors (copper indium gallium selenide CIS/CIGS, cadmium telluride CdTe) and dye sensitization solar cells (DSSC) [17]. Although thin film solar cells have low energy conversion efficiency, low mass production yield, and high costs, there are many advantages such as material savings since they can be fabricated using inexpensive glass or plastic substrates, can be customized and offer greater flexibility for structural applications.

The tandem cell is a PV cell which uses two solar cells with different absorption characteristics enabling a wider range of the solar spectrum to be converted to energy. A transparent titanium oxide (TiOx) layer separates and connects the two cells. The TiOx layer serves as the electron transporting and collection layer for the first cell, and is the foundation that enables the fabrication of the second cell to complete the tandem cell architecture [18]. The technical difficulty of the tandem battery is that the current generated must match and the currents generated by the two layers of the battery are not easy to synchronize. High concentration PV technology has received international attention due to advantages of efficient high-power generation, a low temperature coefficient, and the potential to reduce power generation costs. PV systems are frequently designed to operate and interconnect with the electric utility grid. The main component in grid-connected PV systems is the inverter, or power-conditioning unit (PCU). The PCU converts the DC power into AC power which is consistent with the voltage and power requirements of the grid and automatically stops supplying power when the grid meets the power demand [19].

Electricity must be used as it is produced, but it can be stored as long as it is converted to another energy form (such as chemical energy in batteries) or used to pump water uphill where the hydrostatic power can be used to power turbines. The limitation of solar power is that the technology of transforming electricity into storable energy has not matured. To overcome the intermittency problem of solar power, a storage medium or energy carrier is required. There are three technologies that are currently used as viable energy storage solutions for solar power, i.e., smart batteries, thermal energy storage and hydrogen fuel cells. First, smart batteries can store energy generated by solar panels, which means there is no waiting for sunshine before starting up machines or appliances. The energy generated during the day can supply power at night. Thermal energy storage is commonly used with thermal solar power plants which generate high temperatures using mirror arrays rather than photovoltaic panels. The stored heat (e.g., molten salt) vaporizes water into steam to activate the turbine and electric generators during the night [20]. Fuel cells can be used as part of a solar–hydrogen energy cycle where a system converts water to hydrogen and oxygen. Hydrogen and oxygen are further stored by a fuel cell to produce electricity without sunlight. Large-scale energy storage solutions are still in their infant stages, yet these technologies will greatly influence the renewable energy industry.

Solar thermal systems concentrate sunlight to generate steam and require isothermal energy storage systems to store the energy. One storage option is the application of phase change materials to absorb or release energy [21]. Zalba et al. [22] provided a review of studies dealing with thermal energy storage using these materials. Kenisarin and Mahkamov [23] reviewed the current state of research in this particular field, focusing on the assessment of the thermal properties of various materials, methods of heat transfer enhancement and the design configurations of heat storage facilities. Some natural substances such as salt hydrates, paraffin, fatty acids and other compounds have high latent heat coefficients which are required for solar storage applications. The limitation of salt hydrates is chemical instability when heated, as they degrade at high temperatures and lose water in every heating cycle. Some salts are chemically aggressive towards structural materials. These two factors, poor stability in thermal cycling and corrosion between the phase change materials and the container, have limited

the widespread utilization of latent heat storage technologies [22]. For parabolic trough power plants, heat storage systems with operation temperatures between 300 and 390 °C are widely used. Tamme et al. [24] developed a solid media heat storage system which was tested in a parabolic trough test loop in Spain. The experimental results show the effects of changing parameters on the storage system. While the effects of the storage material properties are limited, the selected geometry of the storage system is important. Weather forecasting errors affect the power and load demand, and the economic performance of the PV power systems. Wang et al. [25] propose an adaptive solar power forecasting model for precise solar power forecasting. The model captures the characteristics of forecasting errors and revises the predictions by combining data clustering, variable selection and neural networks. The combined model approach uses the improved k-means clustering algorithm, the least angular regression algorithm and back propagation neural networks.

PV storage systems can be divided into off-grid, on-grid and hybrid systems. The off-grid system, or stand-alone system, consists of battery packages, photovoltaic charge and discharge controllers, battery packs, off-grid inverters and AC/DC converters. The controller manages the charging and discharging of the battery and protects the battery from over charging and completely discharging. The function of the off-grid inverter is to convert the DC power into AC power and provide it to a system or a utility grid. The design of a stand-alone system must take into account the capacity of the battery to be used at night, knowing the power load, predicting cloudy days and determining the requirements of solar cell module boards. The design is more complicated and more expensive. The typical application is used in high mountain areas, outlying islands or undeveloped areas without power grids. Figure 2 shows the operation concept of an off-grid storage system [26].
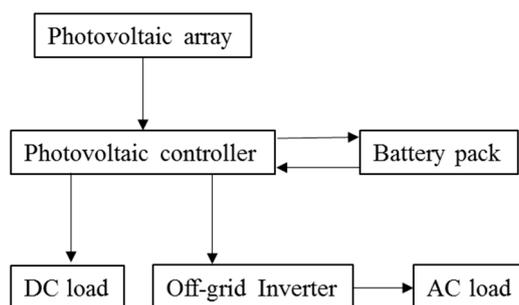

Figure 2. Schematic view of an off-grid system and key components.

The on-grid system consists of a PV array, a PV controller, battery packs, a battery management system, an inverter, an energy storage unit and a dispatch control system [27]. Solar panels convert light energy into electricity which charges the lithium battery pack. DC power is converted to AC power through the inverter. The controller continuously switches and adjusts the working state of the battery pack according to changes in sunshine intensity and the load status. The electricity is sent to the DC or AC converter for immediate use or the excess DC power is sent to the battery pack for storage. When power generation cannot meet the load demand, the controller uses power from batteries to ensure the continuity and stability of the system. The on-grid inverter system consists of several inverters, which convert the DC power from the battery into a standard voltage for the user-side low-voltage grid or for transmission to the high-voltage grids. The advantages are a safe and simple design, easy maintenance, with efficient solar energy generation that is higher than stand-alone systems. Figure 3 shows the operation concept of the on-grid storage system [26].

The hybrid solar photovoltaic system combines the on-grid system with more battery modules. The PV system generates power and supplies the load to charge the batteries simultaneously in the daytime, and then the power company supplies electricity at night. There is sufficient battery backup which makes the system suitable for public facilities. Hybrid systems are more complex to design and more expensive to set up. The system architecture is shown in Figure 4 [28].
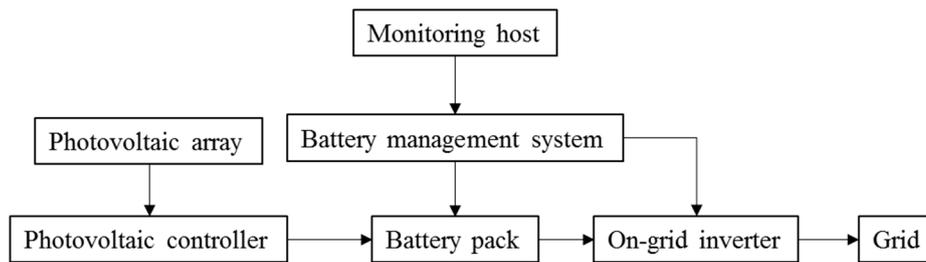
Figure 3. Schematic view of an on-grid system and key components.
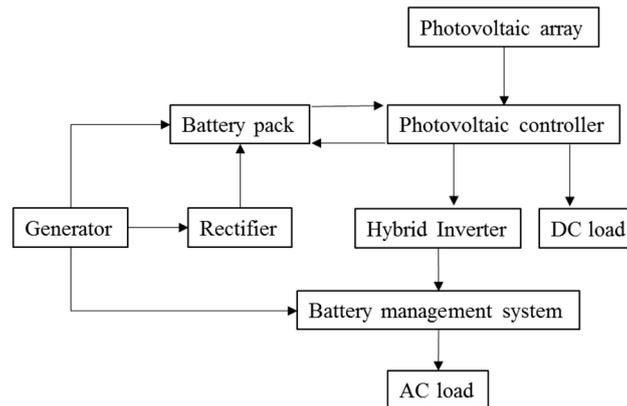

Figure 4. Schematic representation of a hybrid system and key components.

Utilizing battery storage systems can reduce the intermittent output of PV generation systems and store larger amounts of energy. Teng et al. [29] designed an optimal charging and discharging schedule for battery storage systems such that the power loss from transmission systems interconnected with large PV generation systems is minimized. A mathematical model to simulate the charging procedures was proposed in this study, and the minimum line loss problem considering intermittent output was built into the operations support system. The optimal charging and discharging scheduling of battery storage systems was obtained using a genetic algorithm. Zahedi [30] also proposed a model for a combined solar PV with batteries and super-capacitors that helps reduce power injection losses to the grid during peak demand.

Research and development of silicon heterojunction solar cells have seen a marked increase since the recent expiry of core patents [31]. Silicon heterojunction solar cells offer additional cost benefits compared to conventional crystalline silicon solar cells. Louwen et al. [31] analyzed the current cost breakdown of heterojunction designs using life-cycle costing and compared the results to conventional diffused junction monocrystalline silicon modules. The study showed that improvements in cell processing and module design results in a significant drop in production costs. The replacement of indium-tin-oxide was not found to contribute substantially to a reduction in module costs.

Stand-alone PV systems require energy storage to supply continuous energy when there is insufficient or no solar radiation. Valve Regulated Lead Acid (VRLA) batteries are sometimes used but supplying a large burst of current such as motor startup degrades the battery plates and can destroy the battery. A method of supplying large amounts of constant current is to combine VRLA batteries with super capacitors to form a hybrid storage system where the super capacitor supplies instant power to the load [32]. Podjaski et al. [33] proposed a type of solar battery material called 2D cyanimide-functionalized polyheptazine imide (NCN-PHI) which combines light harvesting and electrical energy storage within one single material. The charge storage of NCN-PHI is based on the photo reduction of the carbon nitride and the charge is stored by adsorption of alkali metal ions within the NCN-PHI layers. The photo reduced carbon nitride can thus be described as a battery anode working as a pseudo capacitor, which can store light-induced charge by trapping electrons for few

hours. The feasibility of light-induced electrical energy storage and release on demand by a single component light charged battery provides a unique solution for energy storage.

Wu and Mathews [34] in 2012 deployed a dataset of solar photovoltaic patents filed in Taiwan, Korea and China over the last 24 years (1984–2008). Their analysis of the knowledge in these patents resulted in a set of 12 International Patent Classification (IPC) technology categories. Commonalities in patterns of knowledge between solar photovoltaic and earlier industries are demonstrated. This study first identifies a comprehensive patent dataset for solar PV technologies then differentiates three generations using a three-stage patent extracting methodology. Scientific linkage is applied to investigate the development of knowledge flows for technologies such as solar cells and examines the causes and effects underlying the pursuit of this knowledge.

By reviewing literature, the ontology of solar power is constructed. Solar power generation technology is divided into three parts, PV technology that uses the photoelectric effect to directly transform sunlight to electricity, concentrated solar power that heats water into steam to power machines such as power turbines, and storage systems (e.g., batteries) for uninterrupted supply of electricity when sunlight is not available. Figure 5 illustrates the concepts and newly derived technology structure identified by the comprehensive reviews. The ontology schema, as a structured knowledge map, is iteratively constructed mainly from literature reviews (and can be updated by the state-of-the-art patent reviews), as detailed descriptions of key solar technologies and their relationships.
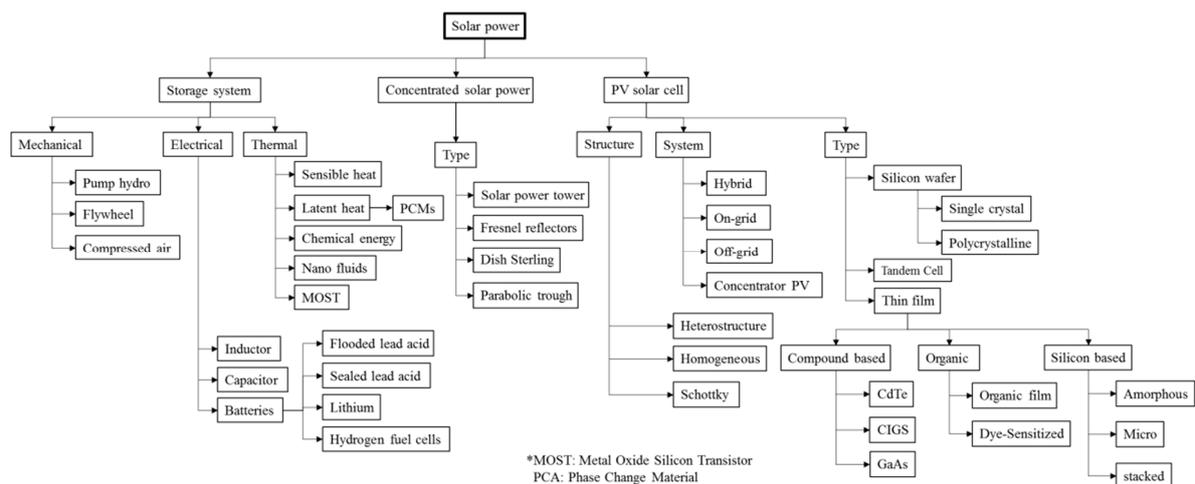

Figure 5. A structured knowledge ontology describes key solar power technologies.

2.2. Text Mining for Patent Analysis

The rapid pace of energy innovation places governments and enterprises in a difficult position to select economically suitable technologies over time. Patents are frequently used for forecasting technology trends and opportunities [35]. Trappey [36] used data mining to categorize renewable energy data and applied analytic hierarchies to evaluate policy goals using a clustering algorithm to segment the characteristics of the policies. Zhang et al. [37] constructed a mixed similarity measurement based on multiple indicators to analyze patent portfolios. Two models are proposed in this method: categorical similarity and semantic similarity. The categorical similarity model emphasizes international patent classifications (IPCs), while the semantic similarity model emphasizes patent text. For categorical similarity, fuzzy set routines are used to translate the IPCs into defined numeric values, and then the similarities between patent portfolios using membership grade vectors to calculate the cosine measures. The semantic similarities are calculated based on comparing the three-level core term tree structures of patent portfolios. A weighting model where values are determined using the analytic hierarchy process and expert knowledge measures the bias between the categorical and semantic similarities. Li et al. [2] proposed a framework that uses patent analysis and

Twitter data mining to monitor the emerging technologies and identify changing technology trends. The authors cluster topics using the Lingo algorithm and two domain experts filter the clustering results and name the cluster topics. Twitter users tend to pay more attention to wearable devices that are designed using environmentally friendly materials.

The technologies identified may also be applied to photovoltaic power generation devices. Sampaio et al. [38] described the technological development of PV cells using patents analysis. The results show that the PV patents are concentrated in three areas: PV semiconductor materials, direct conversion of light energy into electric energy, and solar panels adapted for roof structures. In addition, organic polymers, carbon nanostructures, compounds III-V and cadmium cells are considered to be the outstanding claims of photovoltaic cells patents.

Trappey et al. [39] proposed a roadmap approach to visualize patent evolution corresponding to multi-party logistic services. The relevant IoT smart logistic patents are analyzed to identify technology-oriented business strengths and strategies. From an industrial perspective, this approach has been proved to be an efficient and consistent way of technology monitoring under conditions of limited time and budget for technology development analysis. The approach reduces the effort required by domain experts to identify technological and helps define R&D strategies using roadmap visualization. Trappey et al. [40] also developed an ontology-based smart retailing patent roadmap analysis and valuation approach for developing competitive strategies. Text mining categorizes the patents as an ontological structure. The valuations of the patent portfolios provide insight of two companies' competitiveness in terms of their innovative business models and intellectual property advantages.

Clustering is an application of unsupervised machine learning and divides documents into groups based on their correlations [41]. A good cluster result has greater similarity within the same group but smaller similarity between different clusters. By exploring keyword terms that appeared in domain patents, patents with similar keyword terms are clustered into the groups. Trappey et al. [42] used Normalized TF-IDF to find the key terms in the corpus of 3D printing patents, considering different lengths of patent documents, for hierarchical clustering, K-means, and K-medoids to better analyze patent sub-technology clusters. Kim and Bae [43] also proposed an approach to forecast promising technologies by clustering patents. A symmetrical patent-patent matrix is constructed by calculating the Pearson's correlation coefficient between patent documents. Then, the k-means algorithm is used to cluster patents with the average silhouette width applied to determine the best number of clusters. The topic for clusters is defined by examining the combination of patent classification categories from each cluster. Finally, patent indicators such as forward citations, triadic patent families and independent claims are analyzed to summarize the promising technologies.

Word embedding was first proposed by Bengio et al. in 2003 [44], which is a technique that converts words in a sentence into a vector. The algorithm constructs a set of features for each word from the text and then distributes the features. This neural network-based language model allows machines to learn the relationship of words by calculating the distance between two vectors [45]. The words are mapped to the other space, which has the characteristics of injective and structure-preserving. When training the neural network model, each word is transformed from a high-dimensional vector into a continuous lower-dimensional vector. In addition to finding correlations between words, word embedding serves as the basis for downstream natural language processing tasks such as text categorization, text clustering, part-of-speech tagging and sentiment analysis [46]. The concept of word embedding has been widely applied in natural language processing and many studies have been proposed such as Google's word2vec, Facebook's fasttext and Stanford's Glove. Mikolov et al. [47] proposed two model architectures for computing continuous vector representations of words from very large data sets. The quality of these representations is measured in a word similarity task, and the results are compared to the previously best performing techniques based on different types of neural networks. Tang et al. [48] proposed a method that learns word embedding for Twitter sentiment classification which encodes sentiment information from the continuous representation of words.

Specifically, three neural networks are developed to effectively incorporate the supervision from sentiment polarity of text in their loss functions.

3. Methodology Applied in This Research

There is some literature reporting their patent analyses using machine learning approaches as reviewed in the previous section (Section 2.2). However, this study develops a novel framework for patent technology mining, combining multiple, un-supervised machine learning algorithms in a specific workflow. The other uniqueness of this research is that the text mining approaches are implemented in both document sets (i.e., patents and literatures), which are analyzed respectively, and then cross-compared to explain the technology trends for enhancing the explanatory capability and reliability. The novel methodology process flow is shown in Figure 6. This research used 2280 global patents and 5610 academic literature documents as the document corpora for technology mining. First, the database containing both patent and literature documents related to solar power technologies were collected. The key machine learning algorithms, including clustering, topic modeling, word embedding, document similarity analysis and technology evolution mapping, were used in sequence to identify key technologies and their patenting evolution. The detailed algorithms and the key references for further theory understandings are described in detail in Sections 3.1–3.4.
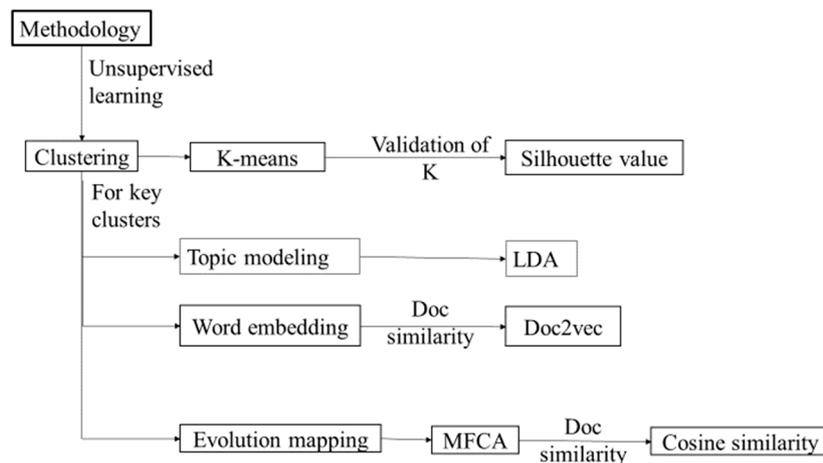

Figure 6. Structure of proposed machine learning methodologies.

3.1. Clustering

In this research, k-means was used as the algorithm for clustering as proposed by Hartigan and Wong in 1979 [49]. The principle is that given a set of observations (x_1, x_2, ... , x_n), where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k ($\leq$n) sets S = {S_1, S_2, ... , S_k} so as to minimize the within-cluster sum of squares variance. The objective function I defined by formula (1), where μ_i is the mean of points in Si.

$$\underset{}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x \in s_i} \|x - \mu_i\|^2. \tag{1}$$

To determine the k values, the number of cluster groups with best performance, Rousseeuw [50] proposed silhouettes as a graphical aid to the interpretation and validation of cluster analysis. Let *b(i)* be the highest average distance of *i* to all points in any other cluster. The cluster with the highest

average dissimilarity is selected as the neighboring cluster of *i* since it is the best fit cluster for point *i*. The silhouette formula (2) is defined as follow:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}},$$

(2)

where $-1 \le S(i) \le 1$, the silhouette value is used to find the best k value or number of clusters. The users defines a range of clusters as an input value to be examined in validation program. After determining the number of clusters, the key terms are placed into a multi-dimension vector matrix and the cosine similarity Equation (3) is used to calculate the relation distance between clusters.

$$\text{similarity}(i, j) = \frac{\langle X_i, X_j \rangle}{\|X_i\| \|X_j\|} = \frac{\sum_{k=1}^{n} X_{ik} X_{jk}}{\sqrt{\sum_{k=1}^{n} X_{ik}^2} \sqrt{\sum_{k=1}^{n} X_{jk}^2}}.$$

(3)

For the research case study, the python scikit-learn package is applied. The basic functions of scikit-learn include classification, regression, grouping, model selection and data pre-processing [51].

3.2. Modified Formal Concept Analysis (MFCA)

Mapping the evolution of patents originated from Wille's method in 1982 and is called Formal Concept Analysis (FCA). By formalizing the common attributes of patents and transforming them into conceptual lattices, analysts are able to define the associations between patents. Modified formal concept analysis (MFCA) was proposed by Lee et al. in 2011 [52] to include time as an attribute linked to key terms to enable the creation of maps to depict the evolution of patents. Trappey et al. [53] proposed a patent evolution method to explore 3D printing innovations. The extracted and ranked key terms are treated as attributes in patent clustering and similarity analysis, which are extracted by normalized term frequency (NTF). MFCA is applied to analyze technology trends, and the results are graphically displayed for four patent clusters. k-means is used as part of the MFCA process and the NTF matrix is built with key terms. The concept of patent evolution is shown in Figure 7. The patents spread out from P1, P2, P3 to P4, P5, P6 and P7. The circles represent the evolution time (year) of patents, the dots represent each single patent, and the lines connecting two dots represent the relation between patents. If the similarity of two patents exceed a threshold value, the line is drawn as a solid line, otherwise as a dotted line.
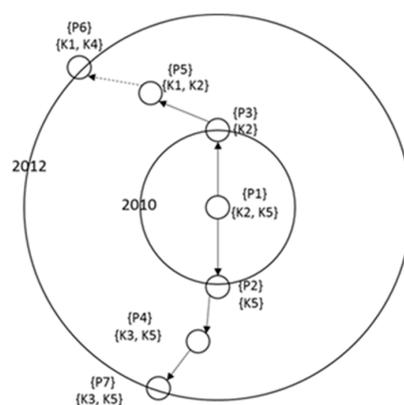

Figure 7. Concept diagram for mapping patent evolution.

3.3. Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a statistical model used to explain the similarity of data. The concept of LDA originated from a population genetics study in 2000 as first proposed by Blei et al. [54]. LDA assumes that each document is a mixture of a few topics and that each word fits within some topic of the document. LDA is a topic model, where each document has a topic

probability distribution, and each topic has a word probability distribution. LDA can better eliminate word ambiguity and assign documents more accurately to topics [55]. Zou [56] developed a smart method for building an ontology using LDA topic modeling and identifying the key phrases under each topic from a large number of patent documents. The patent documents are clustered using the k-means and hierarchical clustering methods and then the LDA topic model is built based on each cluster. The number of topics is determined by researchers by observing the model training results. After the topic models are established, the key phrases under each topic become the output and the ontology of the domain patents is constructed. The ontology architecture is shown in Figure 8 [56]. The topics are held constant and the time information within the model is treated as a variable and used to discover these hidden topics [57]. Changes in words along with changes in time are used to detect topic patterns. Doucet et al. provide an approach which allows a model to change over time using sequential importance sampling or particle filtering. Canini et al. describe an implementation of online LDA framework with particle filters that yields better results than multiple LDA runs.
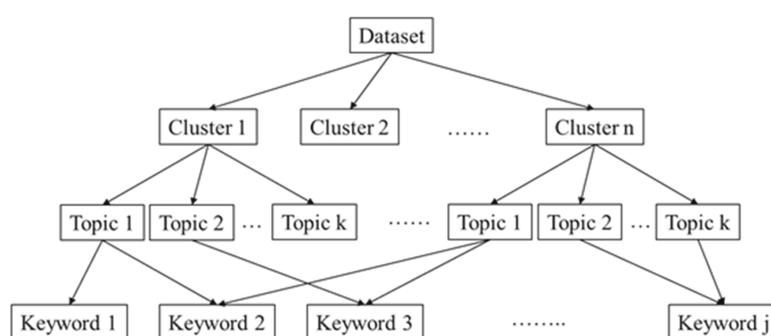

Figure 8. Ontology architecture diagram.

Most topic modeling algorithms that address the evolution of documents over time use the same number of topics which means that new topics arise and old ones disappear. Wilson and Robinson [58] proposed an algorithm to model the birth and death of topics within an LDA-like framework. The user first selects an initial number of topics, and then the new topics can be created or retired without supervision. The algorithm of this research provides initial topics. The first step computes the drift of any topic with respect to its counterpart, where each topic is a probability distribution. This allows the application of the Hellinger convenient divergence measure. After computing the drift for all topics in a specific epoch t, it can determine if any have changed enough to generate a new topic. The modified Z score is used to identify the central tendency of each topic, and to determine which topics have drifted too far and need to be split. On the other hand, old topics are combined into a larger discussion or dropped entirely. The method measures the number of tokens assigned to each topic. Topics with fewer tokens are placed on probation. If a topic stays on probation for more than 10 epochs, then it is marked as closed.

In this research, topic models are built under each cluster in order to better describe a cluster using the word distribution list. After the topic models are generated under each cluster, patent documents are assigned to each topic model by matching the keywords of patents and topic and calculating the similarity scores. The original concept of the assignment measure is to check the number of words appearing in the top patent keywords and topic model which indicates the two documents are more similar. The sequence of keywords is considered using the modified assignment measure weighting algorithm shown in Figure 9. First for each cluster, T(i) is set to be the top 50-word list of topics(i); then for each patent document k, P(k) is set to be the top 50 keyword list of patent(k) and W is the intersection words between P(k) and T(i). The similarity score of patent(k) and topic model (i) is the accumulated score of the 50-sequence of W in P(k) and T(i)), where an intersection word is assigned a higher score if it located higher in the keyword list. If the word score is higher than the threshold number, the patent (k) is assigned to topic(i).

```
for each cluster:
    set T(i) = top 50 words list of topic(i)
    for each patent document k:
        set P(k) = top 50 keywords list of patent(k)
        set W = intersection words between P(k) and T(i)
        word score += ( 50 - sequence of W in P(k) and T(i) )
        if word score > threshold:
            topic(i) ← patent(k)
    Print(topic(i))
```

Figure 9. Algorithm of topic modeling and patent topic assignment.

3.4. Doc2vec

Doc2vec was proposed by Le and Mikolov [59] as an unsupervised algorithm that learns fixed-length feature representations from variable-length pieces of text such as sentences, paragraphs and documents. The algorithm maps each document into a dense vector which is trained by a neural network to predict words in the document. The approach overcomes the weaknesses of bag-of-words models by ignoring the order and semantic information of words and is an extension of word2vec [60]. The learned vectors can be used to find the similarity between the terms, paragraphs and documents by calculating the distance, which is further applied for text clustering.

4. Solar Technology Patent Review and Analysis

This research focuses on the domain-specific patent review and analysis of the state-of-the-art solar (PV) technology. Following the technology mining process flow in Figure 1, both domain patents and academic papers (literature) were searched and collected as document corpora related to solar power technology. The literature review part was described thoroughly in Section 2.1 to form the solar technology ontology. This section will describe the detailed study of the solar technology patent review and analysis. First, the search query for patent dataset was determined, and then the statistical patent analysis was conducted. The text mining for both patents and literature are introduced in the following sections.

4.1. Patent Search

This section describes the search strategy, statistical and text mining analysis for solar power related patents. The search shown in Table 1 is the patent search query related to the technology specifications. The focus is on the energy generation, supply and storage systems for solar power. The geographical scope includes the United States of America, China, Europe, World Intellectual Property Organization (WIPO) and Australia over a period of 10 years (2008–2018). The result provided by the Derwent Innovation search platform yielded 2280 patents, which are systematically analyzed in the following sections.

Table 1. Patent search query strategy and strings.

Title DWPI	**(batter* or storage* or material* or generation or supply) and (solar or concentrate* or photovoltaic or PV or CSP)**
Abstract DWPI, First claim DWPI, Title/Abstract/Claim	solar and (batter* or storage* or material* or generation or supply) and (concentrate* or photovoltaic or PV or CSP) and (electri* or energy or power) and ((silicon or (thin film) or compound or organic or dye-sensitized) or (hybrid or on-grid or off-grid) or (thermal or heat or sensible or latent or pump or (nano fluids)))
Year	2008–2018
Country	US, China, Europe, WIPO, Australia

4.2. Statistical Analysis of Patent Metadata

Statistical information provides experts with a preliminary understanding of emerging patent trends. In this research, there were a total of 2280 patents found that matched the search domain and included 2054 DWPI (Derwent World Patents Index) families. Based on this search result, statistical analysis and text mining were applied. The objective was to analyze the leading assignees, IPCs, countries and patent publishing trends. From these results, it is possible to illustrate and describe global industrial development of solar power. China owns the most patents in this domain, which accounts for over 90% of overall patents. The US owns the second most patents, but only accounts for 4.3%. China is the world's largest market for solar photovoltaics and solar thermal energy and is the second largest country in energy consumption. Greater energy demand may be driving the Chinese government's efforts to revise their energy policies to support sustainable energy development [61].

The patent publishing trend shows that the number of patents published have continued to increase over the last 10 years with China the largest contributor of solar power technology. China is the world's leading solar PV installer, and the solar photovoltaic industry in China is a growing industry with more than 400 companies. During the period of the 11th Five-Year Plan (FYP) from 2006 to 2010, China's PV industry developed rapidly and became one of the few industries that could compete globally. China has been the largest PV manufacturing nation since 2008 when it became the largest producer of solar panels in the world. During 2011 and 2012, in China government implemented a series of incentives, including direct subsidies for solar PV installations and a national feed-in tariff scheme [62]. China's domestic PV market has seen steady growth with increasing cumulative installed capacity. However, China's PV products rely heavily on foreign markets. The orders from the European and American markets have declined due to the economic crisis and tariff protection around 2008, which decreased the number of patents published. In 2012, the US Department of Commerce ruled on imposing anti-dumping duties on Chinese solar photovoltaic module products exported to the US [63]. Along with the PV subsidy policy cancellation in European countries, the reduced orders led to China's overcapacity and subsequent decline in R&D patenting in 2014 [64]. In November 2016, The Commonwealth Scientific and Industrial Research Organization [65] signed a technology patent transfer agreement with the Chinese solar company Thermal Focus. The concentrating solar thermal power generation technology (CSP) developed in Australia was transferred to China which may have influenced the patent publishing numbers as they dropped in China during 2017.

The top seven assignees are all from China including four academic institutions (universities) and three energy-oriented technology companies. It means that China owns the largest share of IP technologies in terms of solar power supported by development and innovation from academia and industry. Wuxi Tongchun Energy Technology Corporation is the assignee that owns the most patents and much more than the top two assignees. The company is creating new energy products, epoxy boards, and other innovations such as sun loungers for the elderly in smoggy weather. The fifth largest assignee, State Grid Corporation, is an institution and state holding company that has been approved by the State Council of China to conduct state-authorized investment. The company constructs and operates the China Power Grid and supplies national electricity. The number seven assignee, Fuzhou Aquapower Electric Water Heater Corporation, is focusing on the development, manufacture and sale of solar water heaters. Wuxi Tongchun used to be the largest technology developer before 2015, however, Tianjin University and State Grid Corporation became the top assignees during the following three years.

The leading International Patent Classification (IPC) is H02S which involves converting infrared radiation, visible light or ultraviolet light to generate electrical power. The second IPC, H02J, relates to circuit devices or systems for power supply or distribution, and electrical energy storage systems. The third IPC, H01L, is related to semiconductor devices or electric solid devices not included in other categories. The fifth most important IPC, F24J, relates to heat generation devices not included in other categories. Detailed IPC statistics show that the leading IPC is H02S004044, which is a method of utilizing thermal energy such as a system generating both warm water and electricity. The second

leader, H02J00735, is a photosensitive battery. The third is H02S004042, which is a cooling method, the fourth is H02J000700 which is circuit device for charging or depolarizing a battery pack or for supplying power from a battery pack to a load. These statistics show that the solar power technology is trending toward transforming solar power into thermal energy by hydroelectric systems. Both H02S and H02J IPCs account for the most patent technologies. A01G appears in the top IPCs but not in the top rank of total IPCs which relate to horticulture, flower cultivation and watering systems. Plant cultivation is becoming one of the most popular applications of solar power. More detailed IPC analysis shows that H02S004042, H02S004044 and H02J00735 are the three dominant technologies which are also in the top three IPCs.

In addition to IPC, the statistical results of the Cooperative Patent Classification (CPC) system was also analyzed [66]. Since 2013, The European Patent Office (EPO) and the United States Patent and Trademark Office (USPTO) officially implemented the CPC, which has become a global patent classification system. The leading CPC Y02E001060 relates to thermal-PV hybrids technologies, the second CPC Y02E001050 relates to PV energy, the third CPC Y02E001044 relates to heat exchange systems, and the fourth most important CPC relates to solar thermal, hybrid systems, PV systems with concentrators and solar thermal energy. The CPC trend of solar technology shows that development is gradually turning to thermal-PV hybrids technologies instead of individual PV or thermal technologies commonly used over the past 10 years.

4.3. Patent Analysis by Text Mining

Python text mining programs are used to generate the unsupervised learning results. The Python packages used are described in this paragraph. Pandas is a package to transform data so that it is easy-to-handle, including the data frame and series where data can be split and combined. Csv and xlrd are used to read Excel files by filename and the Re and Nltk manage text preprocessing such as reduction, tokenization and stop word or punctuation removal. Numpy is widely used to store data into arrays for mathematical operations. Gensim has many functions for text mining and natural language processing. LDA and Doc2vec are used in this research. Sklearn is used for data mining and data analysis including tfidf, cosine similarity and K-means.

For clustering, the number of k is set using the Silhouette validation approach where the clustering results improve with higher values [67]. Some variables were adjusted to optimize the clustering results including adding more stopwords and revising the searching query string to improve the dataset. Better clustering results provide a clearer interpretation with respect to the technology content in each cluster. The values of five to ten clusters yielded values of 0.153, 0.159, 0.164, 0.128, 0.131 and 0.133, whereas seven clusters have the highest Silhouette value.

The collected patents are distributed into seven clusters and the publishing trend for each cluster is shown in Figure 10. Most clusters have an increasing trend with some dropping in the same period, which follows a similar trend for total published patents (Figure 10). Clusters 2 and 5 are assigned the most patents, with a continuous increase of patent numbers representing the largest global technical field in solar power. In addition, these two clusters are growing with stability compared to other groups.

For these seven clusters, the topic contents are defined and symbolic key words are listed in Table 2. Cluster 1 describes grid-connected energy storage systems and silicon based or lithium ion solar cells which are identified by the keywords, such as battery, connect, storage, inverter, charge, grid, wire, install, switch, conduction, AC and DC. Clusters 2 and 5 describe solar hydropower storage systems. Cluster 2 focuses on transferring waterpower into electricity especially for irrigation systems (keywords: water, pump, storage, valve, pipeline, circulation, seawater, irrigation and condenser), but Cluster 5 relates to light absorbing materials of CSP (keywords: heat, thermal, exchanger, collector, light, material, surface, absorb, steam, evaporator and medium). Cluster 3 aggregates solar battery modules for both thermal hydro and PV solar systems (keywords: module, battery, circuit, light, voltage, charge, sensor, inverter, board, array and signal). Clusters 4 and 7 describe some new materials of thin film PV cells. Cluster 4 focuses on silicon based and compound based thin film

cells (keywords: layer, film, silicon, material, oxide, thin, structure, electrode, coat, insulation, metal, PV and Nano), but Cluster 7 focuses on silicon based and organic materials (battery, light, film, silicon, surface, component, thin, material, organic, crystal). Cluster 6 describes air processing systems of CSP, especially for fluid heat conduction mediums (keywords: air, heat, storage, water, dust, indoor, channel, purification, sensor and greenhouse).

Table 2. Description of the patent clusters.

| Cluster	Topic Content	KEYWORDS/Key Phases	Patent Numbers (Total, CN, US, WO, EP)
1	Grid-connected energy storage system, silicon based or lithium ion solar cell.	Battery, connect storage, inverter, charge, grid, wire, install, switch, conduction, AC, DC, conductive wire, lithium ion	321, 317, 3, 1, 0
2	Solar hydropower storage system, focusing on transferring waterpower into electricity especially for irrigation systems.	Water pump, storage tank, valve, pipeline, circulation, seawater, irrigation, condenser, pipe connect, heat collector	619, 615, 3, 0, 1
3	Solar battery module for both thermal hydro and PV solar systems.	Module, battery module, circuit, light, voltage, charge, sensor, inverter, board, array, signal	156, 148, 5, 1, 2
4	Material of thin film PV cells (focusing on silicon based and compound based).	Battery, thin film, crystal silicon, material layer, oxide, coat, electrode, insulation, metal, PV, Nano, glass layer	197, 175, 20, 0, 2
5	Solar hydropower system focusing on light absorbing materials of CSP.	Heat, thermal, heat exchanger, heat collector, light, material, surface, absorb, steam, evaporator, medium	454, 402, 36, 8, 8
6	Air processing system of CSP especially for fluid heat conduction.	Air conditioner, heat, storage, water, dust remove, indoor, channel, purification, sensor, greenhouse	162, 155, 7, 0, 0
7	Material of thin film battery and PV cells focusing on silicon based and organic materials.	Battery board, light, thin film, surface, component, material, organic, crystal silicon	371, 342, 24, 1, 4
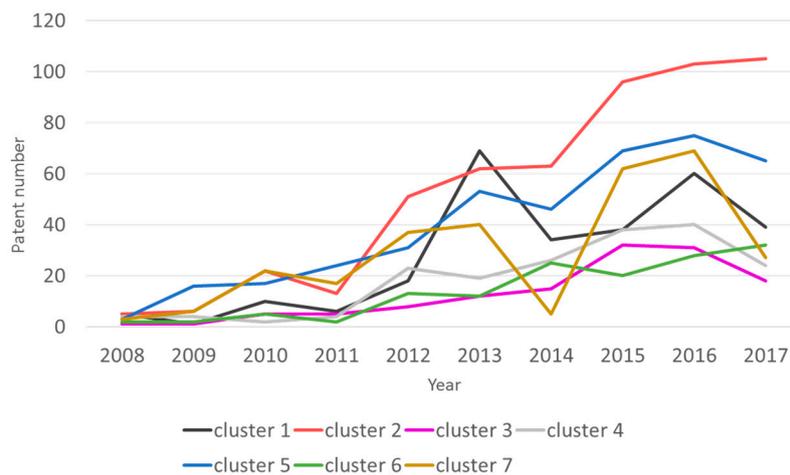
Figure 10. Patent trend of all clusters.

4.4. Literature Analysis by Text Mining

Clustering is also applied to academic literature. A total of 5610 documents are distributed into seven clusters. The academic publishing trend for each cluster is shown in the Figure 11. The seven clusters have similar increasing trends with Clusters 3 and 2 showing similar increases in research activity.

The topic contents for each cluster and the corresponding key words or phrases are listed in Table 3. Cluster 1 relates to CSP thermal collecting and storage systems with keywords of CSP plant, steam, efficiency, parabolic trough, thermal storage, heat and receiver. Clusters 2 and 5 are related to the simulation of grid-connected energy storage and supply systems, with Cluster 2 targeting on-grid systems and DC electricity supply (keywords: gGrid, load, voltage, PV module, simulation, DC, battery, capacity, network, electricity, demand and installation). Cluster 5 collects research on off-grid systems and AC electricity supply management (keywords: grid connect, PV, battery, off grid, simulation, management, converter, charge, network, AC, smart grid). Clusters 3 and 6 are related to integration of renewable energy generation systems (hybrid systems) such as the combination of wind power and solar power. Cluster 3 focuses on the economic impact of the major renewable systems and the integrated application of solar power and the other systems (keywords: renewable, electricity, battery, capacity, grid, electricity, wind, reduce, carbon, emission, PV, integration). Cluster 6 integrates hybrid systems of wind and solar or other renewable power generation systems combined with grid networks for electricity storage (keywords: hybrid, wind, PV, battery, diesel, generator, renewable, grid, resource, storage). Cluster 4 defines novel materials of PV cells (especially organic thin film), with high conversion and absorption efficiency (keywords: charge, cell, conversion efficiency, electron, organic, material, light, polymer, layer, acceptor, absorption, perovskite, PV). Cluster 7 discusses novel materials for heat collection with high heat transfer efficiency and light absorption (keywords: thermal storage, material, efficiency, collector, cycle, molten salt, fluid, heat transfer, absorber).

Table 3. Description of literature clusters.

| Cluster	Topic Content	Keywords/Key Phases
1	CSP thermal collecting and storage system	CSP plant, steam, efficiency, parabolic trough, thermal storage, heat, receiver
2	Simulation of grid-connected energy storage systems and DC electricity supply	Grid, load, voltage, PV module, simulation, DC, battery, capacity, network, electricity, demand, installation
3	Integration of renewable energy generation system such as wind and solar power.	Renewable, electricity, battery, capacity, grid, electricity, wind, reduce, carbon, emission, PV, integration
4	Material of PV cells (especially organic thin film), with high conversion and absorption efficiency.	Charge, cell, conversion efficiency, electron, organic, material, light, polymer, layer, acceptor, absorption, perovskite, PV
5	Simulation of on-grid and off-grid (stand alone) network systems and AC electricity supply management	Grid connect, PV, battery, off-grid, simulation, management, converter, charge, network, AC, smart grid
6	Hybrid system of wind and solar or other renewable power generation systems combined with grid networks to store energy.	Hybrid, wind, PV, battery, diesel, generator, renewable, grid, resource, storage
7	Material of heat collector, with high heat transfer efficiency and light absorption	Thermal storage, material, efficiency, collector, cycle, molten salt, fluid, heat transfer, absorber
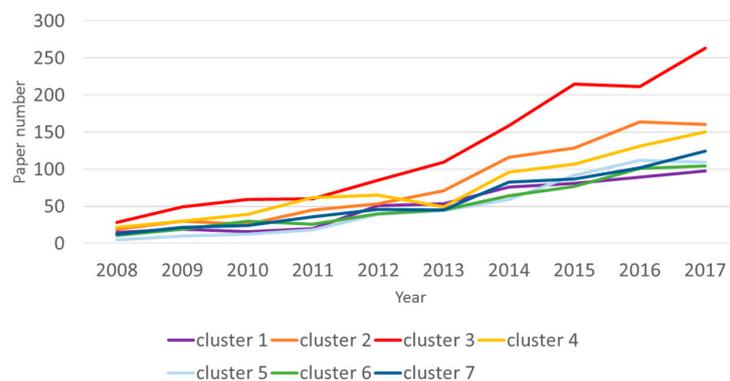
Figure 11. Literature paper trend of all clusters.

4.5. Cross-Comparison analysis

The technology development between patents and academic literature is quite different. As previously stated, most patents describe solar hydropower storage systems with a variety of subsystems relevant to indirect solar collection technology. Academic literature most frequently proposes new frameworks or algorithms for grid-connected electricity supply systems. The proposed systems use novel light absorbing materials for photovoltaic panels to lower the cost of the system and improve manufacturability. New system frameworks must be carefully planned for precise implementation and integration since new approaches require complex examination processes and other various factors such as social and government acceptance. Systems are relatively easy to implement if they are improvements based on existing systems. Conversely, it is very difficult to implement innovative systems and algorithms as a first attempt. Therefore, it is reasonable that there are more novel technologies describing integration of renewable energy generation systems and simulation of grid-connected energy storage systems in the literature, while technologies describing solar hydropower storage system are presented in patents.

The doc2vec model is trained using all the patent files in this research domain, so the user can put any group of files within a specific technology domain to generate a list of the most relevant patents. The number of output patents can also be determined by the user. The input data for Case 1 contains 107 academic articles selected from the key field literature dataset. The target field covers Topic 1 under Cluster 3 between 2016 and 2018, which is the largest topic group from the largest cluster. The 10 output patents having the highest similarity with the input data are shown in Table 4. The output patents are evenly distributed over the years 2010 to 2018, indicating the technology has developed constantly. The content of the patents focuses on the application of PV solar cells such as mobile power supply systems, mosquito killing devices, greenhouse roof heating systems and immune identification systems.

Table 4. Doc2vec output for Case 1.

Patent Number	Year	Title
CN201590776U	2010	Mobile photovoltaic power supply system.
CN201983524U	2011	Multi-energy complementary ground source heat pump system based on independent power generation.
CN202519366U	2012	Pipeline for silicon wafers diffusion of polycrystalline silicon PV cells.
CN103461299A	2013	Solar power mosquito killing device.
CN104032982A	2014	Heating and cooling system of power generation for solar greenhouse roof.
CN204762528U	2015	Zero energy consumption constant temperature greenhouse lighting system.
CN102496650B	2015	Manufacturing method for solar cell module.
CN107374775A	2017	PV intelligent identification for immune system.
CN104032982B	2018	Heating and cooling system of power generation for solar greenhouse roof.
CN206838588U	2018	PV panel cleaning device.

For Case 2, the integration of smart grid and intelligent electricity management systems are reviewed. The search was for novel technologies combining solar power systems (including power generation, storage and supply systems) with cyber-physics systems and big data generation. The input data contained 24 articles. The input literature is selected from Cluster 2 of the literature dataset which describes the simulation of grid-connected energy storage systems and DC electricity supply. In order to obtain the latest technologies, the target literature is published in the year 2018. By training the

doc2vec model, Table 5 shows 10 output patents which are most similar to the input documents. The 10 recommended patents are quite new (all published after 2013) compared to the whole patent dataset (published from 2008 to 2018). Second, most patents are from Cluster 1 which describes grid-connected energy storage systems which conforms to the case target domain. All of the patents have an intelligent adjustment function and process automation. For instance, WO2017210402A1 developed a self-balancing photovoltaic energy storage system to manage the energy storage and supply. CN103452164A invented a new type of automatic solar air intake device with temperature sensor to make the adjustment task intelligent. CN106169777A developed a micro-grid system with DC and AC hybrid power supply by connecting a micro-grid battery, DC distribution manager and photovoltaic power generation inverter in parallel to achieve intelligent transmission of power.

Table 5. Doc2vec output for Case 2.

Number	Title	Year	Cluster
CN102913978A	Renewable energy and building integrated utilization system	2013	1
CN103452164A	A new type of automatic solar air intake device	2013	7
CN204575198U	Optical signal sensor for high concentration solar power generation	2015	5
CN205427579U	Intelligent irrigation water and fertilizer integrated device	2016	1
CN106169777A	Micro-grid system with DC and AC hybrid power supply	2016	1
CN106440008A	Integrated energy utilization station for tea garden and comprehensive energy utilization method	2017	1
WO2017210402A1	Self-balancing photovoltaic energy storage system and method	2017	1
CN106525130A	A Bluetooth technology based on wireless quality integrated sensor device	2017	1
US20170370250A1	Combined energy supply system of wind, photovoltaic, solar thermal power and medium-based heat storage	2017	2
CN206853416U	Solar-assisted coal-fired unit carbon capture heating energy management system	2018	1

The clusters and evolution pathways for solar energy patents using the concept lattice algorithm are shown in Figure 12. A total of 52 patents from three categories (grid-connected energy storage systems, solar hydropower storage systems and thin film battery and PV cells) between 2014 and 2018 are displayed in the five concentric circle. Each data point represents a patent document, and is connected to the others in terms of the cosine similarity value. The green solid lines identify the connection where the similarity value is larger than 0.99. The green dotted lines identify the connection where the similarity value is between 0.9 and 0.99. From these patents, this research focuses on the category of grid-connected energy storage systems, which is most relevant to the Internet of Things technology. Figure 13 displays the stem path in this category with the key terms of each plotted patent. Patent CN104184394A describes household off-grid photovoltaic power generation systems (key terms: battery group, AC module, off-grid, switch, light, radiation, convert). Patent CN204179991U describes solar photovoltaic power generation devices (key terms: storage, battery, array, inverter, double switch, generate, off-grid, charge, load, efficiency). Patent CN105743429A describes off-grid photovoltaic power generation systems based on Internet of Things (key terms: inverter, array, load, DC converter, wireless receiver, transmitter, emitter, grid internet, connect). Patent CN106169777A describes a micro-grid system with DC and AC hybrid power supply (key terms: battery, connect, generate, assembly, hybrid, inverter connect, manager, AC, DC, allocation). Patent WO2017210402A1 describes self-balancing photovoltaic energy storage systems and methods (key terms: storage, DC, hybrid cell, connect, self-balance, direct, inverter, maximum, algorithm, plurality, conversion). Patent CN106525130A describes a Bluetooth technology using a

wireless sensor device (key terms: water pump, battery, liquid, sensor, seawater, electrode, Bluetooth, wireless, integrate, supplementary). Patent CN206834764U describes photovoltaic power generation systems (key terms: connect, grid connect, battery, inverter, connect inverter, array module, convert, distribution). The assignee of patent CN106525130A, Tianjin University, is the top three assignee which implies the high value of this patent as well as the evolution analysis.
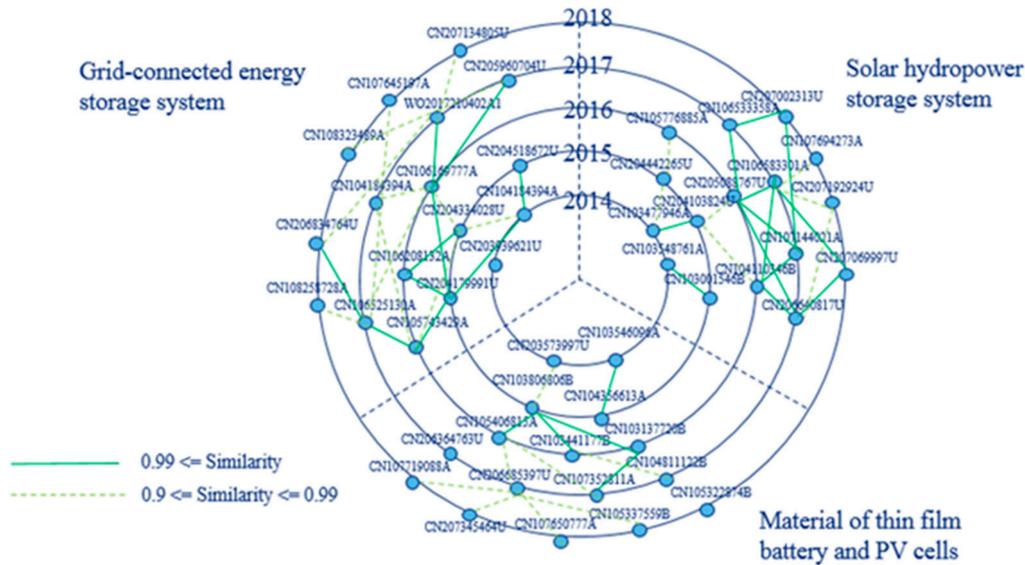

Figure 12. Evolution graph for three clusters of solar patents.
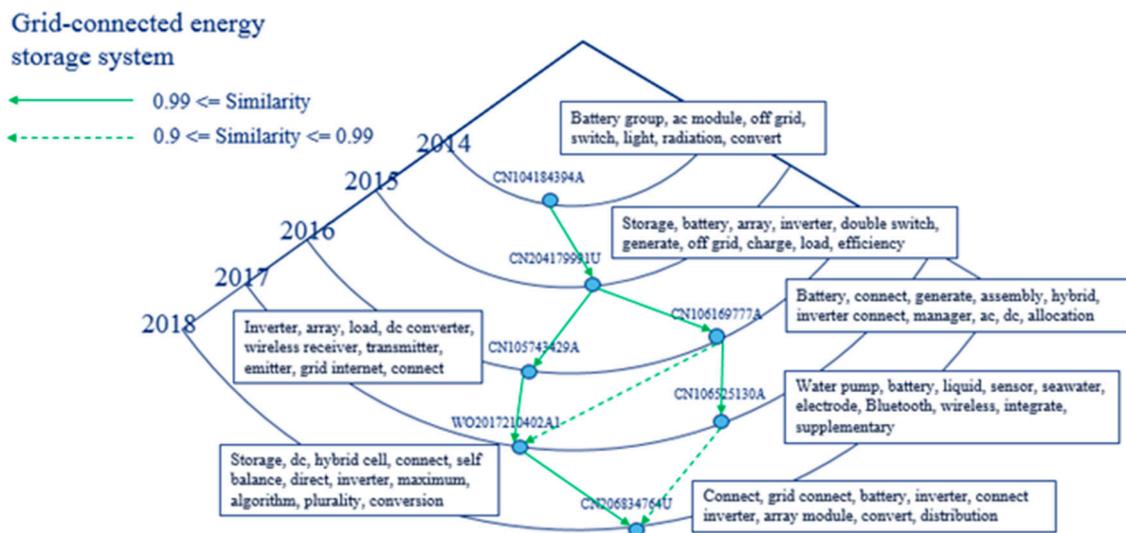
Figure 13. The key terms of the evolving patents in the grid-connected energy storage technology cluster.

The mainstream technology evolves from off-grid to grid-connected systems including technologies such as self-balance storage systems and wireless integrated sensors which are critical for smart grid networks. The evolution graph shows that solar technology is trending toward intelligent energy supply systems. The smart grid electricity supply system can be integrated with cyber-physics systems and the renewable resources industry. Smart battery management and supply balance systems are essential parts of the cyber physical system.

When comparing the seven patents (in Figure 13) to the related literature clusters and topics (in Table 4), we found that the patents are relevant to literature Clusters 2 and 5, which consist of articles in the (modular) simulation systems of on-grid and off-grid (stand-alone) networks and DC electricity supplies. More specifically, the papers under Topic 4 in Cluster 2 are most similar to

the target patents, covering the technical issues of "grid-connected solar energy storage systems or their intelligent management systems." Among these papers, earlier papers focus more towards the off-grid related technologies. For instance, a study in 2009 proposed a fuzzy logic control module of stand-alone PV system with battery storage [68]. Another study [69] in 2012 proposed a control method of the stand-alone direct-coupling PV-water electrolyzer. However, papers published recently focus more on grid-connected network systems and the real time algorithms. For example, in 2014, Sridhar and Meera [70] developed a grid-connected solar PV system using a real time digital simulator. Furthermore, Li et al. [71] in 2018 investigated the performance of a grid-connected residential PV-battery system focusing on enhancing self-consumption and peak shaving in Japan. Petrollese et al. in 2018 [72] also studied coordinated control for grid integration of a PV array, battery storage and super-capacitor. The patent clustering and evolution trends, as depicted in the case study, match well with the results of literature clustering and topic mining when cross referencing comparisons are conducted. The results strengthen the reliability of the technology mining and reviews for the solar power energy industry.

5. Conclusions

In this study, the academic literature and patents are reviewed to construct the knowledge domain ontology for solar energy and the derived subcategories for PV solar cells and concentrating solar thermal power generation technology (CSP). The ontology defines the relationships between the existing technologies of solar energy. By analyzing the statistical data and using text mining, the key development fields are discovered. The statistical data provides critical information attributes including top assignees, the top IPCs and the publishing year. The values are consolidated so that the technology R&D evolution trends can be tracked. For text mining, both patents and academic literature are collected to define the current research and development in terms of solar energy generation. The comparison between academic and industry R&D strategies are compared by text mining the results of both datasets. For text mining, there are four machine learning approaches, i.e., clustering, topic modeling, doc2vec and patent evolution graphs, used for deriving analytical results.

Clustering is used to group both academic literature papers and the patent documents into clusters. The trend for each patent and literature cluster is shown in Figures 10 and 11. Extracting keywords is an important process of this research. Without proper and precise keywords, the documents will not be properly assigned to the corresponding clusters. Therefore, normalized term frequency-inverse document frequency (NTF-IDF) is used to identify the key terms and avoid the document length problem. Further, the "silhouette value" is calculated to examine the cluster performance for the ideal number of clusters for any given document set. The technologies are divided into seven groups and topic models are generated under each cluster, for both patents and research articles. Using the key word distribution within each technology cluster, key clustered technologies are defined. For example, a user interested in the most mature technology can select the largest cluster with the most patents. In this cluster, the user is able to see the patent publishing trends and understand the document content in the cluster by checking the key words ranking and IPC ranking. To classify the documents in more detail, the user can refer to the topic model output within clusters. After the key domain technologies are discovered, results can be mapped to the ontology to identify research development opportunities.

Using the word2vec and doc2vec approaches, this research retrieves or recommends the best related patents or literature that match the target (input) documents. For the patent evolution graph using MFCA, the target documents are plotted in concentric circles by years and connected based on their conceptual similarity. The result depicts the evolution of key technologies, which also cross-reference to the cluster(s) and topic(s) of best related literature. This research helps enterprises easily discover existing technologies related to solar energy-oriented research. The proposed machine learning approaches and technology mining process flow are generic, which can be applied to reviews and analyses of other technology domains.

To summarize the contribution of this research, the readers can better understand the detailed technologies under each category by the construction of a machine learning program system and knowledge ontology. The proposed methodology framework can be referenced for further exploration of other technical aspects of solar technology. For instance, the recommendation system based on doc2vec can be applied to describe the novel research or patents for the solar materials used on the panel surfaces. The results help energy companies review and select technologies related to their key technical strengths and R&D interests. For future work, this research can be extended by further combination of machine learning or deep learning approaches to explore the application and development of other types of renewable energy technologies.

References

1. Kabir, E.; Kumar, P.; Kumar, S.; Adelodun, A.A.; Kim, K.H. Solar energy: Potential and future prospects. *Renew. Sustain. Energy Rev.* **2018**, *82*, 894–900. [CrossRef]
2. Li, X.; Xie, Q.; Jiang, J.; Zhou, Y.; Huang, L. Identifying and monitoring the development trends of emerging technologies using patent analysis and Twitter data mining: The case of perovskite solar cell technology. *Technol. Forecast. Soc. Chang.* **2018**. [CrossRef]
3. Trappey, A.J.C.; Trappey, C.V.; Wang, D.Y.; Li, S.J.; Ou, J.J. Evaluating renewable energy policies using hybrid clustering and analytic hierarchy process modeling. In Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Hsinchu, Taiwan, 21–23 May 2014; pp. 716–720.
4. Sahraei, N.; Looney, E.E.; Watson, S.M.; Peters, I.M.; Buonassisi, T. Adaptive power consumption improves the reliability of solar-powered devices for internet of things. *Appl. Energy* **2018**, *224*, 322–329. [CrossRef]
5. Thomson, R. *Engineering Sections, Derwent World Patents Index*®*, Thomson Reuters*, 6th ed.; 2016; Available online: http://clarivate.com/ (accessed on 30 August 2018).
6. Trappey, C.V.; Wu, H.-Y. An evaluation of the time-varying extended logistic, simple logistic, and Gompertz models for forecasting short product lifecycles. *Adv. Eng. Inform.* **2008**, *22*, 421–430. [CrossRef]
7. Yilanci, A.; Dincer, I.; Ozturk, H.K. A review on solar-hydrogen/fuel cell hybrid energy systems for stationary applications. *Prog. Energy Combust. Sci.* **2009**, *35*, 231–244. [CrossRef]
8. Cavallaro, F. Fuzzy TOPSIS approach for assessing thermal-energy storage in concentrated solar power (CSP) systems. *Appl. Energy* **2010**, *87*, 496–503. [CrossRef]
9. Ferrara, M.A.; Striano, V.; Coppola, G. Volume Holographic Optical Elements as Solar Concentrators: An Overview. *Appl. Sci.* **2019**, *9*, 193. [CrossRef]
10. Gentry, B. Holographic Optical Elements. HARLIE. NASA. Archived from the original on 15 February 2013. Retrieved 9 August 2018. Available online: harlie.gsfc.nasa.gov (accessed on 9 August 2018).
11. Philibert, C.; Frankl, P.; Tam, C.; Abdelilah, Y.; Bahar, H.; Mueller, S.; Waldron, M. *Technology Roadmap: Solar Thermal Electricity*; International Energy Agency: Paris, France, 2014.
12. Ogunmodimu, O.; Okoroigwe, E.C. Concentrating solar power technologies for solar thermal grid electricity in Nigeria: A review. *Renew. Sustain. Energy Rev.* **2018**, *90*, 104–119. [CrossRef]
13. Mehos, M.; Turchi, C.; Vidal, J.; Wagner, M.; Ma, Z.; Ho, C.; Kruizenga, A. *Concentrating Solar Power Gen3 Demonstration Roadmap*; National Renewable Energy Laboratory (NREL): Golden, CO, USA, 2017.
14. Ma, T.; Yang, H.; Lu, L. Feasibility study and economic analysis of pumped hydro storage and battery storage for a renewable energy powered island. *Energy Convers. Manag.* **2014**, *79*, 387–397. [CrossRef]

15. Philibert, C.; Frankl, P.; Tam, C.; Abdelilah, Y.; Bahar, H.; Marchais, Q.; Wiesner, H. *Technology Roadmap: Solar Photovoltaic Energy*; International Energy Agency: Paris, France, 2014.

16. Notton, G.; Voyant, C.; Fouilloy, A.; Duchaud, J.L.; Nivet, M.L. Some Applications of ANN to Solar Radiation Estimation and Forecasting for Energy Applications. *Appl. Sci.* **2019**, *9*, 209. [CrossRef]

17. Lee, W.W. *Thin-film Solar Battery Technology Development Trend Analysis*; Industry, Science and Technology International Strategy Center, Industrial Technology Research Institute (ITRI): Hsinchu, Taiwan, 2008.

18. Kim, J.Y.; Lee, K.; Coates, N.E.; Moses, D.; Nguyen, T.Q.; Dante, M.; Heeger, A.J. Efficient tandem polymer solar cells fabricated by all-solution processing. *Science* **2007**, *317*, 222–225. [CrossRef]

19. Selvaraj, J.; Rahim, N.A. Multilevel inverter for grid-connected PV system employing digital PI controller. *IEEE Trans. Ind. Electron.* **2009**, *56*, 149–158. [CrossRef]

20. Bassetti, M.C.; Consoli, D.; Manente, G.; Lazzaretto, A. Design and off-design models of a hybrid geothermal-solar power plant enhanced by a thermal storage. *Renew. Energy* **2018**, *128*, 460–472. [CrossRef]

21. Steinmann, W.D.; Tamme, R. Latent heat storage for solar steam systems. *J. Sol. Energy Eng.* **2008**, *130*, 011004. [CrossRef]

22. Zalba, B.; Marın, J.M.; Cabeza, L.F.; Mehling, H. Review on thermal energy storage with phase change: Materials, heat transfer analysis and applications. *Appl. Therm. Eng.* **2003**, *23*, 251–283. [CrossRef]

23. Kenisarin, M.; Mahkamov, K. Solar energy storage using phase change materials. *Renew. Sustain. Energy Rev.* **2007**, *11*, 1913–1965. [CrossRef]

24. Tamme, R.; Laing, D.; Steinmann, W.D. Advanced thermal energy storage technology for parabolic trough. *J. Sol. Energy Eng.* **2004**, *126*, 794–800. [CrossRef]

25. Wang, Y.; Zou, H.; Chen, X.; Zhang, F.; Chen, J. Adaptive Solar Power Forecasting based on Machine Learning Methods. *Appl. Sci.* **2018**, *8*, 2224. [CrossRef]

26. Appen, J.; Stetz, T.; Braun, M.; Schmiegel, A. Local voltage control strategies for PV storage systems in distribution grids. *IEEE Trans. Smart Grid* **2014**, *5*, 1002–1009. [CrossRef]

27. Riffonneau, Y.; Bacha, S.; Barruel, F.; Ploix, S. Optimal power flow management for grid connected PV systems with batteries. *IEEE Trans. Sustain. Energy* **2011**, *2*, 309–320. [CrossRef]

28. Kim, S.K.; Jeon, J.H.; Cho, C.H.; Ahn, J.B.; Kwon, S.H. Dynamic modeling and control of a grid-connected hybrid generation system with versatile power transfer. *IEEE Trans. Ind. Electron.* **2008**, *55*, 1677–1688. [CrossRef]

29. Teng, J.H.; Luan, S.W.; Lee, D.J.; Huang, Y.Q. Optimal charging/discharging scheduling of battery storage systems for distribution systems interconnected with sizeable PV generation systems. *IEEE Trans. Power Syst* **2013**, *28*, 1425–1433. [CrossRef]

30. Zahedi, A. Maximizing solar PV energy penetration using energy storage technology. *Renew. Sustain. Energy Rev.* **2011**, *15*, 866–870. [CrossRef]

31. Louwen, A.; Van Sark, W.; Schropp, R.; Faaij, A. A cost roadmap for silicon heterojunction solar cells. *Sol. Energy Mater. Sol. Cells* **2016**, *147*, 295–314. [CrossRef]

32. Glavin, M.E.; Chan, P.K.; Armstrong, S.; Hurley, W.G. A stand-alone photovoltaic supercapacitor battery hybrid energy storage system. In Proceedings of the 2008 13th Power Electronics and Motion Control Conference (EPE-PEMC), Poznan, Poland, 1–3 September 2008; pp. 1688–1695.

33. Podjaski, F.; Kröger, J.; Lotsch, B.V. Toward an Aqueous Solar Battery: Direct Electrochemical Storage of Solar Energy in Carbon Nitrides. *Adv. Mater.* **2018**, *30*, 1705477. [CrossRef] [PubMed]

34. Wu, C.Y.; Mathews, J.A. Knowledge flows in the solar photovoltaic industry: Insights from patenting by Taiwan, Korea, and China. *Res. Policy* **2012**, *41*, 524–540. [CrossRef]

35. Chang, S.B.; Lai, K.K.; Chang, S.M. Exploring technology diffusion and classification of business methods: Using the patent citation network. *Technol. Forecast. Soc. Chang.* **2009**, *76*, 107–117. [CrossRef]

36. Trappey, A.J.C.; Trappey, C.V.; Wang, Y.C.; Ou, J.R.; Li, S.J. An integrated self-organizing map and analytic hierarchy process modeling approach for evaluating renewable energy policies. *Int. J. Electron. Bus. Manag.* **2015**, *13*, 3–14.

37. Zhang, Y.; Shang, L.; Huang, L.; Porter, A.L.; Zhang, G.; Lu, J.; Zhu, D. A hybrid similarity measure method for patent portfolio analysis. *J. Informetr.* **2016**, *10*, 1108–1130. [CrossRef]

38. Sampaio, P.G.V.; González, M.O.A.; de Vasconcelos, R.M.; dos Santos, M.A.T.; de Toledo, J.C.; Pereira, J.P.P. Photovoltaic technologies: Mapping from patent analysis. *Renew. Sustain. Energy Rev.* **2018**, *93*, 215–224. [CrossRef]

39. Trappey, A.J.C.; Trappey, C.V.; Fan, C.-Y.; Hsu, A.P.T.; Li, X.K.; Lee, I.J.Y. IoT patent roadmap for smart logistic service provision in the context of Industry 4.0. *J. Chin. Inst. Eng.* **2017**, *40*, 593–602. [CrossRef]

40. Trappey, A.J.C.; Trappey, C.V.; Chang, A.C.; Li, X.K. Deriving competitive foresight using an ontology-based patent roadmap and valuation analysis. *Int. J. Semant. Web Inf. Syst.* **2018**. [CrossRef]

41. Sato, Y.; Iwayama, M. Interactive constrained clustering for patent document set. In Proceedings of the 2nd International Workshop on Patent Information Retrieval, Hong Kong, China, 6 November 2009; pp. 17–20.

42. Trappey, A.J.C.; Trappey, C.V.; Chung, L.S. IP portfolios and evolution of biomedical additive manufacturing applications. *Scientometrics* **2017**, *111*, 139–157. [CrossRef]

43. Kim, G.; Bae, J. A novel approach to forecast promising technology through patent analysis. *Technol. Forecast. Soc. Chang.* **2017**, *117*, 228–237. [CrossRef]

44. Bengio, Y.; Ducharme, R.; Vincent, P.; Jauvin, C. A neural probabilistic language model. *J. Mach. Learn. Res.* **2003**, *3*, 1137–1155.

45. Santos, C.D.; Tan, M.; Xiang, B.; Zhou, B. Attentive pooling networks. *arXiv*, **2016**; arXiv:1602.03609.

46. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv*, 2013; arXiv:1301.3781.

47. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 3111–3119.

48. Tang, D.; Wei, F.; Yang, N.; Zhou, M.; Liu, T.; Qin, B. Learning sentiment-specific word embedding for twitter sentiment classification. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MD, USA, 22–27 June 2014; Volume 1, pp. 1555–1565.

49. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Society. Ser. C (Appl. Stat.)* **1979**, *28*, 100–108. [CrossRef]

50. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]

51. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

52. Lee, C.; Jeon, J.; Park, Y. Monitoring trends of technological changes based on the dynamic patent lattice: A modified formal concept analysis approach. *Technol. Forecast. Soc. Chang.* **2011**, *78*, 690–702. [CrossRef]

53. Trappey, A.J.C.; Trappey, C.V.; Lee, K.L.C. Tracing the evolution of biomedical 3D printing technology using ontology based patent concept analysis. *Technol. Anal. Strateg. Manag.* **2017**, *29*, 339–352. [CrossRef]

54. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

55. Girolami, M.; Kabán, A. On an equivalence between PLSI and LDA. In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, ON, Canada, 28 July–1 August 2003; pp. 433–434.

56. Zou, C.H. Using Non-Supervised Machine Learning Approach to Generate Knowledge Ontology for Patent (Advisor: A.J.C. Trappey). Master's Thesis, Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan, 2018.

57. Wang, X.; McCallum, A. Topics over time: A non-Markov continuous-time model of topical trends. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, 20–23 August 2006; pp. 424–433.

58. Wilson, A.T.; Robinson, D.G. *Tracking Topic Birth and Death in LDA*; Sandia National Laboratories: Carlsbad, NM, USA, 2011.

59. Le, Q.; Mikolov, T. Distributed representations of sentences and documents. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014; pp. 1188–1196.

60. Lau, J.H.; Baldwin, T. An empirical evaluation of doc2vec with practical insights into document embedding generation. *arXiv*, 2016; arXiv:1607.05368.

61. Li, Z.S.; Zhang, G.Q.; Li, D.M.; Zhou, J.; Li, L.J.; Li, L.X. Application and development of solar energy in building industry and its prospects in China. *Energy Policy* **2007**, *35*, 4121–4127. [CrossRef]

62. Zhang, S.; He, Y. Analysis on the development and policy of solar PV power in China. *Renew. Sustain. Energy Rev.* **2013**, *21*, 393–401. [CrossRef]

63. Zhao, Z.Y.; Zhang, S.Y.; Hubbard, B.; Yao, X. The emergence of the solar photovoltaic power industry in China. *Renew. Sustain. Energy Rev.* **2013**, *21*, 229–236. [CrossRef]

64. Yang, D.; Wang, X.; Kang, J. SWOT Analysis of the Development of Green Energy Industry in China: Taking solar energy industry as an example. In Proceedings of the 2018 2nd International Conference on Green Energy and Applications (ICGEA), Singapore, 24–26 March 2018; pp. 103–107.
65. Csiro.au. Australian Solar Tech to Help China Reach Clean Energy Targets—CSIRO. 2016. Available online: https://www.csiro.au/en/News/News-releases/2016/ (accessed on 29 October 2018).
66. EPO and USPTO. CPC Cooperative Patent Classification System Annual Report 2016. 2017. Available online: https://www.cooperativepatentclassification.org/publications/AnnualReports (accessed on 30 October 2018).
67. Campello, R.J.; Hruschka, E.R. A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets Syst.* **2006**, *157*, 2858–2875. [CrossRef]
68. Lalouni, S.; Rekioua, D.; Rekioua, T.; Matagne, E. Fuzzy logic control of stand-alone photovoltaic system with battery storage. *J. Power Sources* **2009**, *193*, 899–907. [CrossRef]
69. Maeda, T.; Ito, H.; Hasegawa, Y.; Zhou, Z.; Ishida, M. Study on control method of the stand-alone direct-coupling photovoltaic–Water electrolyzer. *Int. J. Hydrogen Energy* **2012**, *37*, 4819–4828. [CrossRef]
70. Sridhar, H.; Meera, K.S. Study of grid connected solar photovoltaic system using real time digital simulator. In Proceedings of the 2014 International Conference on Advances in Electronics Computers and Communications, Bangalore, India, 10–11 October 2014; pp. 1–6.
71. Li, Y.; Gao, W.; Ruan, Y. Performance investigation of grid-connected residential PV-battery system focusing on enhancing self-consumption and peak shaving in Kyushu, Japan. *Renew. Energy* **2018**, *127*, 514–523. [CrossRef]
72. Petrollese, M.; Cau, G.; Cocco, D. Use of weather forecast for increasing the self-consumption rate of home solar systems: An Italian case study. *Appl. Energy* **2018**, *212*, 746–758. [CrossRef]