Solar Energy 168 (2018) 60–101

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

Review

History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining

Dazhi Yang[a,*], Jan Kleissl[b], Christian A. Gueymard[c], Hugo T.C. Pedro[b], Carlos F.M. Coimbra[b]

[a] Singapore Institute of Manufacturing Technology, Agency for Science, Technology and Research (A*STAR), Singapore
[b] Center for Renewable Resources and Integration, Center for Energy Research, Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA, USA
[c] Solar Consulting Services, Colebrook, NH, USA

ARTICLE INFO

Keywords:
Text mining
Solar forecasting
Review
Photovoltaics

ABSTRACT

Text mining is an emerging topic that advances the review of academic literature. This paper presents a preliminary study on how to review solar irradiance and photovoltaic (PV) power forecasting (both topics combined as "solar forecasting" for short) using text mining, which serves as the first part of a forthcoming series of text mining applications in solar forecasting. This study contains three main contributions: (1) establishing the technological infrastructure (authors, journals & conferences, publications, and organizations) of solar forecasting via the top 1000 papers returned by a Google Scholar search; (2) consolidating the frequently-used abbreviations in solar forecasting by mining the full texts of 249 ScienceDirect publications; and (3) identifying key innovations in recent advances in solar forecasting (e.g., shadow camera, forecast reconciliation). As most of the steps involved in the above analysis are automated via an application programming interface, the presented method can be transferred to other solar engineering topics, or any other scientific domain, by means of changing the search word. The authors acknowledge that text mining, at its present stage, serves as a complement to, but not a replacement of, conventional review papers.

1. Introduction: Towards a new reviewing paradigm

The history of solar irradiance forecasting can be said to have started in the late 19th- and early 20th-century when numerical weather prediction (NWP) began. It is remarkable how pyrheliometers—the primary instrument to measure direct normal irradiance (DNI), still in common use today as a reference instrument—had already been developed and employed as a forecasting tool by then (Marvin and Kimball, 1926). However, it was not until the advent of mainframe computers and simulations that computation time was reduced to less than the forecast horizon. Today, solar irradiance forecasting and photovoltaic (PV) power forecasting (both referred to as "solar forecasting" in what follows) receive unprecedented attention from various scientific communities. This is because of the importance of forecasting the variability of solar and wind power for their grid integration, which constitutes a major challenge to a successful transformation of the conventional fossil fuel-based energy sector into a 100% renewable one. To give perspective, Google Scholar searches for "solar irradiance forecasting" and "PV power forecasting" return 15,700 and 6340 results for the year 2016 alone.

Considering this abundant literature on solar forecasting, many review papers have been written in recent years. The primary purpose of review papers is to familiarize students and researchers with a relatively new topic and facilitate the use of a number of new and powerful tools. A list of recent review papers on solar forecasting is shown in Table 1. Reviews compile, summarize, critique, and synthesize the available information on a subject (Suter, 2013). Despite the obvious benefits of reviews, they nevertheless have three main drawbacks:

1. The number of references considered in each review is still small relatively to the total available publications on the subject.
2. It is often unclear what methods review authors applied to search the literature, identify publications, extract information, and generate insights (Suter, 2013).
3. Since each review is only read by a handful of scientists (authors, reviewers, and possibly journal editors) before its publication, the content may be biased and/or subjective.

Analogically speaking, review papers behave like local optima in an optimization problem, while actually the global solution is sought. As in optimization, there are ways to escape from the local optima, but it often takes years of experience before a reader can critically interpret

* Corresponding author.
E-mail addresses: yangdz@simtech.a-star.edu.sg, yangdazhi.nus@gmail.com (D. Yang).

https://doi.org/10.1016/j.solener.2017.11.023
Received 23 October 2017; Accepted 9 November 2017
Available online 03 February 2018

Table 1
Review papers on solar forecasting. The number of citations is taken from Google Scholar at the time of manuscript submission.

Review	Title	Journal	#Refs.	#Pages	#Citations
Barbieri et al. (2017)	Very short-term photovoltaic power forecasting with cloud modeling: A review	Renewable & Sustainable Energy Reviews	94	22	7
van der Meer et al. (2017)	Review on probabilistic forecasting of photovoltaic power production and electricity consumption	Renewable & Sustainable Energy Reviews	140	29	1
Voyant et al. (2017c)	Machine learning methods for solar radiation forecasting: A review	Renewable Energy	105	14	18
André et al. (2016)	Predictive spatio-temporal model for spatially sparse global solar radiation data	Energy	29	10	2
Antonanzas et al. (2016)	Review of photovoltaic power forecasting	Solar Energy	151	34	44
Raza et al. (2016)	On recent advances in PV output power forecast	Solar Energy	123	20	28
Kashyap et al. (2015)	Solar radiation forecasting with multiple parameters neural networks	Renewable & Sustainable Energy Reviews	65	11	19
Qazi et al. (2015)	The artificial neural network for solar radiation prediction and designing solar systems: A systematic literature review	Journal of Cleaner Production	54	12	38
Ren et al. (2015)	Ensemble methods for wind and solar power forecasting–A state-of-the-art review	Renewable & Sustainable Energy Reviews	55	10	35
Wan et al. (2015)	Photovoltaic and solar power forecasting for smart grid energy management	CSEE Journal of Power and Energy Systems	82	9	54
Law et al. (2014)	Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting—A review	Solar Energy	165	21	46
Diagne et al. (2013)	Review of solar irradiance forecasting methods and a proposition for small-scale insular grids	Renewable & Sustainable Energy Reviews	50	12	170
Inman et al. (2013)	Solar forecasting methods for renewable energy integration	Progress in Energy and Combustion Science	293	42	265
Kleissl (2013)	Solar energy forecasting and resource assessment	Book	–	–	124

and synthesize these reviews to derive an objective assessment of the state-of-the-art.

In this paper, an assistive method—text mining—is primarily considered as a potential replacement for, or addition to, conventional literature reviews. Since text mining is an automated process of deriving information from text, it is not limited by the amount of input data, thus providing a remedy for the first aforementioned drawback. In each of the sections below, the methods used to collect, group and analyze publications are elaborated with justification. Such elaboration is believed to improve the transparency of the present results. In turn, this should provide greater assurance of the quality of the review process, and hence close the second gap mentioned earlier. Lastly, to reduce the unavoidable biases in any review, a group of domain experts—five associate/subject editors of Solar Energy[1] on the subject of solar resources & energy meteorology—are interpreting the text mining results and co-writing this paper. Furthermore, Google Scholar search results are herein considered. Since Google Scholar ranks a publication based on (i) where it was published, (ii) who it was written by, as well as (iii) the count and recency of its citations,[2] the search results essentially reflect the prevailing confidence in popularity and publication quality (as suggested by crowdsourcing). Based on this assessment, the combination of Google Scholar data and supervision from domain experts is expected to mitigate the third drawback.

2. Introducing a new toolkit for literature review

2.1. Working with Google Scholar data

Google Scholar is one of the most important free academic search engines (Ortega and Aguillo, 2014), and often provides a more comprehensive coverage of resources in various scientific disciplines as compared to Web of Science or Scopus (Harzing, 2013). By mining and analyzing the environment of a large number of publications (e.g., titles, authors, abstracts, citations, and Google Scholar profiles), valuable information and insights on an academic field can be obtained.

Much research has been done in various fields using Google Scholar data. For instance, Chen et al. (2017) collected more than 400,000 Google Scholar profiles across various disciplines and analyzed the demography of these scholars. A co-authorship network was built to study the collaboration among authors and its resulting link to citation metrics. It was found that the ranking of a page is strongly correlated with the h-index.[3] From a different perspective, Shariff et al. (2013) utilized Google Scholar to help physicians to retrieve clinical evidence and to guide the care of their patients. In the field of knowledge management, Google Scholar was used to discover growing, stable and declining research trends (Serenko and Dumay, 2015). Google Scholar data has also been used in solar engineering. Yang (2016) compared citations of 15 papers on irradiance transposition modeling through years, and filtered out the less-cited models for that study.

Despite all its potentials and benefits, Google Scholar has its downside: the lack of transparency is the main reservation of bibliometricians to use it as a research evaluation database (López-Cózar et al., 2014). Because Google Scholar automatically retrieves, indexes, and stores any form of text-based scientific material (paper, presentation slides, or even personal memo) uploaded by an author without much quality control, information such as citation counts may be inflated. López-Cózar et al. (2014) performed an experiment by uploading several false papers with abundant citations of publications from their

[1] For clarity, journal and author names (only when not in a citation) are noted with Small Caps.

[2] https://scholar.google.com/intl/en/scholar/about.html.

[3] The h-index was proposed by Hirsch (2005) to characterize the scientific output of a researcher. It is defined as the number of papers with citation number ⩾h. Google Scholar separately calculates the h-index of all scientists based on their whole career and on the latest 5-year period.