

ebadi_2020_application_of_machine_learning_techniques_
to_assess_the_trends_and_alignment_of_the_funded_resea
rch_output

Year

2020

Author(s)

Ashkan Ebadi and Stephane Tremblay and Cyril Goutte and Andrea Schiffauerova

Title

Application of machine learning techniques to assess the trends and alignment of the
funded research output

Venue

Journal of Informetrics

Topic labeling

Manual

Focus

Secondary

Type of contribution

Established approach

Underlying technique

Manual labeling

Topic labeling parameters

\

Label generation

We asked three domain experts to manually examine the extensive set of keywords for each topic and assign a representative label to the topic.

The labeled extracted topics are: 1) nanoscience, 2) environment, climate, and oceanology, 3) physics and mechanics, 4) mathematics, 5) analytics, 6) genomics, 7) cognition and physiology, and 8) healthcare and disease.

Motivation

"Additionally, in our study, we were interested in finding the main research topics, i.e. more generic and inclusive topics, preferably with a simple and interpretable name rather than a phrase."

Topic modeling

Structural Topic Model (STM)

Topic modeling parameters

Nr of topics = {8, 10}

Nr. of topics

8

Label

Manually assigned single or multi-word labels representing the main research areas at an abstract level.

Label selection

First, the experts separately analyzed the list of keywords and identified the topic labels. Next, all three experts shared their labels and tried to convince other researchers if there was an incompatibility.

Finally, the label with the highest vote on a topic was chosen as the topic's label. We used more than one expert to reduce the subjectivity effect.

Label quality evaluation

\

Assessors

Three domain experts

Domain

Paper: Funded research

Dataset: Funded research

Problem statement

Research and development activities are regarded as one of the most influencing factors of the future of a country.

Scientific publications are considered as one of the main outputs of research investment.

A detailed analysis is required to monitor the research trends and assess whether they are in line with the top research priorities of the country.

Using structural topic models, in this paper we evaluated the trends in academic research performed by federally funded Canadian researchers during the time-frame of 2000–2018, covering more than 140,000 research publications.

Corpus

Origin: Scopus

Nr. of documents: 140,966

Details:

- Focus on researchers who were funded by the main federal funding agency in Canada, i.e. NSERC, from 2000 to 2018.
- Publications that acknowledged NSERC as a source of funding from the Elsevier's Scopus

Document

Metadata related to a single publication: article id, title, abstract, year of publications, and coauthors

Pre-processing

- Publications with no abstract available are filtered out
- Special characters and parsing affiliations are removed
- Author disambiguation is performed to link the funding and publications datasets. For this purpose, full author names and affiliations provided by the NSERC funding data as well as current and past affiliations of authors provided by Scopus are used. Next, similarity measure is defined based on author names, affiliations, and research areas. The measure was then used by a machine learning algorithm to disambiguate authors and link the funding and publication datasets.
- Publications titles and abstracts are merged

Text processing steps:

- Text is transformed to lower case
- Removed stop words
- Numbers, html tags, punctuation marks, and words with less than three characters are eliminated.
- Tokenized the textual data
- Created a document-term frequency matrix.

```
@article{ebadi_2020_application_of_machine_learning_techniques_to_assess_the_trends_and_alignment_of_the_funded_research_output,
```

```
    abstract = {Research and development activities are regarded as one of the most influencing factors of the future of a country. Large investments in research can yield a tremendous outcome in terms of a country's overall wealth and strength. However, public financial resources of countries are often limited which calls for a wise and targeted investment. Scientific publications are considered as one of the main outputs of research investment. Although the general trend of scientific publications is increasing, a detailed analysis is required to monitor the research trends and assess whether they are in line with the top research priorities of the country. Such focused monitoring can
```

shed light on scientific activities evolution as well as the formation of new research areas, thus helping governments to adjust priorities, if required. But monitoring the output of the funded research manually is not only very expensive and difficult, it is also subjective. Using structural topic models, in this paper we evaluated the trends in academic research performed by federally funded Canadian researchers during the time-frame of 2000--2018, covering more than 140,000 research publications. The proposed approach makes it possible to objectively and systematically monitor research projects, or any other set of documents related to research activities such as funding proposals, at large-scale. Our results confirm the accordance between the performed federally funded research projects and the top research priorities of Canada.},

author = {Ashkan Ebadi and Stéphane Tremblay and Cyril Goutte and Andrea Schiffauerova},

date-added = {2023-03-13 19:44:45 +0100},

date-modified = {2023-03-13 19:44:45 +0100},

doi = {https://doi.org/10.1016/j.joi.2020.101018},

issn = {1751-1577},

journal = {Journal of Informetrics},

keywords = {Text mining, Topic modeling, Machine learning, Funded research, Publications, Government research priorities, Canada},

number = {2},

pages = {101018},

title = {Application of machine learning techniques to assess the trends and alignment of the funded research output},

url = {https://www.sciencedirect.com/science/article/pii/S1751157718301901},

volume = {14},

year = {2020}}