Disentangling Research Trends

Neural Network Vector Embeddings for Detecting R&D Change

Abstract

We propose a novel computational framework to proactively monitor research trends across domains. It distinguishes genuine changes from natural growth and quantifies shifts in research topics and collaborative networks. Employing advanced vector embedding methods, the system provides valuable insights for researchers, policymakers, and research managers, supporting data-driven decision-making for strategic research directions.

Introduction

Problem Statement: Research trends reflect complex dynamics shaped by scientific breakthroughs, societal challenges, and shifting funding priorities. Traditional trend analyses often struggle to separate natural growth from meaningful shifts in direction or semantic changes in underlying topics of research interest.

Limitations of currently applied methods:

- Expert-defined taxonomies limit scalability and cross-domain analysis.
- Surface-level metrics may not reveal deep transformations within evolving research themes.
- Quantifying changes in the collaborative networks that underpin shifts in research remains challenging.

Methods

Overall Approach: We employ a multi-faceted vector embedding approach to quantify dynamic patterns in research trends using multiple publication-based data sources:

A) Trend Normalization: Publication counts per year are embedded as vectors and normalized to facilitate comparison across trends. This normalization isolates meaningful deviations from natural growth patterns.

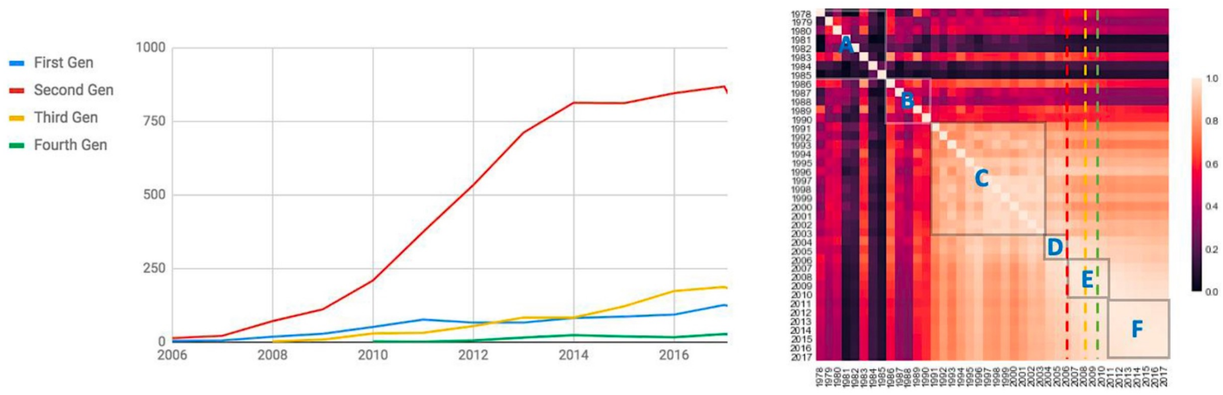
B) Semantic Analysis: A state-of-the-art foundational LLM generates embeddings for individual research documents. Cosine similarity quantifies the degree of semantic change between these embeddings over time and per scoped segment (e.g. , topic, affiliation, categories, etc.).

C) Collaborative Network Analysis: Co-authorship networks (authors/affiliations) are analyzed using the RV-Coefficient to measure structural changes across time. This reveals potential shifts in collaboration patterns.

Preliminary Results

In Pedro Parraguez, et al., Technological Forecasting & Social Change (https://doi.org/10.1016/j.techfore.2019.119803) we demonstrated the impact of a previous version of this approach in the bioenergy domain. In that study, our method successfully:

- Unveiled pre-trends, identifying biofuel generations before their formal definition in the literature.
- Detected potential declines of once-promising research avenues.
- Demonstrated the potential for cross-domain adaptation (energy to medicine, etc.).



Left panel, number of records containing references to each of the four biofuel generations. Right panel, year-to-year similarity matrix between 2006 and 2018, with colored vertical lines for when the first mentions to each generation occurred and the previously described clusters A to F as reference points. Image from Pedro Parraguez, et al., Technological Forecasting & Social Change (https://doi.org/10.1016/j.techfore.2019.119803)

Relevance and Impact

New Insights: This approach provides data-driven insights into the evolving nature of research, facilitating better strategic decision-making for:

- Researchers: Identify emerging research areas and potential blind spots.
- Policymakers and Funders: Focus support on promising new directions or understand reasons for the decline of once-popular themes.
- Research managers: Optimize resource allocation and facilitate impactful collaborations.

Conference/Portal Alignment: This work directly supports the themes of the Research Portal Denmark launch conference by:

- Demonstrating the power of open data analysis in research evaluation.
- Offering methods to enhance research monitoring and support the portal's goals.

Learn more at

https://pattr.nz/nora-poster

Pedro Parraguez & Nelson Guaman
pedro@pattrnz.net, www.pattrnz.net

Poster Session, Launch of Danish Research Portal Denmark, 21 March, 2024

pattrnz www.pattrnz.net

Powered by Data verz