Research Trends in Data Augmentation Algorithms and Applications
A Systematic Literature Review

Joao Fonseca[1*], Fernando Bacao[1]

[1]NOVA Information Management School, Universidade Nova de Lisboa

*Corresponding Author

Postal Address: NOVA Information Management School, Campus de Campolide, 1070–312 Lisboa, Portugal

Telephone: +351 21 382 8610

1 Introduction

Introduction goes here.

2 Theory

This part is going to be the very last thing to be done, since it doesn't seem absolutely necessary.

3 Methodology

- General overview of the steps taken to produce the visualizations and analyses.
- Use diagrams and visualizations.
- Discuss the PRISMA methodology.
- Discuss embedding techniques used.

3.1 Keyword Identification and Search

- Goal: Avoid Preconceived biases on relevant keywords

- Primary keyword: “Data Augmentation”
- Append to query OR rules using top n keywords not domain specific
- Refine results based on research domain
- Discuss topics included (computer science, mathematics, engineering etc)

Started with a single keyword: “Data Augmentation”, 4618 results.

Limited results to documents written in English, 4517 results.

Only include articles and conference papers, 4443 results.

Consider papers published in the past 15 years, 4281 results.

Final query:

KEY (”data augmentation”) AND (LIMIT-TO (LANGUAGE , ”English”)) AND (LIMIT-TO (DOCTYPE , ”cp”) OR LIMIT-TO (DOCTYPE , ”ar”)) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012) OR LIMIT-TO (PUBYEAR , 2011) OR LIMIT-TO (PUBYEAR , 2010) OR LIMIT-TO (PUBYEAR , 2009) OR LIMIT-TO (PUBYEAR , 2008) OR LIMIT-TO (PUBYEAR , 2007) OR LIMIT-TO (PUBYEAR , 2006))

Due to the limitations in the Scopus data export (maximum 2000 documents per export), the data was extracted in four parts: 2021, 2020, 2019 and 2018 – 2006.

3.2 Repositories

- Repositories: Scopus

3.3 Bibliometric Analysis

- Discuss data preprocessing done.
- Concatenate data, extract features (if possible), data cleaning etc

3.4 Data Analysis Software for Bibliometric Research

- Python
- Streamlit

- Text Mining and Embedding libraries
- Data Visualization libraries
- VOSviewer?

4 Results

Results go here.

4.1 PRISMA Flow Diagram

- Create a flowchart with the data cleaning process and describe it.

4.2 Terms Frequency

4.3 Topics Discovered

4.3.1 Main Journals

4.3.2 Main Conference Proceedings

4.4 Author Co-occurrence Analysis

Not sure whether to keep this one.

4.5 Title and Abstract Text Occurrence Analysis

4.6 Most Cited Publications

4.7 Application and Method Analysis

5 Discussion

Discussion goes here.