



Covid 19 Literature Exploring

Unsupervised/NLP Project

Jhih-Rou Huang

Introduction

The literature of Covid 19 has increased rapidly

Problem

- Difficult to keep up with the newest information
- Waste too much time searching for articles





Goal

Explore the Covid Literatures and build a topic model let the researchers find the papers easily and faster

Data

Resource: COVID-19 Open Research Dataset Challenge (CORD-19)

Papers in 2021: 599,616 articles

Use the Abstract to do the topic modeling

Tools



Langdetect: use English articles only (94,984 papers)



spaCy, pandas, numpy



Gensim: topic models for the LDA



NLTK: stop word removal, tokenization



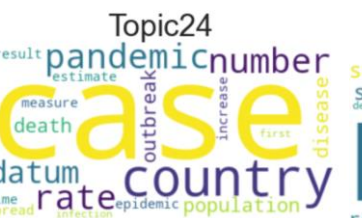
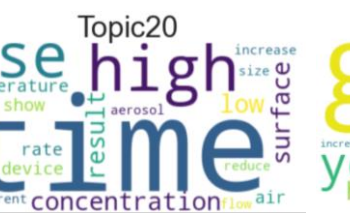
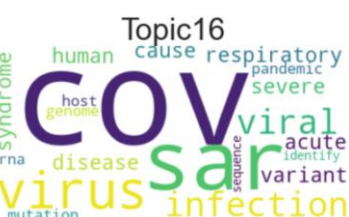
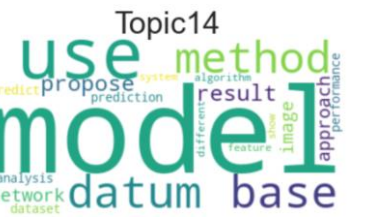
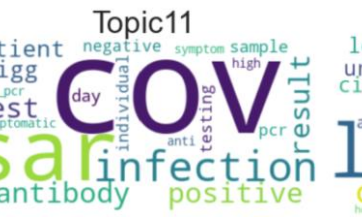
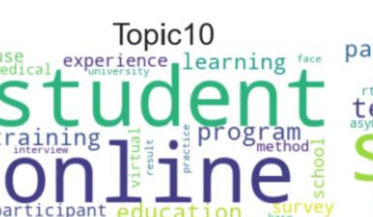
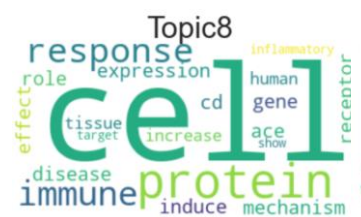
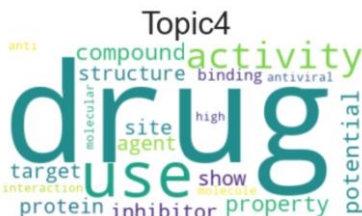
WorldClouds: visualization

The best score is 0.39 with 26 topics

Coherence Scores



World Cloud



Keywords

- Topic 1 Cancer Tumor
- Topic 2 Risk
- Topic 3 Physical activity
- Topic 4 Drug Use
- Topic 5 Financial Economic
- Topic 6 Research system
- Topic 7 Vaccine
- Topic 8 Cell Protein
- Topic 9 Public Health
- Topic 10 Student Online
- Topic 11 and 16 Cov infecti
- Topic 12 Lockdown
- Topic 13 Care Patient

- Topic 14 Use Model
- Topic 15 Antibiotic Bacterial
- Topic 17 Drug Treatment

- Does insurance demand react to economic policy uncertainty and geopolitical risk?
- Implications of the 2020 Global Economic Crisis for Sustainable Development and Financial Mechanisms for Their Management
- Short-term financial constraints and SMEs' investment decision: evidence from the working capital channel

- Topic 21 Group age
- Topic 22 Case Symptom
- Topic 23 Trial Study

- Impact of the COVID-19 lockdown on roadside traffic-related air pollution in Shanghai, China
- Have any effect of COVID-19 lockdown on environmental sustainability? A study from most polluted metropolitan area of India
- Air pollution and COVID-19 lockdown in a large South American city: Santiago Metropolitan Area, Chile



To Do List

- Comparing different models
- Building a website for the covid articles searching

Thank you!