UNIVERSITY OF GENOA
DEPARTMENT OF INFORMATICS, BIOENGINEERING, ROBOTICS AND SYSTEMS ENGINEERING
MASTER'S PROGRAM IN BIOENGINEERING

Thesis Project Form

Title (tentative):   Capturing the landscape of neurological research through machine learning and data visualization

Thesis advisor(s):   Barla Annalisa

E-mail: Annalisa.Barla@unige.it

Address:

Phone: (+39) 010 353 6602

Description

Motivation and application domain

As we move further into the 21st century, the pace of scientific progress is accelerating faster than ever before. As a scientist, it can be a challenge to stay current with the latest advances, especially in the life science and biomedical fields that draw on multiple disciplines. With so much new research being published every day, it can be difficult to filter through the noise and stay focused on the most relevant information.

General objectives and main activities

We will collect a vast amount of scientific literature on neurology, including peer-reviewed journal articles, conference papers, and other relevant publications, using various databases such as PubMed. We will then preprocess the collected data by removing irrelevant information, such as duplicate articles, and cleaning up the data, such as removing special characters and stopwords and storing the clean data in a structured data format. Next, we will apply natural language processing techniques to analyze the text data. We will also use deep learning techniques, such as convolutional neural networks and recurrent neural networks, to extract features and patterns from the text data. Using topic modeling techniques, we will group the articles into relevant topics and identify the most significant keywords for each topic. We will also use network analysis to identify the relationships between the topics and the articles. Finally, we will use data visualization techniques to visually represent the results of our analysis.

Training Objectives (technical/analytical tools, experimental methodologies)

The candidate is expected to learn:
- how to handle large scale data and prepare it for a machine learning analysis
- collect data (i.e. scientific corpus of publications from public repositories)
- apply NLP techniques, including the most advanced ones base on deep learning, to obtain a numerical representation of text that also keeps into account its semantics
- use topic modeling to identify and stratify the data at hand
- present the results of the analysis with effective and appropriate data visualization techniques

Place(s) where the thesis work will be carried out: UniGe | MaLGa - Machine Learning Genoa Center

Additional information

Pre-requisite abilities/skills: familiarity with machine learning methods and their application to natural language processing. Fundamentals of data visualization are a plus

Maximum number of students: 1