

NOTES

- Each section must be 9-12 sentences
- Text orientation on poster once complete
- Add images and/or graphs
- **Final Version Due: Saturday, March 21st**

Using natural language parsing and artificial intelligence techniques to initiate a phase change to biological knowledge

1 Abstract

The Biological Navigation and Visualization Tool (BioNaVisT) is an artificial intelligence system that can read, interpret, and organize information from textbooks and research papers. The project is currently being developed to tackle the rapid growth of biological discoveries and literature. Today as individuals, it is becoming infeasible to keep up with the amount of information that is published at an ever increasing rate. The BioNaVisT that is being developed may allow us to interact with very large sets of data. The system draws together vast amounts of information disseminated by researchers, and creates a visual user-interface to show concentrations of related data using a natural language processing program and a customized open-source graphing program. The system will utilize artificial intelligence to establish connections between different blocks of information, and allow the researcher to see clusters of information and display the relative importance of each node in the system. This system could help shed light on very complex biological process such as the human aging process.

2 Introduction

- Motive: No complete model on human aging. Several different theories, lots of controversy.
- Problems:
 - **Difficulty in Quantifying Aging:** Complex subject, hard to investigate experimentally. Too many systems involved, dynamic biological interactions, too much variation between individuals, too much variation over time.
 - **"Big Data":** Large number of abstracts on Pubmed. New articles on aging published daily. Data stored in inaccessible formats, as semantically unstructured text or images.
- Goal:
 - **Efficiency:** Automate text curation process and extract relations from Pubmed article abstract
 - **Reliability:** Improve relation extraction with machine learning on biological aging-related rules
 - **Functionality:** Provide passage for concept modelling and hypothesis generation

3 Background

- **Quantifying Aging Processes:** Quantify interactions in human body (growth in biological aging, is a function of time). Interested in biological events to build our model. CREA is modelling aging from a systems biology standpoint. Begin by looking at interactions between biological objects, how they give rise to a model for function and behaviour of a particular system, and how the sum of their parts give rise to function and behavior of all systems as a whole.
- **Algorithmic Approach:** NLP to extract unstructured text from PubMed article abstracts. Store relations in structured digital format. Map data using graphs to visualize connections between unlinked data.
- **Expected Advantages:** Improve pattern discovery. Provide knowledge base accessibility, permit data manipulation (i.e applying functions that predict biological change in a process)

4 Methods

A: Knowledge Extraction and NLP

- Use a NLP engine to process and extract relations of PubMed article abstracts
- Relations are a 3-tuple of (subject, predicate, object)
- Filter these relations
- Plot them on a graph
- Analyse the graph to find patterns, or possible optimizations to the algorithm
- Repeat

B: Bootstrapping and Semi-Supervised Learning

- Iteration: Use patterns to get more instances and patterns
- Example 1: Target relation: chemical interaction
 - 1. Seed tuple: [named_entity=progesterone, named_entity=dopamine]
 - * i.e. progesterone decreases dopamine
 - * i.e. progesterone regulates dopamine
 - 2. Seed tuple: [named_entity=dopamine, named_entity=prolactin]
 - * i.e. dopamine regulates prolactin
 - * i.e. dopamine inhibits prolactin
- Example 2: Target relation: excitation event
 - 1. Seed tuple: [named_entity=dopamine, verb=excite]
 - * i.e. cannabinoid excites dopamine
 - * i.e. opiod excites dopamine
- Example 3: Target relation: inhibitory event
 - 1. Seed tuple: [named_entity=dopamine, verb=inhibit]
 - * i.e. dopamine inhibits prolactin
 - * i.e. L-dopa inhibits dopamine
 - * i.e. antipsychotic medication inhibits dopamine
 - * i.e. dopamine inhibits activated T cells

C: Entity Classification

- Biological Events
 - Pathways (i.e. Wnt signaling)
 - Biomolecular Interaction (i.e. glucose increase insulin release)
 - Biological Phenomena (i.e. Apoptosis - cell death)
 - Environmental Events (i.e. Acetaminophen digestion)
 - Others (i.e. Binding in cytosol)

5 Results

- To be discussed Saturday

6 Discussion

NLP: Effectiveness of using computer science data structures (particularly graph structures) to store knowledge

- Explore connections of known relations to find new ones mentioned in new papers
- Find papers that back up a relation link
- Find papers that support or contradict each other edge
- Find the importance of something, i.e. its multitude of relations to others histogram

Web App: Supporting transparency of our research

- Visualizing the relations to easily identify connections graph
- A tool to find large-scale trends and relations previously overlooked
- An simpler interface for anyone to use interaction
- A way to test and reflect the well-roundedness of this method

7 Conclusion

- Describe what we learned about aging as a result of the methods we used.
- How the methods we used help us to accomplish our task, and the possibility of generalizing it to help apply it to related studies.
- Describe how we can improve the process.

8 Bibliography

9 Appendices