Systems Literature Analysis Engine (SLAE)

AI @ UNC Projects

Hanqi Xiao[1], Robert Peters[2], Davyd Voloshyn[1], Baeden Vester[1], Kristen Hassmiller Lich[2*]

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

1. UNC-Chapel Hill College of Arts and Sciences - Computer Science
2. UNC-Chapel Hill Gillings School of Global Public Health - Health Policy and Management
*Principal Investigator

Introduction

Public health challenges are often dynamically complex.[1] Dynamic complexity includes several characteristics: 1) Web of cause-and-effect relationships 2) Varying time delays 3) Feedback loops 4) Nonlinear relationships.

These characteristics make understanding systems and responding to them effectively very difficult. They are the source of policy resistance, the process by which a complex system that is imperfectly understood undermines the effectiveness of interventions.[2]

The first step to combating policy resistance is understanding a system's structure. Researchers in public health often visualize dynamically complex system structure using Causal Loop Diagrams (CLDs) as shown in Figure 1.[3] CLDs show relationships between variables with an S (or +) to show causal linkages where variables move in the same direction and an O (or -) to show causal linkages in which variables change in opposite directions. Feedback loops are defined as reinforcing (R) or balancing (B).

Typically, relationships in CLDs are informed by stakeholders working close to the issue of interest, but strong models are informed by the scientific literature as well.[3] However, the vast number of relationships present in robust CLDs limits researchers' capacity to undertake literature reviews to support each relationship. Thousands of papers would need to be reviewed and synthesized, an impractical goal without AI.

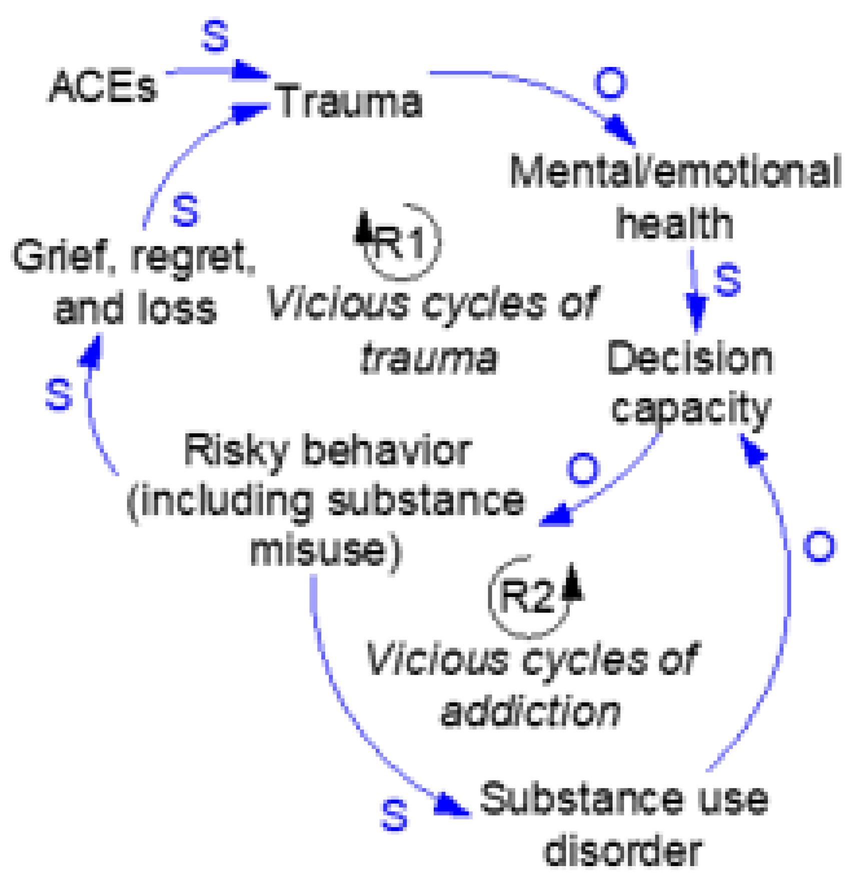

Figure 1: An example of a simple causal loop diagram hypothesizing reinforcing loops through which trauma begets more trauma.

Aims

We aim to pilot a novel tool that automates the extraction of key relationships from scientific literature and presents them in a CLD using generative AI and NLP pipelines. This approach is tested on a small dataset of scientific literature relating to the determinants of Major Depressive Disorder. This topic was chosen for its public health significance.

Second, we generalize pipelines into a Modular NLP Pipeline and aim to host an NLP contest to crowdsource innovative ideas and improve the accuracy of the pipeline.

Methods

SLAE was piloted on a narrow test dataset of literature about depression using search terms derived from Wittenborn 2015.[5] Abstracts were screened for papers that studied multiple causal relationships through experimental methods, meta-analysis, or systematic review. 28 papers (400 relations) were identified and coded.

One reviewer (RP) read each full text and extracted unique relationships. Coded fields included the variables of interest, the relationship between them (direct, indirect, independent), whether that relationship was proposed to be causal, and any additional attributes.

Modular NLP Pipeline

The modular NLP pipeline intakes a CLD diagram and a list of relevant academic literature. The CLD diagram is decomposed into a list of relationships (often 100+ relationships). The pipeline matches literature to a list of relationships (often 10-15 relationships) through semantic similarity.

An accuracy evaluation system automatically decodes coded papers (see methods) by extracting coding comments from pdf files. Extracted files are used to validate predictions by matching correlation predictions (see introduction) between ground truth and pipeline predictions.


```
{
    "PaperTitle": "test_paper_2",
    "PaperContents": "The review establishes several correlational relationships in the context of
    "Relations": [{
        "VariableOneName": "Completing Computerized Working Memory Training (CWMT)",
        "VariableTwoName": "self-reported cognitive functioning ",
        "RelationshipClassification": "direct",
        "IsCausal": "yes",
        "Attributes": "low sample size ",
        "SupportingText": "Completers showed large significant improvements in subjective cognitive
    }]}
```


Figure 2: I/O format for modular NLP Pipeline

Contest to Improve accuracy of extractions

The extraction task is difficult and irregular for LLMs. An NLP contest was run to improve the initial **baseline accuracy of 5.882%**. In a competition including 38 entrants the **top performer reached 45.8%**. Several techniques emerged to increase accuracy, such as batch-processing relationships through LLMs, using alternative classifiers such as BERT, and programmatically fixing common JSON formatting issues from LLMs.

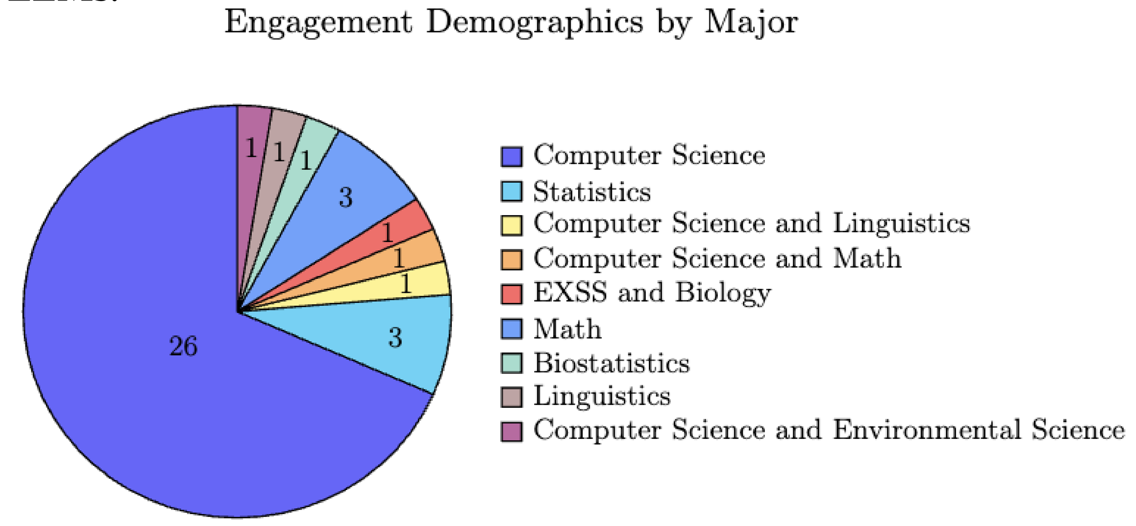

Figure 3: istribution of Participants by Major (38 responses)

Rank	1	2	3	4	5
Accuracy	45.79439	43.92523	42.05607	28.03738	22.42990

Table 1: Top Five Results for Accuracy and Rank

User Prototype

A prototype was developed that allowed us to demonstrate a proof-of-concept user experience of the system. In our prototype, the user can start with a map from either of the two most popular platforms for CLD diagramming (Kumu or Vensim). The system will convert the graph to our unique format and pair each relationship/connection with relevant papers from an existing dataset.

A modular NLP pipeline is then triggered to make predictions based on these pairings. The result is reserialized into kumu format as represented in **Figure 2**. Red indicates that the analyzed literature disagrees with the original classification. Size indicates the amount of relevant literature analyzed. Explanations for extractions are attached to each relation, as shown in the left sidebar.

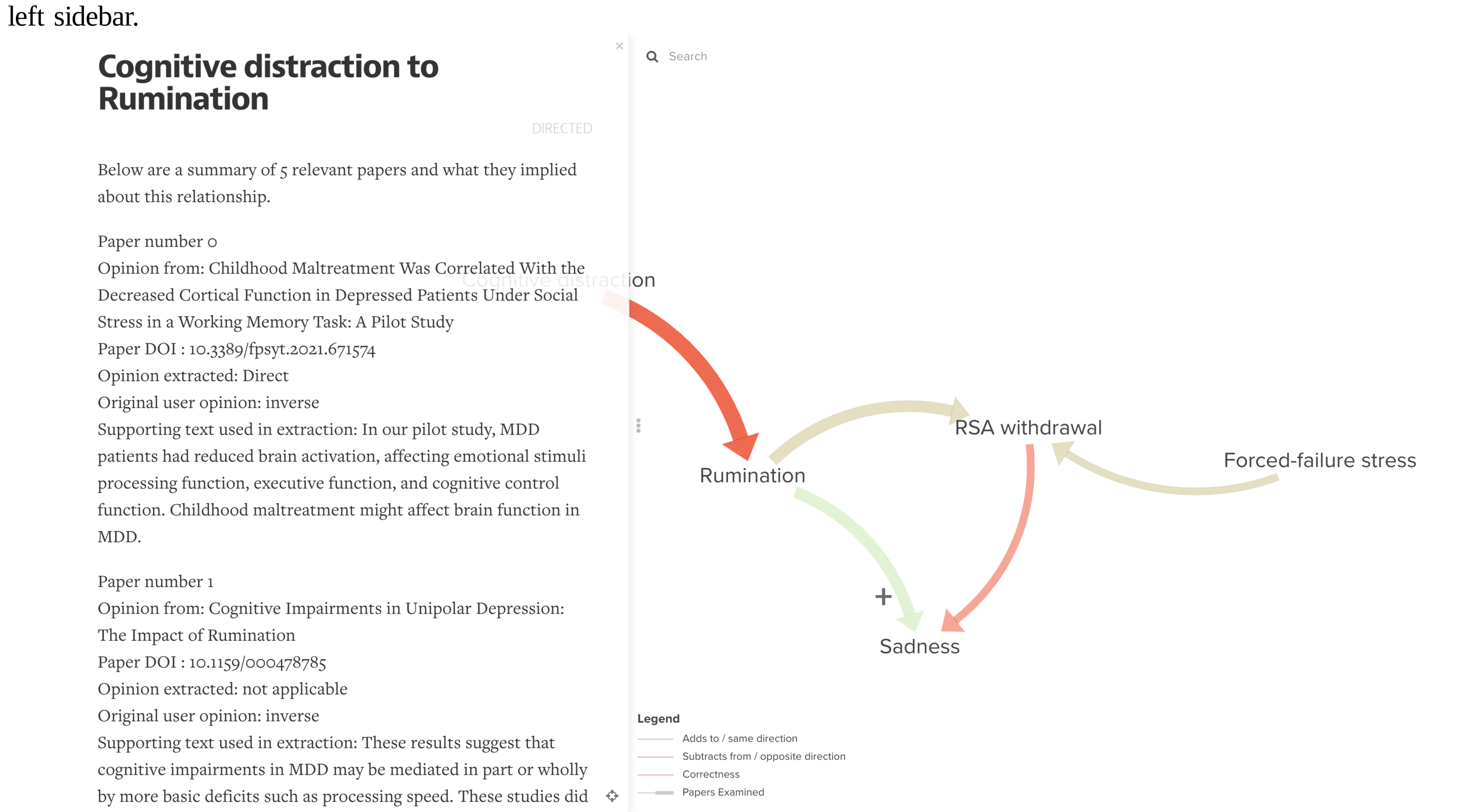

Figure 4: CLD output in Kumu from the SLAE with relationship description in the sidebar. (see User Prototype)

Discussion and Conclusion

SLAE, demonstrated proof-of-concept for reading a scientific paper and producing a CLD in Kumu that showed relevant relationships. The accuracy of the model was quite poor at baseline but improved **eight-fold** over the competition.

SLAE's strengths are its adaptability and scalability. The extraction pipeline is fully modular, easily evaluable, and captures SOTA advancements in machine learning with few changes. SLAE can analyze academic literature much faster than humans.

Future Directions

The SLAE team will be working with Dr. Snigdha Chaturvedi to improve the accuracy of the modular NLP Pipeline.

The current training data is limited to 400 relationships over 28 papers. Model efficiency and accuracy can be improved with more training papers. SLAE is currently seeking funding to support research on improving model accuracy. Rapid improvement is expected.

References

1. Sterman J. Business dynamics: systems thinking and modeling for a complex world. *Irwin McGraw-Hill*. 2000.
2. Sterman JD. Learning from evidence in a complex world. Am J Public Health. 2006 Mar;96(3):505-14. doi: 10.2105/AJPH.2005.066043. Epub 2006 Jan 31. PMID: 16449579; PMCID: PMC1470513.
3. Littlejohns L, Hill C, Neudorf C. Diverse approaches to creating and using causal loop diagrams in public health research: recommendations from a scoping review. Public Health Rev. 2021 Dec 14;42:1604352.
4. Wittenborn AK, Rahmandad H, Rick J, Hosseinichimeh N. Depression as a systemic syndrome: mapping the feedback loops of major depressive disorder. Psychol Med. 2016 Feb;46(3):551-62. doi: 10.1017/S0033291715002044. Epub 2015 Dec 1. PMID: 26621339; PMCID: PMC4737091.

Acknowledgements

AI@UNC (Artificial Intelligence Club at UNC)
UNC Gillings School of Global Public Health
UNC Department of Computer Science
1789 Student Venture Fund

Contacts

Hanqi Xiao – hanqix@ad.unc.edu
Robert Peters – rrober@ad.unc.edu
Dr. Kristen Hassmiller Lich – klich@unc.edu
AI@UNC Club - ai-unc.com

Artificial Intelligence Club at UNC