HEBI LI

lihebi.com@gmail.com | lihebi.com | lihebi

EDUCATION

Iowa State University *Aug. 2014 - Aug 2019 (Expected)*
PhD Student in Department of Computer Science

University of Science and Technology of China *Aug. 2010 - Jun. 2014*
B.E. in Electrical Engineering

RESEARCH AREA

- Artificial Intelligence, Causality
- Natural Language Processing
- Programming Languages & Analysis

PROGRAMMING SKILLS

- Lisp & Scheme
- C/C++
- Python

RESEARCH PROJECTS

NLP: Semantic metric for abstract doc summarization (Python, Tensorflow) *Spring 2018 - Present*

- Website & Code: `https://github.com/lihebi/anti-rouge`
- Description: ROUGE is the de facto criterion for summarization research. However, its two major drawbacks: 1. favors lexical similarity instead of semantic similarity 2. require a reference summary which may be expensive to obtain. Therefore, we introduce a completely new end-to-end metric system for summary quality assessment by leveraging recently developed **deep sentence embedding**.

  We develop two **negative sample generation** approaches: *random mutation* and *cross-pairing*. We apply and evaluate three sentence-embedding models: **Universal Sentence Encoder** by Google and **InferSent** by Facebook, and a vanilla word-embedding model **Grove**. We evaluate 3 neural network architectures atop the embedding, **Fully-connected**, **CNN**, and **LSTM**.
- **Hebi Li**, Qi Xiao, Yinfei Yang, Forrest Sheng Bao, *"End-to-end semantics-based summary quality assessment for single-document summarization"*, In submission.

AI: Causal Discovery From High Dimensional Data (Python, Tensorflow) *Spring 2018 - Present*

- Description: Causal relationships are fundamental for predictions of the consequences of actions. However, most causality research are done in low-dimensional data. We hence are interested in **causal discovery in high dimensional data** such as images and text. In particular, we are interested in mapping low-level raw pixels (micro variables) into high-level features (macro variables) that have explicit causal relations. The model is built atop **Variational Causal Encoder (VAE)**.

PL: Demand-driven Dynamic Program Analysis (C++, LLVM/Clang, Racket) *Summer 2015 - Spring 2018*

- Website: `https://helium.lihebi.com/`. V1 Code: `https://github.com/lihebi/Helium` (723 commits, 22k C/C++). V2 Code: `https://github.com/lihebi/helium2` (143 commits, 1.7k Racket, 3k C/C++).
- Description: We develop Helium, a framework that debugs programs on-demand, preventing running time overhead by running just enough code. It analyzes a buggy program and generates a much smaller partial program that retains the same bug. It features a **syntactic patching algorithm** that find the extra code in addition to the user selection that is necessary for a valid partial program. It also features a **demand-driven context search algorithm** to find smaller partial programs that preserve a given program property.
- Implementation highlights: The framework uses **LLVM/Clang** framework as the underlying parser, and performs **AST** modelling and manipulation, and generates partial programs together with dynamic tests. The framework is provided as **docker** image that is easy and ready to use. Partial program ASTs are visualized through **Graphviz** framework.
- **Hebi Li**, Wei Le, *"Enabling Dynamic Analysis for Partial Programs Via Syntactic Patching"*, In submission.
- **Hebi Li**, Wei Le, *"Demand-Driven Dynamic Analysis for Automatic Benchmark Building"*, In submission.