HEBI LI

1865 Long Rd, Unit A, Ames IA 50010
hebi@iastate.edu | 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, Deep Learning, Causality**
- **Natural Language Processing**
- **Programming Languages & Analysis, Software Engineering**

PREFERRED PROGRAMMING SKILLS

- **Lisp & Scheme dialects**: Racket, Common-Lisp, Emacs Lisp
- **C/C++**: familiar with LLVM/Clang, Boost
- **Python**: familiar with Tensorflow, BeautifulSoup

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**.

  Existing data set like CNN/DM contains only positive samples. We hence develop two **negative sample generation** approaches: *random mutation* and *cross-pairing*. We apply and evaluate different sentence-embedding models: **Universal Sentence Encoder** by Google and **InferSent** by Facebook, and a vanilla word-embedding **Grove** model. We evaluate 3 kinds of networks, Fully-connected, CNN, and LSTM.
- Publications: Hebi Li, Qi Xiao, Yinfei Yang, Forrest Sheng Bao, *"End-to-end semantics-based summary quality assessment for single-document summarization"*, In submission to NAACL-2019.

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

- Description: Causality is believed by many to be the right way for Artificial General Intelligence. However, most causality research are done in low-dimensional data. We hence are interested in discovering causal relations embedded inside 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&SE: Dynamic Program Analysis on Demand (C/C++, LLVM/Clang, Racket) *Summer 2015 - Spring 2018*

- Website: `https://helium.lihebi.com/`
- Project Code:
  - `https://github.com/lihebi/Helium` (723 commits, 22k C/C++)
  - `https://github.com/lihebi/helium2` (143 commits, 1.7k Racket, 3k C/C++)

- Description: Helium is a framework that debugs your program on-demand, preventing running time overhead by running just enough code. It generates a small partial program from the original big buggy program by performing data and control flow analysis and dynamically test to verify the 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**. **Program Generation** steps outputs partial programs, including **Unit test generation** and **dynamic test data generation**. The framework is provided as **docker** image that is easy and ready to use. Partial program AST are visualized through **Graphviz** framework.
- Publications:
  - 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.

USTC Bachelor Thesis: Named Data Network *Summer 2013 - Summer 2014*

Publications: Hebi Li, Xiaobin Tan *"TreeSync: A Distributed Message Synchronization Algorithm Using Topology-Related Hierarchy over Named Data Network"*, USTC Thesis, 2014

OPEN SOURCE PROJECTS

- **hn.el (ELisp)**: A Hacker News Client for Emacs. `https://github.com/lihebi/hn.el`
- **simple-drill.el (Elisp)**: A Flashcard Program for Emacs. `https://github.com/lihebi/simple-drill.el`
- **smart-scholar.el (Elisp, Python)**: An Emacs plugin for downloading and managing bibs for CS conferences. `https://github.com/lihebi/smart-scholar.el`

TEACHING EXPERIENCE

TA in Network Programming, Applications, and Research Issues (ISU CS587)	*Spring 2019*
TA in Advanced Design and Analysis of Algorithms (ISU CS511)	*Spring 2019*
TA in Principles of Programming Languages (ISU CS342)	*Fall 2016, Fall 2017*
TA in Introduction to Database System (ISU CS363)	*Fall 2014, Spring 2015*