

# AHRIM HAN, PH.D.

## Summary

I am working to transit from academia to industry to work as a data scientist. I have strong research backgrounds to be a good data scientist. Over 14 years, I have researched software engineering in computer science for improving software design/code quality using statistics and machine learning techniques. While doing research, I realized that I am good at dealing with big data and finding new insights. I am fast and love to learn new technologies. Recently, I have developed a deep learning model to automatically classify movie reviews in Python with Keras. During my research days, I have focused on software-related problems, and now I have special interests for human-related problems such as "user behavior prediction" and "user preference prediction".

## Employment

### Springboard

Data Science Career Track - Student

Oct. 2018 to Current

- Mastering skills in Python, SQL, data wrangling, data visualization, exploratory data analysis, and machine learning methods
- Performed two data science related capstone projects: "*Prediction of Scores for Public Schools in California*" and "*Sentimental Analysis of Movie Reviews using a Deep Learning Neural Network*"

### Korea University

Research Professor

Seoul, South Korea

Sept. 2013 to Apr. 2018

- Advised graduate students to develop research topics and conduct the experiments
- Received the two National Grants for my research projects, "*Efficient Refactoring Candidate Identification*" (Principal Investigator)
- Published the research results in the top conferences and journals such as *IEEE Transactions on Software Engineering* (Google Scholar)

### Korea Advanced Institute of Science and Technology (KAIST)

Graduate Researcher

Daejeon, South Korea

Mar. 2005 to Aug. 2013

- Actively performed my research project "*Improvement of Change-Proneness Prediction*" (papers cited 64 times)
- Mentoring and Taught in Computer Science courses: "Introduction to JAVA programming" and "Capstone Projects in CS"
- Implemented program structure analysis and metric measurement tools using **Java** and **Python**

### Peace Corps (Headquarters)

Intern

Washington D.C, USA

Aug. 2004 to Oct. 2004

Responsible for organizing the intranet web pages in the Technical Infrastructure and Support Team under the Chief Information Officer

## Projects

### Sentimental Analysis of Movie Reviews using a Deep Learning Neural Network (Springboard)

Mar. 2019 to June 2019

We have developed a deep learning model to automatically classify movie reviews as positive or negative in Python with Keras. We especially identify and deal with overfitting and use a pre-trained embedding in a neural network model to improve accuracy.

### Prediction of Scores for Public Schools in California (Springboard)

Oct. 2018 to June 2019

I applied many of the data science-related techniques for data wrangling, exploratory data analysis, data visualization, and machine learning modeling. I provided the prediction models using the regression and classification algorithms for finding the inferior schools that need help.

### Efficient Refactoring Candidate Identification (Korea University)

Mar. 2008 to Apr. 2017

I developed an **efficient refactoring candidate recommendation system** that helps software developers change code more easily. The main challenge is the heavy computation for evaluating large refactoring candidates, therefore, I developed the following methods in order to increase computational efficiency.

- I developed the *fast graph-based coupling metric* for calculating the effects of suggested refactorings based on matrix computation. This sacrifices some degree of precision (e.g., Connectivity: precision = 0.52, recall = 0.96) but **significantly reduces the computation complexity**, which can be helpful in analyzing large-scale software (e.g., our approach: **154 sec** vs. EPM (previous approach): **28,622 sec**).
- I suggested the *two-phased search-based refactoring identification method* predicting the probabilities of the candidates that are likely to improve "maintainability" and evaluating only the candidates with higher chances. Compared to the previous research of the no-reduction approach, the two-phased refactoring identification approach is **2.6 (min) to 13.5 (max) times faster in computation time**.

### Improvement of Change-Proneness Prediction (KAIST)

Mar. 2006 to Feb. 2010

I developed a **more accurate change-proneness prediction model** to help software developers to focus their efforts on more critical maintenance activities. I developed the **new behavioral dependency metrics** that capture the dynamic aspects of the program. Using dependency metrics in conjunction with existing program structural metrics, we built the regression model for change-proneness prediction. **The accuracy (R-square) of our model is increased 8% than the model using only structural metrics.**

## Awards

National Research Foundation of Korea - Individual Basic Science & Engineering Research Program

Nov. 2014

National grant awarded to top researchers and received \$125,000 for my research project "*Efficient Refactoring Candidate Identification*"

National Research Foundation of Korea - Post-Doctoral Fellowship Grant

Nov. 2013

National grant awarded to top Post-doc researchers and received \$33,000 for my research project "*Efficient Refactoring Candidate Identification*"

## Contact

✉ ahrimhan@gmail.com

🌐 ahrimhan.github.io

☎ (949)910-8751

📍 913 Piccolo, Irvine, CA, 92620  
(Permanent Resident and Eligible to Work in the U.S)

in ahrimhan

🔗 ahrimhan

## Education

Korea Advanced Institute of Science and Technology (KAIST)

Feb. 2007 to Aug. 2013

Ph.D. Computer Science 2013

Korea Advanced Institute of Science and Technology (KAIST)

Sept. 2004 to Feb. 2007

M.S. Computer Science 2007

Sogang University

Feb. 2000 to Feb. 2004

B.E. Computer Science 2004

## Skills

### DATA ANALYSIS, VISUALIZATION, AND PREDICTIVE MODELING

Pandas

Numpy

Scipy

Matplotlib

Pyplot

Seaborn

Scikit-Learn

Keras

Tensorflow

### STATISTICAL METHODS AND MACHINE LEARNING

Classification

Regression

Clustering

Hypothesis Testing

### PROGRAMMING LANGUAGES AND TOOLS

Python

SQL

Java

R

Fortran

Assembly

Markdown

Latex

SPSS

### RESEARCH AREAS

Software Engineering

Refactoring

Software Quality Improvement

Software Design Measurement