于悦

http://blog.csdn.net/m0_37924639

18811395702   617302261@qq.com

教育背景

| 北京邮电大学 | 智能科学与技术专业 | 2016.09 - 2019.06 |
| 北京邮电大学 | 通信工程专业 | 2012.09 - 2016.06 |

项目经历

基于语义的软件缺陷预测技术研究   2017.12 - 2018.9

利用CBOW模型预训练特征token，构建LSTM模型进行软件缺陷预测

- 从Java程序的语法树中抽取函数调用，类实例创建和控制流结点组成特征序列
- 利用word2vec将特征序列中的元素（函数名和类名）映射成词向量
- 以词向量序列和程序标签构建LSTM模型，对程序进行是否包含缺陷的分类

基于机器学习的软件脆弱性分析技术研究   2017.05 - 2018.05

将程序token序列输入DBN模型生成程序语义特征，构建机器学习模型进行软件缺陷预测

- 从C/C++和Java程序的抽象语法树中抽取程序结构特征构成特征序列，对Linux平台下与Windows平台下的二进制程序（ELF，PE）进行逆向，从中抽取关键库函数调用，作为程序的特征(token)序列
- 将token序列输入深度信念网络(DBN)模型生成语义特征
- 整个项目在linux下开发完成

基于弱监督的语义对齐技术研究   2016.01 - 2016.05

利用隐马尔可夫模型进行文本和文本语义信息的对齐

- 利用隐马尔可夫模型对文本和文本的语义信息之间的关系进行建模
- 利用EM算法对模型参数进行求解
- 在弱监督（半监督）情况下对模型的准确性进行研究

其他经历

2018中国高校计算机大赛——大数据挑战赛 排名前10% (120/1400)

利用快手用户30天内的行为数据预测未来一周的活跃用户

特征为最后一次登录日期、登录天数、30天内播放、点赞、关注等行为的最大值、最小值、标准差等

分类模型选择lightgbm模型，最终得到F1值为81.9%

个人技能

熟悉常用机器学习算法：LR, RF, GBDT, XGBoost等

熟悉常用深度学习算法：LSTM, word2vec等

熟悉python, C/C++, linux shell及深度学习框架Tensorflow

CET-6成绩618分，能够较熟练用英语进行听说读写

Yue Yu

http://blog.csdn.net/m0_37924639

18811395702   617302261@qq.com

Education

Beijing University of Posts and Telecommunication Intelligence science and technology   2016.09 - 2019.06

Beijing University of Posts and Telecommunication Communication Engineering   2012.09 - 2016.06

Experience

Learning Semantic Features for Software Defect Prediction   2017.12 - 2018.9

Learning semantic features from code tokens using CBOW model and build LSTM to perform software defect prediction.

- Extract method invocation, class instance creation and control flow nodes from Java Abstract Syntax Trees to generate token sequences.
- Transform tokens into real-valued vectors using word2vec.
- Build LSTMs with vector sequences and the corresponding labels and perform defect prediction.
- LSTMs are implemented with tensorflow.

Software Defect Prediction via Machine Learning   2017.05 - 2018.05

Generating program features by feeding code tokens into DBN models and building machine learning models with these features to perform defect prediction.

- Extract programs' code tokens from C/C++ or Java Abstract Syntax Trees, or API calls from binaries (PE/ELF) to generate token sequences.
- Feed token sequences into Deep Belief Nets to generate programs' semantic features.
- This project is implemented on Linux.

Learning Semantic Correspondences with Less Supervision   2016.01 - 2016.05

Learning the correspondences between text and semantic representations with HMMs.

- Model the correspondences between text and semantic representations via Hidden Markov Models.
- Calculate the parameters by Expectation Maximization Algorithm.
- Evaluate the accuracy of the model with less supervision.

Skills

CET-6   618