Mining Heterogenous Relationships from Pubmed Abstracts Using Weak Supervision

This manuscript (permalink) was automatically generated from greenelab/text_mined_hetnet_manuscript@d3fa3f1 on June 26, 2019.

Authors

- **David N. Nicholson**
  0000-0003-0002-5761 · danich1
  Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania · Funded by GBMF4552

- **Daniel S. Himmelstein**
  0000-0002-3012-7446 · dhimmel · dhimmel
  Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania · Funded by GBMF4552

- **Casey S. Greene**
  0000-0001-8713-9213 · cgreene · GreeneScientist
  Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania · Funded by GBMF4552 and R01 HG010067

Abstract

This is a **rough draft** of a manscript on label function reuse for text mining heterogenous relationship from Pubmed Abstracts.

#Introduction Set introduction for paper here Talk about problem, goal, and significance of paper

Recent Work

Talk about what has been done in the field in regards to text mining and knowledge base integration

#Materials and Methods

##Dataset Talk about dataset - Pubtator Talk about preprocessing Pubtator Talk about hand annotations for each realtion

Label Functions

describe what a label function is and how many we created for each relation

Training Models

Generative Model

talk about generative model and how it works ### Word Embeddings mention facebooks fasttext model and how we used it to train word vectors ### Discriminator Model talk about the discriminator model and how it works ### Discriminator Model Calibration talk about calibrating deep learning models with temperature smoothing

Experimental Design

talk about sampling experiment

Results

Random Sampling of Generative Model

place the grid aurocs here for generative model

Discriminator Model Builds Off Generative Model

place the grid of aurocs here for discriminator model

Random Noise Generative Model

place the results of random label function experiment

place figure of number of new edges that can be added to hetionet as well as edges we can reconstruct using this method

Discussion

Here mention why performnace increases in the beginning for the generative model then decreases

Discuss discriminator model performance given generative model

Mention Take home messages

1. have a centralized set of negative label functions and focus more on contstructing positive label functions

Conclusion and Future Direction

Recap the original problem - takes a long time to create useful label function

Proposed solution - reuse label functions

Mention incorporating more relationships Mention creating a centralized multitask text extractor using this method.

References