ENTITY RELATION RANKING METHODS

{ SUMIT ASTHANA, DR. ASIF EKBAL } IIT PATNA, DEPT. OF CSE - 1301CS42

OBJECTIVES

This project deals with developing a model for ranking knowledge base entity relations based on their relevance. For simplicity we take person-profession relations for developing the model. E.g Barack Obama is a politician and Lawyer so these get a high score during ranking, but law professor would get a low score.

MOTIVATION

The problem of ranking entities in relation is a new problem and is of importance in today's time when search engines are utilizing more and more knowledge bases for showing their results.
By ranking the entities related to a particular entity, we can extract only the most useful one's.

PROBLEM AND DATASET

- The inputs are rated person profession-pairs and their Wikipedia articles.
- Outputs are Relevance scores for each person-profession pair.

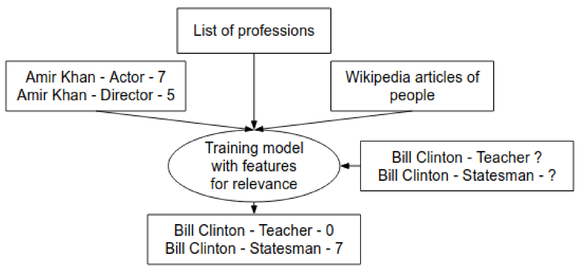


Figure 3: Model Overview

Training pairs	513
Testing pairs	511
Total profession entities	200
Total persons with their possible professions	343,329

Table 2: Dataset statistics

- We propose a feature based approach as well as a **novel** CNN-based learning approach to predict relevancy of profession.

RESULTS - FEATURE BASED

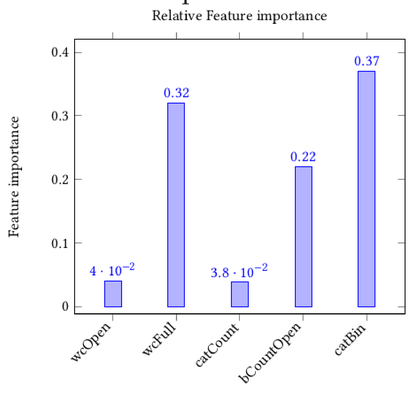
Three metrics were used:

- Accuracy.
- Average score difference.
- Kendall's tau.

Method	Accuracy	ASD	Kendall's
Counting	0.69	1.9	0.39
SVM	0.70	1.86	0.35
Random Forest	0.71	1.78	0.33

Table 3: Scores with classifiers

Feature importance table:



FEATURE BASED LEARNING

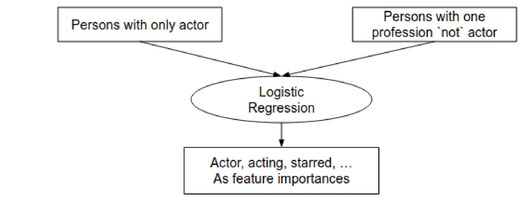
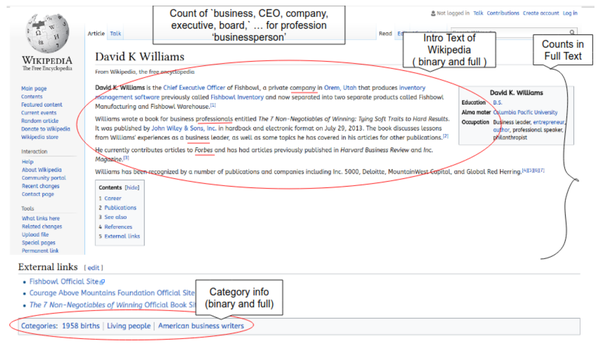


Figure 1: Capturing word importance for one profession



- **Left: Finding important words for each profession.**
- **Bottom-left: Calculation of features from Wikipedia text of person using important words.**
- **Bottom: Feeding these features to a classifier to generate a score.**

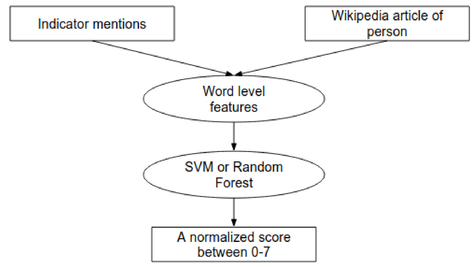
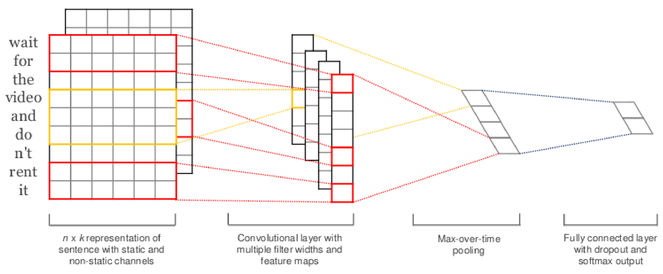


Figure 2: Important words and Wikipedia articles fed to a classifier.

DEEP CNN NETWORK FOR CLASSIFICATION



- 300 dimension Google News word vectors as input to the CNN network.
- The network learned features to predict P(e | t), how much an entity belongs to a type.
- The precision in the next table represents the accuracy of prediction of a single profession as relevant or non-relevant.

Samples	22638
Hidden Layers	2
Hidden Neurons	512, 64
Embedding Dimension	300
Precision	82%

Table 1: CNN statistics

REFERENCES

[1] Yoon Kim. Convolutional neural networks for sentence classification. In *EMNLP*, pages 1746–1751. ACL, 2014.

[2] Hannah Bast, Björn Buchhold, and Elmar Haussmann. Relevance scores for triples from type-like relations. In *SIGIR*, pages 243–252. ACM, 2015.

FUTURE RESEARCH

- The use of CNN based model for ranking is a novel one and a lot of experiments can be done with them especially as they're promising.
- Further, attention can be stacked on top of the model to improve the accuracy.
- Features can be combined with CNN to get a new model for generating rankings.

CONTRIBUTIONS

- Open source cmd-line tool for ranking profession-person pair.
- Experiments with feature based and Deep network based learning.
- Novel Approach of CNN for this problem, which showed good results.