



COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

MSc Defence

Friday August 31, 2018 at 10AM in MacKinnon, Room 318

Automatically Coding Occupation Titles to a Standard Occupation Classification

Negin Nahoomi

Chair: Dr. Mark Wineberg

Advisor: Dr. Fei Song

Co-Advisor: Dr. Gary Grewal

Non-Advisory Committee: Dr. Luiza Antonie

Abstract:

Occupation Coding is the process of classifying occupation titles into one of multiple occupation categories that are usually organized into a hierarchical structure. Historically, the task of classifying occupation titles to standard classifications was done manually. However, the drawbacks of manual coding such as time and cost ineffectiveness led researchers to develop automatic methods for occupation coding. Occupation coding has applications in different fields. For instance, it is useful for identifying diseases and injuries related to occupation classes, inferring social classes, and analyzing labour market statistics and wage regressions such as gender wage gap. We compare the classic machine learning approaches and the deep learning approaches on classifying occupation titles to Standard Occupational Classification (SOC). We implement flat models with Naïve Bayes, Maximum Entropy (MaxEnt), Support Vector Machines (SVM), and Convolutional Neural Networks (CNN) as well as hierarchical models with SVM and MaxEnt to code occupation titles to SOC. For this purpose, 65,962 SOC annotated job titles are collected from publicly available sources. The job titles are extremely short with an average of three words per job title. Our experimental results show that MaxEnt, SVM, and CNN perform similarly and are better than Naïve Bayes on coding occupation titles to SOC.