Semantic texual Similarity Analysis Based Recommender System with Deep Learning

by

Mingzhe Du

Ph.D. Proposal in

Computer Science and Engineering

The College of Engineering and Computing

University of South Carolina

2016

Dr. Jose M. Vidal, Major Professor

Dr. Maton M. Matthews, Committee Member

Dr. Jianjun Hu, Committee Member

Dr. Rose?, Committee Member

Dr. Barry Markovsky, Committee Member

CHAPTER 1

Introduction

With the continuous expansion of Internet activities and online merchandise, recommender systems play an increasingly critical role in the interactive Internet environment[1-5]. Recommender systems apply data analysis in order to help users find their most wanted products from online merchandise sites. For instance, a personalized recommender system on Amazon (www.amazon.com) suggests music and books to customers based on the user's personal shopping experience, hobbies and areas of concern. Non-personalized recommender systems, like Zagat(www.zagat.com) and Yelp(www.yelp.com), provide a general restaurant guide based on the input of millions of individuals. The same reviews and rating scores are presented to users no matter who is looking up their sites.

In our research, we have a system with a lexicon which consists of words and definitions. One word may have multiple definitions. The system recommends pre-defined existing {term: definition} pairs when a user tries to add a new {term: definition} pair to the lexicon. The system consists of a repository of sociology terms while trying to keep a minimal lexicon and eliminating redundant definitions. Instead of using reviews and ratings data, our recommendation is based on the semantic textual analysis of definitions. In recent years, Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Tree-Structured LSTM, Deep Structured Semantic Model (DSSM) and some similar deep learning frameworks have been used to compute semantic similarity between two text snippets[6-11].

# Add a background story of wikitheoria

We propose to develop an algorithm that will calculate and output one or more formal representations of an agent’s behavior, similar to a Markov decision process, but with support for cycles, as needed, from the collected information, as well as a tool that will easily allow the alteration of each decision process to meet the requirements of the modeler. This tool will then produce a file in a standard format that can be efficiently read and used in a variety of applications, in order to create initial decision processes for believable behaviors.