



UMass

Dartmouth

COLLEGE OF ENGINEERING

Computer and
Information Science
Department

MASTER'S THESIS (SUMMER 2017)

TOPIC: *Automated Product Feature Extraction Using Recurrent Neural Networks*

PRESENTOR: Vipanch Erram

ADVISOR: Dr. Haiping Xu

DATE & TIME: Tuesday, September 5, 2017, 2:00 PM

LOCATION: Dion 303 (Lab)

COMMITTEE MEMBERS: Dr. Jan Bergandy and Dr. Iren Valova

ABSTRACT

E-commerce websites nowadays provide platforms for their customers to present feedback that would be useful for future customers to compare their options and make their best selections. With the rapid growth of e-commerce, the size of customer feedback data is also growing tremendously. Such big data can provide valuable information that is not only useful to the buyers in making informed decisions, but also manufacturers in quality control. However, manually reading and analyzing product reviews is a very tiresome task. Classifying review sentences into categories based on the topic discussed in a sentence is one of the major steps for feature-specific sentiment analysis. Traditionally, identifying topics in a sentence requires a pre-defined set of product features. In this thesis, we use a deep learning method, called Recurrent Neural Networks (RNN), to automatically identify product features in a review sentence. As a deep learning approach, RNN can effectively capture dependency relationships of words in a sequence by encoding the context of a data point using an internal state. Instead of using manual syntactic or semantic patterns, we employ two RNNs that model the patterns from the raw data, and extract product features from individual sentences. We discuss two techniques to combine the pieces of evidence derived from the two RNNs. A technique to rank the extracted product features based on their relevance is presented, and hierarchical agglomerative clustering is used to cluster the extracted product features into a manageable number of categories. In a case study, we demonstrate the effectiveness of our approach on automated product feature extraction using real product review data collected from Amazon within various product domains.