TABLE OF CONTENTS

LIST OF FIGURES v

1. Introduction 1

2. Related Work 5

    2.1 Review Mining 5

    2.2 Topic Model 5

    2.3 Sentiment Analysis 7

3. Background Knowledge 10

    3.1 Distributions 10

    3.2 Topic Model 17

        3.2.1 Latent Semantic Analysis (LSA) 17

        3.2.2 Probabilistic Latent Semantic Analysis (pLSA) 17

        3.2.3 Latent Dirichlet Allocation (LDA) 19

    3.3 Stanford CoreNLP 22

        3.3.1 Part-Of-Speech Tagger 23

        3.3.2 Sentiment Analysis Toolkit 23

4. Sentence Level Topic Model 29

    4.1 Basic Framework 29

    4.2 Review Text Preprocessing 31

        4.2.1 Review Crawler 32

4.2.2 Review Level Preprocess 32

4.2.3 Sentence Level Preprocess 33

4.3 Topic Model 34

4.3.1 Model Creation 36

4.3.2 Topic Identification 40

4.3.3 Incrementally Learning 49

4.4 Sentiment Analysis 53

4.5 Review Summary 54

5. Case Study 57

5.1 Product Feature Predefine 57

5.2 Review Data Collection and Preprocessing 60

5.3 Experiment 1 63

5.4 Experiment 2 69

6. Conclusion and Future Work 74

Reference 75