A Study on Deep Learning Models for Sentiment Analysis in Hospitality Media

1710121 TRAN Xuan Thang

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology

[Knowledge Science]

March, 2018

Keywords: sentiment analysis, machine learning, deep leaning models, hospitality media, hotel industry.

With the high-speed development of the Internet in this information age, more and more people over the world can easily access the Internet and use websites to share their opinions, feelings, experiences or complains about products or services on social media (e.g., forum discussion, blogs and social networks, travel websites). For both businesses and customers, these contents (such as reviews, tweets, blogs,…) are the worthiest sources for not only enhancing their services, launching new products, making business campaigns but also making their decisions when choosing a specific service or product. However, the explosion of number websites on the Internet had led to the difficulty of manually monitoring the massive volume of opinionated reviews. Additionally, the analyses of experts towards those opinionated text of the services and products are diverse and heterogeneous and may cause biases. To overcome these limitations, researchers have explored a new research direction in natural language processing area named sentiment analysis or opinion mining that is able to analyze opinionated text systematically and automatically.

According to The World Travel and Tourism Council, Travel and Tourism is the fast-growing industry and become the key to every country's economy. Tourism becomes one of the world's largest economic sector by supporting over 300 million global jobs and contributing more than 10 percent of world GDP. In the tourism sector, hospitality industry has occupied the most significant percentage. Specifically, hospitality industry covers accommodations, food and beverages, travels, transportations, and other fields of tourism

sector. Hospitality industry contributes multibillion-dollar annually by providing specific services, products, and experiences for customers based on the leisure time and disposable income of customers. Simultaneously with the explosion of social media, travelers and customers who used those services and products regularly post their comments, opinions, and experiences on websites. Reviews of travelers regarding different characteristics of a service or product bring the advantages for the managers to improve their business, and for the potential customers before choosing a service or product. In finding a way to improve the practical experience of both buy-side and sell-side in the hospitality market, we apply document-level sentiment analysis for hospitality data collected from a user-generated content site named TripAdvisor. Typically, from big data including both quantitative data and qualitative data of customer's reviews, we apply some deep learning models to identify the customers' sentiment opinions expressed in the text reviews.

The contributions of this thesis are first conducting a comprehensive investigation into sentiment analysis and recent related work. Secondly, we investigate the architecture of some deep learning models, including Recurrent Neural Network, Long Short-Term Memory, Gated Recurrent Unit, Bidirectional Long Short-Term Memory, and Convolutional Neural Network, which frequently applied on current research of sentiment analysis using machine learning approach, and they had archived the state-of-the-art results in these tasks. Additionally, we collected a new hospitality media dataset including more than 75,000 reviews of 410 hotels in Ho Chi Minh City, Vietnam up to February 2017 from a well-known, world's largest travel site named TripAdvisor. Finally, we carried out experiments using those above deep learning models with the new dataset to evaluate the performance of them for the complicated and complex task of sentiment analysis task.

As the results showed, although we just employ some general deep learning models with slight configurations in the architectures, they can perform well with the complicated task. Deep learning models averagely classified 89.95 percent correctly of the sentiment opinions on the reviews given from the hospitality media dataset. They completely overperformed some baseline models, including Support Vector Machine and Naïve Bayes, which were also famous in sentiment analysis field. This study adds more contributions to finding the emerging opinions of customers towards the hotel reviews by applying deep learning algorithms which ultimately benefits for further studies in this area.