Cross-Platform Reputation Generation System Based on Aspect-Based Sentiment Analysis

ABSTRACT

The active growth of Internet-based applications such as social networks and e-commerce websites leads people to generate a tremendous amount of opinions and reviews about products and services. Thus, it becomes very crucial to automatically process them. Over the last ten years, many systems have been proposed to generate and visualize reputation by mining textual and numerical reviews. However, they have neglected the fact that online reviews could be posted by malicious users that intend to affect the reputation of the target product. Besides, these systems provide an overall reputation value toward the entity and disregard generating reputation scores toward each aspect of the product. Therefore, we developed a system that incorporates spam filtering, review popularity, review posting time, and aspect-based sentiment analysis to generate accurate and reliable reputation values. The proposed model computes numerical reputation values for an entity and its aspects based on opinions collected from various platforms. Our proposed system also offers an advanced visualization tool that displays detailed information about its output. Experiment results conducted on multiple datasets collected from various platforms (Twitter, Facebook, Amazon . . . ) show the efficacy of the proposed system compared with state-of-the-art reputation generation systems.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, different type of algorithms is trained to make classifications or predictions, and to uncover key insights in this project. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics.

Machine learning algorithms build a model based on this project data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of datasets, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

**EXISTING SYSTEM**

Poria *et al.* presented the first deep learning approach for the AE task in opinion mining. The authors employed a 7-layer deep convolutional neural network to tag each word in the textual opinions as either aspect or non-aspect word. The authors also proposed a set of heuristic linguistic patterns and integrated them with the deep learning classifier which significantly improves the accuracy compared with the previous SOTA methods. In [19], the authors proposed an attention-based long short-term memory (LSTM) [20] for aspect-level sentiment classification. The idea is to learn aspect embeddings and let aspects participate in computing attention weights.

The proposed model can focus on different parts of a sentence when different aspects are given so that they are more competitive for aspect-level classification. The proposed model achieved better results compared with the standard LSTM on the SemEval 2014 Task 4 dataset [21]. In [22], Wei and Toi improved the deficiencies of the previous LSTM approaches by proposing convolutional neural networks [23] and gating mechanisms (GCAE) based model, which has been proved to be more accurate and efficient. The novel Gated Tanh-ReLU Units can selectively output the sentiment features according to the provided aspect or entity. The architecture of the proposed model is much simpler than the attention layer used in the previously existing models.

The experiments on SemEval datasets show a performance improvement compared with the LSTM based models. The authors in [24] proposed an interactive multi-task learning network (IMN) capable of jointly learning multiple related tasks simultaneously at both the token-level as well as the document-level. The IMN introduces a message passing mechanism that allows informative interactions between tasks, enabling the correlation to be better exploited. Experiments on three benchmark datasets, taken from SemEval2014 and SemEval 2015 [25] show that IMN outperforms other baselines by large margins. Since most existing methods ignore the position information of the aspect when encoding the sentence, authors in [26] proposed a hierarchical attention-based position-aware network (HAPN), which includes position embeddings to learn the position-aware representations of sentences to generate the target-specific representations of contextual words. HAPN achieved the SOTA performance on SemEval 2014 dataset compared with the previous methods.

Xu *et al.* [27] presented a review reading comprehension (RRC) task where they adopted BERT [28] as a base model, and proposed a joint post-training and fine-tuning approach for ATE, APC. Experimental results show that the proposed post-training approach is very effective. Later in [29], the authors proposed a novel architecture called BERT Adversarial Training (BAT) to employ adversarial training for AE and APC by generating artificial data which is carried out in the embedding space. The proposed model outperforms the standard BERT as well as the in-domain post-trained BERT in both AE and APC tasks. In [30], the authors exploit domain-specific BERT language model fine tuning in addition to supervised task-specific fine tuning to produce a new SOTA performance on the SemEval 2014 Task 4 restaurants dataset. The authors also showed that cross-domain adapted BERT model performs better than strong baseline models such as XLNet-base [31] and vanilla BERT-base. In [32], the authors compared the induced trees from pre-trained models and the dependency parsing trees on various popular models for the ABSA task.

They found that the induced tree from fine tuned RoBERTa [33] (FT-RoBERTa) outperforms the parser-provided tree. The experiments show that the RoBERTa-based model can outperform or approximate the previous SOTA performances on six datasets across four languages including SemEval 2014 task 4. Recently, authors in [34] proposed a multi-task learning model named LCF-ATEPC for ABSA based on the multi-head self-attention and the local context focus (LCF) [35] mechanisms. The proposed model is multilingual and applicable to the classic English review SA task, such as the SemEval-2014 task4. The proposed model can automatically extract aspects and determine their sentiment polarities. Since the LCF-ATEPC model currently achieves SOTA performance on AE and APC tasks,2 it was selected to be employed in this paper.

**Disadvantages**

* An existing system not implemented Aspect term extractor which performs

the basic token-level classification for each token, which means that each token will be given a label, and a classification is performed to predict the aspects in the sentence.

* An existing system is not implemented Local Context Focus in which Local context is a new technique that can be adapted to most ne-grained NLP tasks.

Proposed System

This system aims at generating a reputation value toward online entities (movies, hotels, restaurants, services, etc.) and computing a satisfaction score toward each aspect of the target entity by processing textual and numerical data collected from multiple platforms. Proposed system presents its architecture. First, we start by collecting users' reviews from different platforms such as Twitter, Amazon, YouTube, etc. Next, an automatic spammers filtering system is employed to detect and eliminate unwanted spam reviews. Then, we apply a SOTA ABSA model to users' textual reviews in order to compute a score based on the sentiment orientation of the extracted aspects from those reviews. Further, we calculate a popularity score and a time score based on statistical features extracted with the textual reviews. Finally, we compute a reputation value based on the previously calculated scores, and we propose a new user-friendly visualization interface that displays in-depth details about the reputation of the target entity.

One of the important features of the proposed system is the ability to collect and process data from various platforms. Previous reputation generation systems gather necessary data from either e-commerce websites such as Amazon, TripAdvisor, or social media platforms such as Twitter and Facebook. In this work, we decided to normalize the features of all platforms in order to create a single merged dataset by classifying the platforms on the Internet into two types: the first type provides the accessibility of extracting the textual review with the number of likes received for that review such as Amazon, YouTube, etc. The second type provides the accessibility of extracting the textual review with the number of likes received for that review along with the number of times the review was shared among the network such as Twitter, Facebook, etc.

**Advantages**

* Multi-Head Self-Attention (MHSA): The multi-head attention mechanism helps the model to learn the words' relevant information in different presentation subspaces. MHSA is based on multiple scale-dot attention that can be used to extract deep semantic features in the context.
* Aspect Polarity Classifier: To perform the sentiment polarity classification, the LCF-ATEPC model combines the local context features and the global context features. Then, the aspect polarity classifier performs a head-pooling on the learned concatenated context features from the feature interactive learning layer.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 7 Ultimate.
* **Coding Language :** Python.
* **Front-End :** Python.
* **Back-End :** Django-ORM
* **Designing :** Html, css, javascript.
* **Data Base :** MySQL (WAMP Server).