Complementary Aspect-Based Opinion Mining

ABSTRACT

Aspect-based opinion mining is finding elaborate opinions towards a subject such as a product or an event. With explosive growth of opinionated texts on the Web, mining aspect-level opinions has become a promising means for online public opinion analysis. In particular, the boom of various types of online media provides diverse yet complementary information, bringing unprecedented opportunities for cross media aspect-opinion mining. Along this line, we propose CAMEL, a novel topic model for complementary aspect-based opinion mining across asymmetric collections. CAMEL gains information complementarity by modeling both common

and specific aspects across collections, while keeping all the corresponding opinions for contrastive study. An auto-labeling scheme called AME is also proposed to help discriminate between aspect and opinion words without elaborative human labeling, which are further enhanced by adding word embedding-based similarity as a new feature. Moreover, CAMEL-DP, a nonparametric alternative to CAMEL is also proposed based on coupled Dirichlet Processes. Extensive experiments on real-world multi-collection reviews data demonstrate the superiority of our methods to competitive baselines. This is particularly true when the information shared by different collections becomes seriously fragmented. Finally, a case study on the public event “2014 Shanghai Stampede” demonstrates the practical value of CAMEL for real-world applications.

**EXISTING SYSTEM**

* Two subtasks are usually involved in this problem, namely, aspect or feature identification and opinion extraction. Most of the early works on aspect identification are feature-based approaches, e.g., applying frequent itemset mining to identify product aspects, which normally exert some constraints on high-frequency noun phrases to find aspects. As a result, they are usually subject to the risk of producing too many non-aspects examples and missing low-frequency aspects. Several early works have applied supervised learning to identify both aspects and opinions, which, however, needs hand-labeled training sentences and thus is very costly.
* Mei el al. [19] propose a topic sentiment mixture model, which represents positive and negative sentiments as language models separating from topics, but both models only capture general opinion words.
* Brody et al. [3] take a two-step approach by first detecting aspects and then identifying aspect-specific opinion words.
* Zhao et al. [35] propose a topic model integrating with a maximum entropy model (MaxEnt-LDA) to jointly capture both aspects and aspect-specific opinion words within a topic model. Detailed discussions about aspect-specific opinion models based on LDA can be found in this system.

**Disadvantages**

* + There is no Dirichlet processes in Opinion Mining and Classification.
  + There is no Sentiment Analysis.

**PROPOSED SYSTEM**

* The system proposes Cross-collection Auto-labeled MaxEnt-LDA (CAMEL), a novel topic model for complementary aspect-based opinion mining across asymmetric collections. To our best knowledge, our work is among the earliest studies in this direction. CAMEL is essentially a type of cross-collection LDA model, which models aspect-level opinions and gains information complementarity by learning both common and specific aspects across different collections. By keeping all the corresponding opinions for both common and specific aspects, CAMEL is also capable of conducting contrastive opinion analysis.
* Moreover, to boost CAMEL, the system proposes AME, an automatic labeling scheme for maximum entropy model, to discriminate aspect and opinion words without heavy human labeling. It is further enhanced to the so-called EAME scheme by employing the word embedding-based similarity. Finally, we propose CAMEL-DP, a nonparametric alternative to CAMEL. CAMEL-DP is based on coupled Dirichlet processes [16], and is capable of automatically estimating the number of common and specific aspects, which might be a headache in practice for parametric models like CAMEL.

**Advantages**

* The system is very fast in opinion Mining due to Enhanced AME with Word Embedding- Based Similarity technique.
* An effective Bayesian nonparametric models, especially those based on Dirichlet processes (DPs), have emerged as an important tool for model selection..

**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 XP
* Coding Language - Java/J2EE(JSP,Servlet)
* Front End - J2EE
* Back End - MySQL