

INTISARI

Pengembangan Model Opinion Mining Pada Berita Berbahasa Indonesia

Oleh

Sigit Priyanta

09/294338/SPA/00265

Perkembangan pemanfaatan internet menyebabkan bertambahnya jumlah dan jenis konten yang ada dalam internet, khususnya text yang menjadi sangat besar dan tersebar di banyak sumber informasi. Informasi textual yang ada di internet dapat dikategorikan ke dalam informasi yang berupa fakta dan informasi yang berupa pendapat. Sumber opini tersebar di berbagai media seperti surat kabar, televisi, forum diskusi dan internet dalam berbagai bentuk (*blog, online review, mailinglist*). Ekstraksi emosi, attitude dan sentiment pada dokumen text dapat dilakukan dengan opinion mining. *Opinion mining* yang dilakukan pada text berita dapat digunakan untuk mengetahui sentiment masyarakat tentang suatu topik tertentu. Penelitian *opinion mining* pada berita Berbahasa Indonesia belum banyak dilakukan terutama yang disertai dengan identifikasi pemilik dan obyek opini.

Opinion mining dapat diimplementasikan dengan menggunakan beberapa proses komputasi, seperti *stemming, part of speech tagging, klasifikasi, named entity recognition* dan penentuan sentimen dari opini yang ada. Penelitian ini menyelesaikan persoalan *opinion mining* dengan menggunakan pendekatan *lexicon* yang dikombinasikan dengan *natural language processing*. Model yang dibangun dikelompokkan ke dalam tiga bidang penelitian yaitu *document subjectivity classifier* dengan *rule based, SVM* dan *NBC*, pengenalan pemilik dan obyek opini dengan *rule based* dan *named entity recognition* dengan *HMM* dan *CRF* dan proses identifikasi dan penilaian sentimen opini.

Telah berhasil dibangun sebuah model *opinion mining* pada berita Berbahasa Indonesia yang terdiri dari komponen *Part of Speech Tagger*, identifikasi kalimat subyektif dengan *rule based*, penentuan pemilik opini dan obyek opini dengan *rule based* dan penilaian akhir sentiment kalimat dengan menggunakan *rule based, fixed windows length* dan jarak obyek opini ke kata-kata penentu sentiment. Berdasarkan pengujian yang dilakukan, proses *rule based sentence subjectivity classifier* memiliki akurasi sebesar 81,45%, presisi 78,4 %, *recall* 96,64% dan *f-measure* sebesar 86,77%, pengenalan pemilik dan obyek opini dapat diselesaikan dengan menggabungkan pendekatan *rule based* dan *named entity recognition* dengan presisi 91,67% dan *recall* 59,19%. Perhitungan jarak ke obyek opini dapat meningkatkan akurasi 2,3%, presisi 2,4, *recall* 3,2% dan *f-measure* 3,7%.

Kata kunci : *opinion mining, analisis sentiment, document subjectivity classifier, named entity recognition, sentiment orientation, Bahasa Indonesia*

ABSTRACT

The Development of Opinion Mining Model at Indonesian News

By

Sigit Priyanta
09/294338/SPA/00265

The growth of Internet's utilization has increased both numbers and types of its content, particularly text, that became very large and spread out in many information sources. Textual information on the Internet can be categorized into fact- and opinion-based information. The sources of opinion are spread out in various media including newspapers, television, discussion forums and Internet in forms such as blogs, online reviews, and mailinglists. The extraction of emotions, attitudes, and sentiments in various text documents can be carried out by using an opinion mining model. The opinion mining performed on news text can be used to find out people's sentiment on a certain topic. Opinion mining research at Indonesian News has not been done with the identification of opinion holder and object of opinion.

Opinion mining can be implemented using several computing processes, such as stemming, part of speech tagging, classification, named entity recognition, and determination of sentiment from the existing opinion. This study examines the issue of opinion mining by combining a lexicon approach and a natural language processing. The model used in the study was classified into three fields of study, namely, document subjectivity classifiers, including rule-based classifier, SVM classifier, and Naive Bayes classifier, opinion holder and object using rule-based and HMM-based named entity recognition, and the identification and assessment of existing opinion sentiment.

This research has successfully develop a model of opinion mining at Indonesian News. Based on the examination, it is concluded that the rule-based document subjectivity classifier had the best results with accuracy value 81,45%, precision 78,4%, *recall* 96,64% and *f-measure* of 86,77%, while the opinion holder and object recognition could be carried out by combining both rule-based and named entity recognition approaches. Additional calculation of the distance of opinion objects to the words as sentiment identifiers can increase accuracy by 2,3%, precision 2,4, *recall* 3,2% and *f-measure* by 3,7% in determination of the sentiment orientation of a sentence.

Keywords: opinion mining, document subjectivity classifier, named entity recognition, sentiment orientation, Bahasa Indonesia