

OPINI SUMMARIZATION UNTUK REVIEW PRODUK DENGAN ALGORITMA POINTWISE MUTUAL INFORMATION - INFORMATION RETRIEVAL (PMI-IR)

Ingga Bia Putri¹, Warih Maharani², Hetti Hidayati³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Semakin berkembangnya teknologi internet saat ini berdampak pada semakin maraknya sistem penjualan barang melalui online shop (e-commerce). E-commerce tersebut biasanya akan meminta para pembeli untuk memberikan review dari produk yang telah dibelinya. jumlah pembeli yang semakin meningkat setiap harinya mengakibatkan semakin meningkat pula review yang diberikan pada sebuah produk. Tugas akhir ini dibuat untuk membangun sistem yang dapat meringkas kumpulan review - review pembeli terhadap suatu produk berdasarkan kualitas fiturnya. Sehingga akan mempermudah calon pembeli yang lain untuk mengetahui kualitas barang yang akan dibelinya berdasarkan review - review produk sebelumnya.

Proses peringkasan review produk yang dilakukan merupakan salah satu proses opinion mining yang diharapkan dapat menghasilkan ringkasan review produk yang berupa kata dengan orientasi kalimat opini positif atau kalimat opini negatif. Proses pertama yang dilakukan pada Tugas Akhir adalah proses ekstraksi review produk berdasarkan fitur produk beserta opini yang mengikutinya, kemudian dilakukan pembuatan aturan (rule) dari hasil ekstraksi yang telah didapat serta perhitungan frekuensi kemunculan rule dalam suatu review, dan terakhir menentukan orientasi kata opini (positif atau negatif). Metode yang digunakan untuk pembuatan rule adalah Association Rule Mining, sedangkan perhitungan frekuensi kemunculan rule dilakukan dengan algoritma Pointwise Mutual Information - Information Retrieval (PMI-IR).

Berdasarkan hasil pengujian yang didapat menunjukkan bahwa algoritma PMI - IR dapat digunakan untuk proses peringkasan review produk dengan menghitung frekuensi tertinggi kemunculan rule yang berupa fitur produk beserta opininya untuk ditentukan orientasinya dengan menggunakan Sentiwordnet.

Kata Kunci : Opinion Mining, Opinion Summarization, Association Rule Mining, Pointwise Mutual Information - Information Retrieval (PMI-IR), Sentiwordnet.

Abstract

The continued development of internet technology is currently more widespread impact on the system of selling goods through online shop e-commerce. E-commerce will usually ask the buyer to provide a review of the product that has been purchased. an increasing number of buyers every day resulting in increasing as well the reviews are given on a product. The final task is to build a system that can summarize a collection of buyer's review of a product based on its quality. That will facilitate other prospective buyers to know the quality of the goods to be purchased based on the product's reviews before.

Process summarizing product's review conducted is one opinion mining process that is expected to generate a summary review of the product, which are opinion words with positive orientation or negative orientation. The first process on the final project is the process of extracting a product review based on the product's features and their opinions are followed, then performed the extraction of rule-making that has been obtained and the calculation of the frequency of occurrence rule in a review, and finally determining the orientation of subjective words (positive or negative) . The method used for the manufacture of rule is Association Rule Mining, while the frequency of occurrence rule calculations performed by algorithms Pointwise Mutual Information - Information Retrieval (PMI-IR).

Based on the test results, it shows that the algorithm PMI - IR can be used to process summarizing product's reviews with the highest frequency of occurrence counting rule in the form of an opinion piece for the product features and their orientation is determined by using Sentiwordnet.

Keywords : Opinion Mining, Opinion Summarization, Association Rule Mining, Pointwise Mutual Information - Information Retrieval (PMI-IR), Sentiwordnet.

1. Pendahuluan

1.1. Latar Belakang

Pemanfaatan teknologi internet pada saat ini menyebabkan banyaknya penggunaan *website* untuk kegiatan transaksi jual beli suatu produk. Hal ini membuat para pembeli dapat dengan bebas memberikan opininya tentang produk yang mereka beli, karena dalam *website* jual beli atau yang lebih dikenal dengan *e-commerce* selalu disediakan ruang untuk pembeli mengeluarkan saran, pendapat atau opininya terhadap barang yang sudah dibeli ataupun barang yang akan dibeli. Data – data opini tersebut tidak hanya berguna untuk pembeli melainkan juga berguna untuk manufaktur penjualan produk. Oleh karena itu butuh dilakukan *review* terhadap suatu produk tertentu untuk mengidentifikasi kelebihan dan kekurangan produk dari sisi pembeli.

Terlalu banyaknya opini yang muncul tentang suatu produk akan menyulitkan pembeli yang potensial untuk membacanya dan mendapatkan informasi dalam hal pengambilan keputusan pembelian sebuah produk. Oleh karena itu akan dilakukan *mining* dan *summarization* pada *customer review* dari sebuah produk untuk mengatasi permasalahan tersebut. *Summarization* yang dilakukan berbeda dengan *text summarization* biasa karena pada *opinion summarization* untuk *review* sebuah produk hanya dilakukan *mining* fitur dari produk tersebut dengan *review* yang berupa opini dari *customer* dan menentukan apakah opini tersebut positif atau negatif.

Dalam proses *summarization* akan dilakukan *mining* terhadap opini – opini yang sudah terkumpul terlebih dahulu dengan pendekatan konsep *association rule*. Konsep ini bertujuan untuk membuat *rule – rule* yang mendefinisikan keterkaitan antara fitur dan opini dari sebuah produk. Penggunaan konsep *association rule* didasari dengan prinsip apriori dimana jika set fitur-opini sering muncul dalam sebuah opini maka semua subset akan mengikutinya. Oleh karena itu konsep ini dapat digunakan untuk menentukan keterkaitan antara opini umum dari suatu produk dan menentukan fitur produk yang sering muncul beserta opini dari fitur tersebut.

Penggunaan data pada sistem yang akan dibuat merupakan data yang bersifat *lexical* atau data yang berhubungan dengan bahasa, sehingga algoritma PMI – IR (*Pointwise Mutual Information – Information Retrieval*) dapat digunakan untuk proses *opinion summarization* karena merupakan algoritma yang menggunakan pendekatan teori informasi untuk menemukan keterkaitan antara kata – kata.. Algoritma PMI – IR menghitung seberapa banyak sebuah kata disebutkan dalam sebuah opini agar dapat ditarik sebuah informasi yang akan kita dapatkan tentang suatu produk. Algoritma PMI – IR merupakan pengembangan dari algoritma PMI dengan menambahkan konsep *Information Retrieval* untuk meningkatkan ketepatan perhitungan kemunculan pasangan kata dalam sebuah opini dan mencatatnya sebagai nilai *hits* yang dihasilkan. Oleh karena itu diharapkan algoritma ini dapat meningkatkan nilai akurasi dari algoritma sebelumnya sehingga dapat menghasilkan *summary* atau kesimpulan yang lebih tepat dalam hal

penentuan kelebihan dan kekurangan suatu produk dari sebuah *customer review*.

1.2. Perumusan masalah

Berdasarkan latar belakang tersebut, permasalahan yang akan dibahas pada tugas akhir ini antara lain :

1. Bagaimana cara meringkas opini berdasarkan fitur dan orientasinya ?
2. Bagaimana cara mengekstrak fitur produk dan orientasi opini dari fitur tersebut dalam sebuah review produk ?
3. Bagaimana cara menentukan pasangan kata fitur dengan orientasi opininya yang muncul bersamaan ?
4. Bagaimana cara menghitung nilai ketergantungan pasangan kata yang merupakan fitur produk dan orientasinya dengan algoritma *Pointwise Mutual Information – Information Retrieval* (PMI – IR) ?

1.3. Batasan masalah

Batasan masalah yang diberikan dalam pengerjaan tugas akhir ini adalah :

1. Penulis tidak melakukan penelusuran web (*web crawling*) untuk mendapatkan data melainkan dengan menggunakan data set online review Amazon.com yang telah tersedia.
2. Data set diasumsikan sudah merupakan sebuah opini sehingga tidak diperlukan tahapan untuk penentuan opini atau bukan dari sebuah *statement* atau pernyataan *customer*.
3. Review produk yang digunakan merupakan review yang secara eksplisit menyebutkan fitur dan orientasinya.
4. Proses pengekstrakan fitur dan opini dari sebuah review produk menggunakan algoritma ekstraksi.
5. Algoritma ekstraksi tidak melakukan proses ekstraksi terhadap kata fitur yang berupa fitur frase.
6. Peringkasan yang dilakukan hanya merupakan pengelompokan kata berdasarkan fitur dan orientasinya.
7. Kata opini yang diekstrak oleh algoritma ekstraksi hanya kata opini yang memiliki tag *adjective*

1.4. Tujuan

Berdasarkan rumusan masalah diatas, tujuan dari Tugas Akhir ini adalah :

1. Menganalisis dan meringkas opini berdasarkan fitur dan orientasinya.

2. Menganalisis dan mengekstrak fitur produk dan orientasi opini dari fitur tersebut dengan algoritma ekstraksi (*extraction algorithm*).
3. Menganalisis dan menentukan pasangan kata fitur dengan orientasi opininya sebagai suatu *itemset* dengan *Association Rule Mining*.
4. Menganalisis dan menghitung nilai ketergantungan pasangan kata yang merupakan fitur produk dan orientasinya dengan algoritma *Pointwise Mutual Information – Information Retrieval* (PMI – IR).

1.5. Metodologi penyelesaian masalah

Metodologi penyelesaian yang digunakan pada penelitian tugas akhir ini terdiri dari 6 tahap yaitu :

1. Studi Literatur

- Pencarian materi yang dibutuhkan dan berhubungan dengan permasalahan, seperti materi tentang *opinion mining*, *opinion summarization*, algoritma PMI - IR dan materi pendukung lainnya.
- Pemahaman materi yang telah diperoleh untuk membantu proses penyelesaian masalah.

2. Pengumpulan data

Pencarian dan pengumpulan data set berupa review produk yang berasal dari online review Amazon.com.

3. Tahap perancangan sistem

Perancangan sistem yang akan dibangun adalah sebagai berikut :

- a. Data cleaning : Melakukan analisis terhadap dataset asli, kemudian dilakukan pembersihan data sehingga menghasilkan data yang siap untuk diproses.
- b. Pre-processing : Melakukan tagging berdasarkan *Part-of-Speech* (POS) oleh sistem ahli dan mengekstraksi kata yang berupa fitur produk dan opini yang berupa orientasi dari fitur tersebut.
- c. Pembuatan Rule : Melakukan pencarian frequent item set dari kata – kata yang telah diekstrak dengan metode *Association Mining Rule*.
- d. Perhitungan PMI–IR : melakukan perhitungan nilai ketergantungan antara fitur dan orientasinya pada *rule* yang telah terekstrak, kemudian dilakukan pemilihan *rule* dengan nilai PMI-IR yang terbaik jika terdapat *rule* dengan fitur yang sama.
- e. Pengorientasian Opini : Mengidentifikasi orientasi kata opini apakah merupakan kata positif atau kata negatif.
- f. Summarization : pengelompokan pasangan kata berdasarkan orientasinya untuk menentukan kelebihan dan kekurangan suatu produk.

5. Tahap Implementasi
Pada tahap ini, akan dilakukan implementasi berdasarkan rancangan yang telah dibuat sebelumnya.
6. Tahap Pengujian dan analisa hasil
Melakukan Pengujian sistem yang dibangun berdasarkan perancangan sebelumnya dengan menggunakan data uji dan menganalisis hasil pengujian berdasarkan performansinya.
7. Tahap penyusunan laporan
Merupakan tahapan terakhir dalam penelitian untuk mengambil kesimpulan berdasarkan hasil pengujian dan dilakukan penyusunan laporan tugas akhir.



5. Kesimpulan dan Saran

5.1. Kesimpulan

Berdasarkan pengujian serta analisis yang telah dibahas dan dilaksanakan pada bab sebelumnya (Bab 4), maka dapat disimpulkan beberapa hal sebagai berikut:

1. Algoritma ekstraksi (*extraction algorithm*) dapat digunakan dengan baik untuk melakukan ekstraksi fitur dan orientasi opini. nilai *precision* yang dihasilkan pada setiap data review produk yaitu mencapai angka diatas 80%.
2. Nilai minimum support yang terbaik pada proses pembuatan *rule* dengan metode *Association Rule Mining* pada masing – masing data review berbeda – beda sesuai dengan karakteristik *rule* yang dihasilkan.
3. Prediksi proses *summarization* dengan menggunakan metode PMI – IR memiliki nilai *precision* yang lebih baik yaitu mencapai 100 %.
4. Tahap perhitungan PMI – IR dipengaruhi oleh jarak antara kata fitur dan opini yang ditentukan oleh variabel *near*. Nilai maksimum variabel *near* yang digunakan pada perhitungan PMI – IR adalah 5.
5. Sentiwordnet dapat digunakan dengan baik untuk melakukan orientasi opini pada hasil *summarization*.

5.2. Saran

1. Pengembangan sistem agar dapat mengidentifikasi orientasi opini dengan tepat pada kalimat dengan opini implisit yaitu kalimat kompleks atau kalimat sindiran.
2. Implementasi metode *preprocessing* yang lain agar dapat menghapus kata-kata yang tidak diperlukan dalam proses orientasi opini.

6. Daftar Pustaka

- [1] Agrawal Rakesh, Imielinski Tomasz, Swami Arun. *Mining Association Rules Between Sets of Items in Large Database*. IBM Almaden Research Center, 650 Harry Road, San Jose, California.
- [2] Esuli Andrea, Sebastiani Fabrizio. *SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining*. Istituto di Scienza e Technologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa.
- [3] Han, Jiawei dan Kamber, Micheline. 2006. *Data Mining : Concepts and Techniques*. 2nd Edition. Morgan Kaufmann Publishers, San Fransisco.
- [4] Hu Mingqing, Liu Bing. *Mining and Summarizing Customer Reviews*. Department of Coputer Science, University of Illinois, Chicago.
- [5] Kaur Ishwinder, Hornof Anthony J. *A Comparison of LSA, Wordnet and PMI-IR for Predicting User Click Behavior*. Computer and Information Science, University of Oregon, Eugene, OR 97403, USA.
- [6] Kim Won Young, Kim Kyu Il, Ryu Joon Suk, Kim Ung Mo. 2009. *A Method for Opinion Mining of Product Reviews Using Association Rules*. Department of Electrical and Computer Engineering, Sungkyunkwan University Suwon 440-746, Republic of Korea.
- [7] Lee Dongjoo, Jong Ok-Ran, Lee Sang-goo. *Opinion Mining of Customer Feedback Data on the Web*. School of Computer Science and Engineering, Seoul National University, Seoul.
- [8] Liddy, E. D. *In Encyclopedia of Library and Information Science*, 2nd Ed. Marcel Decker, Inc
- [9] Manning Christopher D., Schutze Hinrich. 1999. *Foundations of Statistical Nature Language Processing*. MIT Press, MA : Cambridge.
- [10] Meiwati Leni, Mustikasari Metty. *Aplikasi Data Mining Menggunakan Asosiasi Dengan Metode Apriori Untuk Analisis Keranjang Pasar Pada Data Transaksi Penjualan Apotek*. Jurusan Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma.
- [11] Ohana Bruno. 2009. *Opinion mining with the SentiWordNet lexical resource*. School of computing, Dublin Institute of Technology.
- [12] Turney Peter D. *Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL*. Institute for Information Technology, National Research Council of Canada, M-50 Montreal Road, Ottawa, Ontario, Canada, K1A 0R6.
- [13] Turney Peter D. *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews*. Institute for Information Technology, National Research Council of Canada, M-50 Montreal Road, Ottawa, Ontario, Canada, K1A 0R6.
- [14] Yusuf W. Yogi, Pratikto F. Rian, T. Gerry. 2006. *Penerapan Data Mining Dalam penentuan Aturan Asosiasi Antar Jenis Item*. Jurusan Teknik Industri, Universitas Katolik Parahyangan, Bandung.