

OPINION RETRIEVAL DENGAN PENDETEKSIAN KUALITAS REVIEW MENGUNAKAN SUPPORT VECTOR MACHINES (STUDI KASUS REVIEW PRODUK)

Muhammad Rizky Ramadhan¹, Ema Rachmawati², Arie Ardiyanti Suryani³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

¹Rizkygeblek@gmail.com

Abstrak

Media online seperti website saat ini berkembang sangat pesat di berbagai kalangan. Website juga dapat dimanfaatkan sebagai media untuk melakukan review dari produk yang ada di pasaran. Dari hal inilah, opinion retrieval dapat dijadikan cara untuk mencari opini-opini yang sesuai dengan kebutuhan user untuk mendapatkan feedback terkait dengan produk yang ingin mereka cari. Namun tidak semua review produk yang ada memiliki kualitas yang bagus, seperti deskripsi produk yang kurang lengkap, tidak spesifik, informasi yang disampaikan tidak akurat, atau tidak membahas fitur dari produk.

Pada Tugas Akhir ini, koleksi dokumen review produk terlebih dahulu diklasifikasikan kedalam suatu kategori atau kelas yaitu best review atau bad review, kemudian diimplementasikan opinion retrieval untuk menemukan opini sesuai inputan user. Metode yang digunakan dalam pengklasifikasian suatu review adalah Support Vector Machine (SVM) dengan Gaussian Kernel. SVM dipilih karena telah terbukti kehandalannya dalam melakukan klasifikasi terhadap data berupa teks yang memiliki dimensi yang tinggi. Hasil dari klasifikasi review tersebut dimasukkan kedalam pembobotan nilai untuk menghitung skor akhir dokumen dalam melakukan ranking.

Berdasarkan hasil pengujian, metode SVM dapat mengklasifikasikan suatu review dengan baik dan dengan rata-rata akurasi diatas 90%. Hasil perankingan dengan menambahkan skor klasifikasi review juga dapat diimplementasikan.

Kata Kunci : opinion retrieval, Support Vector Machine, SVM

Abstract

Online media such as website nowadays is growing rapidly in various circles. Website can also be used as a media to write a review of the product on the market. From that point, opinion retrieval can be used as a technique to get relevant opinions to obtain feedback related to the product that user want to search. But not all the existing reviews of products have good quality, such as incomplete product descriptions, not specific, the information submitted is not accurate, or does not discuss the features of the product.

In this final project, collections of product reviews are first classified into a category or class of best review or bad review, then opinion retrieval implemented to find the opinions matching to user's input. The method used in this review classification is Support Vector Machine (SVM) with Gaussian Kernel. SVM was chosen because its reliability has been proven in classifying high dimensional text. The result of that classification are computed into weighted score to calculate an overall score of the document for doing the ranking.

Based on test results, the SVM method can classify a review well and with an average accuracy above 90%. Ranking the results by adding the score of the review classification may also be implemented.

Keywords : opinion retrieval, Support Vector Machine, SVM

1. Pendahuluan

1.1 Latar Belakang Masalah

Dari tahun ke tahun, media *online* seperti website berkembang sangat pesat di berbagai kalangan. Berbagai macam orang dapat mencari berbagai sumber informasi yang dibutuhkan, baik itu informasi untuk mencari suatu peristiwa terbaru ataupun untuk keperluan pribadi. Website juga dapat dimanfaatkan sebagai media untuk melakukan *review* dari produk yang ada di pasaran. Hal ini sangat banyak dilakukan oleh orang yang memakai internet sebagai ajang untuk saling bertukar informasi dari produk yang telah mereka beli. *Review* produk yang telah dibuat oleh seorang *user* atau konsumen yang telah memakai produk tersebut dapat dijadikan sebagai bahan pertimbangan oleh calon konsumen yang lainnya. Dari hal inilah, *opinion retrieval* dapat dijadikan cara untuk mencari opini-opini yang sesuai dengan kebutuhan *user* untuk mendapatkan informasi atau opini terkait dengan produk yang ingin mereka cari.

Opinion retrieval merupakan sebuah cara untuk mendapatkan opini-opini yang ingin dicari oleh *user* dari beberapa dokumen yang ada. Dalam studi kasus *review* produk ini, *opinion retrieval* digunakan oleh *user* untuk mencari opini yang ada dalam berbagai *review* produk. Namun, yang menjadi masalah adalah tidak semua *review* produk yang ada memiliki kualitas yang baik. Maksud dari kualitas baik atau kurang baik disini bukanlah terletak pada kualitas produknya, tetapi pada isi dari *review*-nya, seperti kelengkapan komentar terhadap produk, pembahasan fitur produk, bukti pendukung, atau pemberian rekomendasi. *Review* yang memiliki kualitas sangat baik tentu akan memiliki nilai lebih dalam melakukan *retrieve* opini. Melihat masalah ini, selain menerapkan metode untuk melakukan *opinion retrieval*, maka pengklasifikasian suatu *review* produk juga perlu dilakukan dengan metode *Support Vector Machines* (SVM) agar hasil opini yang didapat sesuai dengan keinginan *user*. Dengan menggunakan pendekatan berbasis klasifikasi tersebut, maka akan didapatkan hasil tingkatan kualitas sebuah *review*, dari kualitas yang terburuk sampai kualitas yang terbaik.

1.2 Perumusan Masalah

Permasalahan yang akan dibahas dalam Tugas Akhir ini antara lain:

1. Bagaimana cara membangun sebuah sistem *opinion retrieval* yang dapat mengambil opini-opini yang tidak hanya dari *review* yang relevan saja, namun juga menjadikan suatu kualitas *review* sebagai faktor dalam melakukan ranking?
2. Seberapa besar tingkat akurasi hasil klasifikasi *review* dan perankingan yang dihasilkan oleh sistem?

1.3 Batasan Masalah

Adapun batasan masalah dalam pengerjaan Tugas Akhir ini diantaranya adalah sebagai berikut:

1. Dokumen yang akan dianalisis adalah dokumen berbahasa Inggris.
2. Data yang dijadikan sebagai bahan analisis yaitu *review* produk laptop dari situs <http://www.amazon.com/> yang telah di-*crawl* dan disimpan dalam *database*.
3. Penelitian ini berfokus pada pengklasifikasian *review* ke dalam kelas “*best review*” atau “*bad review*” yang akan digunakan untuk perankingan dokumen, serta bagaimana cara melakukan proses *retrieve* opini dari *review* yang telah terklasifikasi.
4. Semua proses dilakukan secara *offline* tanpa koneksi internet.

1.4 Tujuan

Tujuan dari Tugas Akhir ini adalah:

1. Mengimplementasikan suatu metode pada sistem *opinion retrieval* yang dapat mengambil opini dari kumpulan *review* yang relevan dan juga memiliki kualitas *review* terbaik.
2. Melakukan pengukuran dan analisis akurasi dari hasil klasifikasi suatu *review* dan perankingan yang dilakukan oleh sistem.

1.5 Metodologi Penyelesaian Masalah

Penyelesaian masalah dilakukan dalam beberapa tahap, secara garis besar adalah sebagai berikut.

1. Identifikasi masalah

Pada tahap ini dilakukan pengidentifikasian beberapa fenomena yang terjadi pada sistem pengklasifikasian teks menggunakan metode *Support Vector Machine* (SVM).

2. Studi literatur

Mempelajari literature-literatur yang berkaitan dengan metode yang digunakan, seperti permasalahan, konsep, dan teknik-teknik yang ada pada klasifikasi teks menggunakan SVM, proses *information retrieval* serta *opinion retrieval* untuk memperoleh opini.

3. Analisis dan perancangan kebutuhan sistem

Tahap ini menganalisa sistem yang dibutuhkan beserta semua komponen-komponen pendukung serta mengumpulkan data. Pengumpulan data *review* produk dari *website* <http://www.amazon.com/> yang akan digunakan sebagai dataset, data kata *sentiment* yang akan dijadikan sebagai kamus *lexicon*, serta data kata *stopword* untuk digunakan sebagai *stopword list*. Data yang terkumpul kemudian digunakan untuk merancang sistem berupa perangkat lunak *opinion retrieval* dengan pendeteksian kualitas menggunakan SVM.

4. Implementasi sistem

Pada tahap ini rancangan sistem yang sudah dibuat sebelumnya akan diimplementasikan ke dalam bentuk aplikasi.

5. Pengujian dan analisis sistem

Sistem yang telah dibuat kemudian akan diuji pada tahap ini untuk mengukur akurasi hasil klasifikasi serta akurasi performansi keseluruhan sistem dan melakukan analisa terhadap hasil yang diberikan.

6. Pengambilan kesimpulan dan pembuatan laporan

Di tahap ini dilakukan pengambilan kesimpulan sesuai dengan analisa yang telah dilakukan dan pembuatan laporan Tugas Akhir beserta pengumpulan beberapa dokumentasi yang diperlukan terkait referensi, proses, maupun hasil penelitian yang dilakukan.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan analisis terhadap pengujian yang dilakukan dalam Tugas Akhir ini, dapat disimpulkan bahwa:

1. Perubahan nilai parameter toleransi tidak terlalu berpengaruh secara signifikan karena secara umum dokumen dapat diklasifikasikan dengan baik, sedangkan nilai parameter γ berpengaruh terhadap akurasi yang dihasilkan karena menentukan *decision boundary*. Untuk data yang ada pada sistem ini, parameter toleransi terbaik berada di nilai 0.99 dan parameter γ dengan nilai 0.125.
2. Hasil perangkingan dokumen dengan ditambahkan faktor kualitas sebuah *review* dapat diimplementasikan dan memiliki hasil yang lebih baik dibandingkan dengan perangkingan tanpa memasukkan faktor kualitas *review*.

5.2 Saran

Saran yang diperlukan dari Tugas Akhir ini untuk pengembangan sistem lebih lanjut adalah sebagai berikut:

1. Memberikan label data pada *training* dan *testing* dengan bantuan *expert judgment* sehingga model yang dihasilkan semakin akurat dan hasil klasifikasi lebih teruji validitasnya.
2. Sistem dapat mengambil data secara *online*.

6. Daftar Pustaka

- [1] Andrea Esuli. “Amazon Parser and Downloader” <http://www.esuli.it/software/amazon-reviews-downloader-and-parser/> (diakses tanggal 3 Februari 2013).
- [2] Guduru, Neelima. (2006). *Text Mining with Support Vector Machine and Non Negative Matrix Factorization Algorithms*. University of Rhode Island.
- [3] Hsu, Chin Wei (et. al.). (2010). *A Practical Guide to Support vektor Classification*. Taiwan: Department of Computer Science National Taiwan University.
- [4] Hu, Mingqing, Bing Liu. (2004). *Mining and Summarizing Customer Reviews*. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004).
- [5] Joachims, Thorsten. *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*. Universitat Dortmund.
- [6] Lewis, David D. (2004). *RCV1: A New Benchmark Collection for Text Categorization Research*. *Journal of Machine Learning Research*.
- [7] Liao, X., Cao, D., Tan, S., Liu, Y., Ding, G., and Cheng X. *Combining Language Model with Sentiment Analysis for Opinion Retrieval of Blog-Post*. Online Proceedings of Text Retrieval Conference (TREC) 2006. <http://trec.nist.gov/>.
- [8] Liu, Jingjing, Yalou Huang. (2007). *Low-Quality Product Review Detection in Opinion Summarization*. Tianjin, China: Nankai University.
- [9] Manning, Christopher D. (2008). *Introduction to Information Retrieval*. New York: Cambridge Univ. Press.
- [10] Nugroho Satriyo Anto, Arief Budi Witarto, dan Dwi Handoko. (2003) “Support Vector Machine-Teori dan Aplikasinya dalam Bioinformatika” dari <http://ilmukomputer.com>.
- [11] Nugroho. A. N. (2008). “*Support Vector Machine: Paradigma Baru dalam Softcomputing*”. KNSI 2008; Bali.

- [12] Nur, Muhammad Yusuf, dan Santika, Diaz D. (2011). *Analisis Sentimen pada Dokumen Berbahasa Indonesia dengan pendekatan Support Vector Machine*. Jakarta.
- [13] Pratiwi, A. Y. (2010). *Implementasi Model Ruang Vektor Pada Sistem Temu Balik Informasi Dokumen Berbahasa Indonesia*. Bandung: Universitas Pendidikan Indonesia.
- [14] Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-wesley, Reading, Pennsylvania.
- [15] Smirnov, Ilia. (2008). *Overview of Stemming Algorithms*. Chicago: DePaul University.
- [16] Susanto, S. (2006). *Pengklasifikasian Artikel Berita Berbahasa Indonesia Secara Otomatis Menggunakan Naive Bayes Classifier*. Depok: Universitas Indonesia.
- [17] Wei Zhang, Clement Yu, Weiyi Meng. *Opinion Retrieval from Blogs*, Department of Computer Science Binghamton University.
- [18] Zhang, Lei, dkk. (2011). *Identifying Noun Product Features that Imply Opinions*. Chicago: University of Illinois.