

SUBJECTIVE TRIGGER MODEL UNTUK OPINION RETRIEVAL BERBAHASA INDONESIA

Rahmi Salendri¹, Warih Maharani², Ade Romadhony³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Opini merupakan pikiran, pendirian atau pandangan pribadi individu yang bersifat subjektif. Opini dapat membantu seseorang dalam pengambilan keputusan, memberikan feedback yang berguna bagi suatu perusahaan dan juga membantu dalam melakukan pemantauan terhadap peminatan suatu produk untuk melihat seberapa besar ketertarikan user terhadap suatu produk tertentu. Oleh karena itu dibutuhkan suatu sistem yang dapat menghasilkan informasi berupa opini didalam suatu dokumen dengan topik atau query tertentu yang dinamakan Opinion Retrieval. Tugas Akhir ini bertujuan untuk menemukan dokumen beropini subjektif dengan menggunakan Subjective Trigger Model. Sistem ini mencari suatu trigger pair yang terdiri atas triggering word dan triggered word untuk mengidentifikasi kata opini subjektif pada dokumen relevan hasil dari Information Retrieval. Sistem ini akan menghasilkan dokumen yang relevan dengan query dan mengandung opini. Untuk evaluasi sistem, digunakan Mean Opinion Score (MOS) untuk mengukur relevansi hasil yang dikeluarkan oleh sistem dan didapatkan hasil MOS sebesar 4,08 dari 5.

Kata Kunci : Opinion Retrieval, Trigger Pair, Subjective Trigger Model

Abstract

Opinion is a thought, conviction, or one person's view subjectively. Opinion can help someone in decision making, giving useful feedback for a company, and help in monitoring product interest to see user's interest in the product. In that matter, a system which producing information as opinion from a document is needed to find subjective opiated document using Subjective Trigger Model. This system will find a trigger pair which consist of triggering and triggered word to identification subjective opinion word in relevant document from Information Retrieval process. This system will produce document which is relevant to the query and opiated. To evaluate the system, Mean Opinion Score (MOS) used to calculating the relevancy of the result, the result of the Mean Opinion Score is 4,08 out of 5.

Keywords : Opinion Retrieval, Trigger Pair, Subjective Trigger Model

Telkom
University

1. PENDAHULUAN

1.1 Latar Belakang

Opini merupakan pendapat, pikiran, pendirian atau pandangan pribadi individu yang bersifat subjektif [3]. Opini dapat membantu seseorang dalam pengambilan keputusan, memberikan *feedback* yang berguna bagi suatu perusahaan dan juga membantu dalam melakukan pemantauan terhadap peminatan suatu produk untuk melihat seberapa besar ketertarikan *user* terhadap suatu produk tertentu. Salah satu media untuk mengekspresikan opini adalah forum, yang didalamnya terdapat berbagai macam topik opini seperti masalah politik, organisasi, hobi, *review* produk dan lain sebagainya. Oleh karena itu dibutuhkan suatu sistem yang dapat menghasilkan informasi berupa opini di dalam suatu dokumen dengan topik atau *query* tertentu yang dinamakan *Opinion Retrieval*. *Opinion Retrieval* merupakan proses melakukan *pe-retrieve-an* opini pada dokumen yang relevan berdasarkan opini mengenai sebuah *query* yang terdapat dalam dokumen tersebut. Dokumen beropini yang relevan memenuhi 2 buah kriteria yaitu relevan dengan *query* dan mengandung opini[21]. Dalam menemukan kata opini subjektif di dalam dokumen, salah satu metode yang pernah digunakan yaitu metode pemodelan bahasa menggunakan *bigram*. Hanya saja jika menggunakan *bigram*, tidak dapat menangani kalimat yang memiliki ketergantungan kata dengan jarak yang jauh[12]. Oleh karena itu untuk menangani kasus tersebut, Lau et al mengajukan suatu metode yang dinamakan *Trigger Model* [20]. *Trigger Model* ini menggunakan *trigger pair* yang berupa *triggering word* dan *triggered word*, untuk mengidentifikasi kata opini subjektif.

Tugas akhir ini berfokus pada *Opinion Retrieval* berbahasa Indonesia. Ada dua asumsi untuk mengetahui dokumen beropini subjektif yaitu asumsi pertama, kalimat mengandung subjek atau objek opini dan ekspresi subjektif dan asumsi kedua yaitu *triggering word* berupa kata ganti (pronomina)[9,13]. Untuk mengetahui dokumen beropini subjektif tersebut, maka digunakan metode yang disebut *Subjective Trigger Model (STM)*. *Subjective Trigger Model* merupakan perpaduan interpolasi linear antara *bigram* dan *Trigger Model*[12]. Dengan menggunakan metode ini, dapat memperbaiki model bahasa *bigram* yaitu dapat menangani kalimat yang memiliki ketergantungan kata dengan jarak yang jauh dan dengan menggunakan *trigger pair* sehingga dapat memudahkan pencarian opini subjektif di dalam suatu dokumen. Dokumen yang dicari merupakan dokumen yang relevan dengan *query* dan memiliki opini subjektif. Dokumen yang mengandung opini subjektif akan ditentukan berdasarkan *trigger pair* dengan *triggering word*nya berupa kata ganti orang (pronomina persona) dan *triggered word*nya berupa ekspresi subjektif[13]. Ekspresi subjektif adalah kata yang dipicu oleh suatu pronomina sehingga kalimat tersebut akan menjadi kalimat beropini.

1.2 Rumusan Masalah

Berdasarkan pada latar belakang di atas, permasalahan yang akan diuraikan dan diteliti adalah:

1. bagaimana proses pengambilan kata beropini subjektif dalam dokumen yang relevan dengan *query* pada dokumen berbahasa Indonesia?

2. bagaimana mengukur akurasi sistem dengan menerapkan *Subjective Trigger Model* pada *Opinion Retrieval*?
3. faktor-faktor apa saja yang mempengaruhi hasil perangkingan dokumen yang ditampilkan oleh sistem?

1.3 Batasan Masalah

Adapun batasan masalah untuk tugas akhir ini adalah:

1. Dokumen yang akan dianalisis terbatas pada dokumen teks bahasa Indonesia yang memiliki ejaan yang benar.
2. Data yang digunakan sebagai bahan analisis yaitu dokumen yang berasal <http://students.itelkom.ac.id/web/forum.php> dan www.tabloidpulsa.co.id.
3. Data akan dianalisis secara *offline learning*, yaitu sistem tidak terhubung ke internet.
4. *Triggering word* yang digunakan didalam sistem hanya pronomina persona tanpa imbuhan yaitu aku, saya, daku, kami, kita, engkau, kamu, kau, anda, dikau, kalian, anda sekalian, kamu sekalian, ia, dia, beliau, mereka[10].

1.4 Tujuan

Tujuan yang ingin dicapai dalam pengerjaan Tugas Akhir ini adalah sebagai berikut:

1. menerapkan *Subjective Trigger Model* dalam dokumen berbahasa Indonesia pada sistem *Opinion Retrieval*;
2. menganalisis dan mengevaluasi hasil akurasi sistem dari penerapan *Subjective Trigger Model* pada *Opinion Retrieval* dengan menggunakan *Mean Opinion Score* (MOS);
3. menganalisis faktor-faktor yang mempengaruhi hasil perangkingan dokumen oleh sistem.

1.5 Metodologi Penyelesaian Masalah

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

1. Tahap studi literatur
Mencari informasi dan mempelajari literatur-literatur yang relevan dengan permasalahan *Information Retrieval* khususnya *Vector Space Model*, model bahasa *bigram*, konsep dan teknik-teknik *Opinion Retrieval* khususnya *Subjective Trigger Model* serta sumber-sumber lain yang relevan untuk menunjang penyelesaian Tugas Akhir ini. Sumber-sumber pustaka didapat dari buku, paper, maupun halaman web.
2. Tahap pengumpulan dataset
Pada tahap ini dilakukan pengumpulan data berupa teks-teks dalam bahasa Indonesia. Teks tersebut berada dalam sebuah dokumen, yang nantinya dokumen tersebut akan digunakan untuk uji coba.
3. Tahap perancangan sistem
Pada tahap ini membuat rancangan sistem dengan metode yang telah dipilih, dimulai dari rancangan untuk melakukan proses *pre-processing* dokumen,

pencarian dokumen yang relevan dengan *query*, dan menghitung bobot dokumen beropini subjektif yang terurut yang dihasilkan oleh sistem.

4. Tahap implementasi

Pada tahap ini akan diterapkan hasil perancangan sistem pada data yang telah dikumpulkan.

5. Tahap analisis

Pada tahap ini dilakukan analisa terhadap hasil penelitian yang telah dilakukan dalam penyelesaian masalah.

6. Pengambilan kesimpulan dan pembuatan laporan

Pada tahap ini dilakukan pengambilan kesimpulan terhadap analisa yang telah dilakukan dan pembuatan laporan Tugas Akhir untuk mendokumentasikan tahap-tahap penelitian yang telah dilakukan sampai terselesaikannya masalah pada Tugas Akhir ini.



5. Kesimpulan dan Saran

5.1 Kesimpulan

1. Panjang kata tidak mempengaruhi banyak atau sedikit ter-retrieve-nya *trigger pair*. Banyak ter-retrieve-nya *trigger pair* dipengaruhi oleh adanya kata pronomina (kata ganti persona) didalam suatu dokumen, dapat dilihat pada tabel 4-3.
2. Semakin besar nilai *threshold*, semakin banyak pula dokumen yang akan ter-retrieve. Hal ini dikarenakan semakin besar nilai *threshold*, sehingga bobot dari *bigram* yang semakin kecil membuat pasangan kata *bigram* akan ter-retrieve yang menyebabkan dokumen tersebut ter-retrieve pula, dapat dilihat pada tabel 4-5.
3. Semakin sedikit kemunculan *triggering word* di dalam dokumen, maka semakin besar bobot dari STM, begitu pula sebaliknya. Dikarenakan pembagi dalam perhitungan bobot *trigger model* adalah frekuensi kemunculan *triggering word* didalam semua dokumen, oleh karena itu semakin sering *triggering word* muncul didalam suatu dokumen maka akan semakin kecil pula bobot *trigger model*, begitu pula sebaliknya dan hal ini akan mempengaruhi nilai dari bobot STM. Sehingga banyaknya *trigger pair* didalam suatu dokumen, tidak menentukan bahwa dokumen tersebut memiliki bobot yang besar, dapat dilihat pada tabel 4-7.
4. Hasil rata-rata *Mean Opinion Score* (MOS) sudah cukup baik, yaitu sebesar 4.079 dari 5 dengan tiga aspek penilaian, yaitu relevansi dokumen, bobot rangking (baik IR maupun OR) serta kerelevanan dokumen beropini berdasarkan *trigger pair*.

5.2 Saran

1. Untuk melihat pengukuran akurasi sistem, sebaiknya menggunakan *relevance judgement*, tidak perlu lagi menggunakan MOS dalam mengukur perfomansi sistem hal ini cukup menyita waktu.
2. Untuk kedepannya, *triggering word*nya dapat diperluas lagi, dengan cara menambahkan pronomina persona dengan imbuhan seperti “-ku”, “-mu”, dan “-nya” agar hasil pencarian dokumen beropini subjektif lebih baik lagi.

6. Referensi

- [1] Adi Wibowo, Andreas Handojo, Minardi Taliwang, "Recommender System di Perpustakaan Universitas Kristen Petra menggunakan Rocchio Relevance Feedback dan Cosine Similarity". Universitas Kristen Petra.
- [2] Ahmad Najibullah, 2011, "Implementasi N-Gram Dalam Pencarian Teks Sebagai Penunjang Aplikasi Perpustakaan Kitab Berbahasa Arab". Institut Teknologi Sepuluh November.
- [3] Buku Bahasa Indonesia SMP kelas Vii, <http://goo.gl/tcb11I>, didownload pada tanggal 02 Maret 2013.
- [4] C. D. Manning, H. Schutze, 1999, "Foundation of Statistical Natural Language Processing". The MIT Press.
- [5] C. D. Manning, Prabhakar Raghavan, Hinrich Schutze, 2009, "An Introduction to Information Retrieval". Cambridge University Press.
- [6] C. Tillmann, H. Ney, 1996, " Selection Criteria for Word Trigger Pairs in Language Modelling". University of Technology Germany.
- [7] Chih-Heng Ke, Ce-Kuen Shieh, Wen-Shyang Hwang, Artur Ziviani, 2008, "An Evaluation Framework for More Realistic Simulations of MPEG Video Transmission". Journal of Information Science and Engineering.
- [8] Dyan Keke, Rian Chikita, Agus Dwi Prayogo, 2012, "Sistem Temu Balik Informasi Algoritma Nazief dan Andriani". Universitas Gadjah Mada.
- [9] G. Zhou, H. Joshi, and C. Bayrak, 2007, "Topic Categorization for Relevancy and Opinion Detection". Computer Science Department.
- [10] Hasan Alwi, dkk, 2003, "Tata Bahasa Baku Bahasa Indonesia". Edisi Ketiga. Jakarta: Balai Pustaka.
- [11] Indra Lukmana, Agus Zainal Arifin, Diana Purwitasari, 2012, " Perangkingan Dokumen Berbahasa Arab Berdasarkan Susunan Posisi Kata dari Query". Teknik Informatika, Institut Teknologi Sepuluh November.
- [12] Kazuhiro Seki, Kuniaki Uehara, 2009, "Adaptive Subjective Triggers for Opinionated Document Retrieval". Organization of Advanced Science Technology Kobe University.
- [13] Kiduk Yang, Ning Yu, Hui Zhang, 2007, "WIDIT in TREC 2007 Blog Track: Combining Lexicon-based Methods to Detect Opinionated Blogs". School of Library and Information Science.
- [14] Metode N-gram, <http://goo.gl/ZoZKce>, didownload pada tanggal 29 November 2012.
- [15] Michael Collins, 2008, "Advanced NLP: Lecture 2-Language Models".
- [16] Min Zhang, Xingyao Ye, 2008, "A Generation Model to Unify Topic Relevance and Lexicon-Based Sentiment for Opinion Retrieval". Tshinghua University.
- [17] Mingjie Zhu, Shuming Shi, Mingjing Li, Ji-Rong Wen, 2008, "Effective top-k Computation with Term Proximity Support". Univesity of Science and Technology of China.
- [18] Novian Anggis Suwastika, 2009, "Analisis dan Implementasi Vector Space Model Pada Information Retrieval". Institut Teknologi Telkom.

- [19] Raymond J. Mooney, 2001, "Natural Language Processing: N-Gram Language Models". University of Texas.
- [20] Raymond Lau, Ronald Rosenfeld, Salom Roukos, 1993, "Trigger-based Language Models: a Maximum Entropy Approach". Massachusetts Institute of Technology.
- [21] Wei Zhang, Clement Yu, Weiyi Meng, 2007, "Opinion Retrieval from Blogs". Department of Computer Science Binghamton University.

