

RELEVANCE FEEDBACK PADA INFORMATION RETRIEVAL DENGAN SUPPORT VECTOR MACHINE

Sri Ulinar Romatua N B¹, Yanuar Firdaus A.w.², Warih Maharani³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Dengan semakin banyaknya informasi yang berkembang saat ini, search engine sebagai salah satu alat untuk mencari informasi di tuntut ketepatannya dalam menemukan kembali informasi tersebut. Seringkali informasi yang didapatkan belum tentu sesuai dengan keinginan user atau si pencari informasi. Relevance feedback adalah salah satu teknik yang dapat membantu meningkatkan kerelevanan informasi dengan keinginan user. Penerapannya pada Information Retrieval System dapat dilakukan dengan banyak metode. Support Vector Machine merupakan salahsatu yang dapat digunakan.

Dengan Support Vector Machine, performansi Information Retrieval System meningkat pada jumlah dokumen feedback tertentu. Dimana dokumen yang menjadi dokumen feedback merupakan sejumlah dokumen relevan teratas yang ditemukan sistem. Dari penerapan metode ini peningkatan terbaik performansi perangkat lunak mencapai 75.60% untuk precision, 60.3 % untuk recall dan IAP 35.83%.

Kata Kunci : information retrieval system, relevance feedback, support vector

Abstract

With consideration of rapid information develop nowadays, search engine is consider as a tools for searching information and that information needs to be exact so it can retrieve the right information. Sometimes the information retrieve is not the exact information that the user wants. One of the techniques to help increase the relevance level for the information is called "Relevance Feedback". The implication of the Information Retrieval System can be done implement in many different ways and Support Vector Machine is one of the method.

With Support Vector Machine, the performance of Inforation Retrieval System increase in the total number of certain feedback. Whereas document can become a document feedback which consider to be top level documents that can be found in a system. From the implementation of this method, the best increasing level at this stage can reach 75.60% at precision level, 60.3% for recall and 35.83% for IAP.

Keywords : information retrieval system, relevance feedback, support vector

Telkom
University

1. PENDAHULUAN

1.1 Latar belakang

Informasi merupakan salah satu hal penting yang dibutuhkan dalam segala aspek kehidupan kita. Sangat banyak cara untuk menemukan informasi pada saat sekarang ini. Pencarian informasi dapat dilakukan secara manual maupun secara otomatis. Pencarian secara manual dapat kita lakukan dengan mencari informasi pada gudang informasi seperti majalah, buku, koran dan yang lainnya dengan manual, sedangkan secara otomatis kita dapat memanfaatkan suatu alat bantu yang dapat membantu kita melakukan pencarian tersebut.

Search engine merupakan sebuah alat berbasis teknologi informasi yang dapat membantu kita dalam pencarian informasi. Pencarian dengan *search engine* ini dilakukan pada gudang data atau sering disebut dengan *document collection* yang jumlahnya sangat besar. Dan dengan bantuan *search engine* ini informasi dapat dengan lebih mudah ditemukan dibandingkan bila kita harus mencari sendiri informasi tersebut.

Sistem temu kembali informasi (*Information Retrieval System*) merupakan sistem yang digunakan untuk mendapatkan atau menemukan kembali informasi atau dokumen-dokumen yang sesuai dari gudang data atau *document collection*. Sistem ini menggunakan kata kunci (*query*) yang diberikan oleh pengguna (*user*) untuk melakukan pencarian pada *document collection*. Dengan sistem temu kembali (*IRS*), pengguna (*user*) mendapatkan informasi yang sesuai dengan kebutuhannya. Pada saat *user* ingin memasukkan kata kunci tersebut sering kali kata kunci yang dimasukkan ambigu dan kadang memiliki arti yang berbeda dengan yang dimaksud. Dengan begitu maka informasi yang diberikan kepada *user* belum tentu sesuai/relevan dengan apa yang dibutuhkan oleh *user*. Untuk lebih meningkatkan ketepatan informasi yang dibutuhkan oleh *user* dapat digunakan *Relevance Feedback*.

Relevance feedback merupakan teknik dimana pengguna melakukan umpan balik terhadap hasil pencarian informasi sebelumnya. Dengan teknik ini, *user* menilai dan memilih beberapa dokumen yang sesuai/relevan dengan kebutuhannya. Kemudian informasi yang telah dipilih tersebut akan digunakan kembali oleh sistem untuk mencari dokumen yang lebih sesuai/relevan dari hasil pencarian sebelumnya. Berdasarkan *feedback* dari *user*, sistem kemudian melakukan formulasi ulang *query* yang digunakan pada proses sebelumnya.

Relevance Feedback merupakan salah satu cara untuk meningkatkan performansi sistem temu kembali (*IRS*) ini. Dengan *relevance feedback* kita dapat meningkatkan tingkat kesesuaian hasil dari sistem dengan keinginan dari *user*. Sehingga waktu yang dibutuhkan oleh *user* untuk mendapatkan informasi yang sesuai/relevan lebih singkat. Terdapat beberapa model untuk menerapkan *relevance feedback*, seperti model probabilistik dan model vektor. Untuk model probabilistik, ada beberapa metode seperti *Bayesian Network* dan *Robertson Spark Jones* sementara untuk model vektor ada *Rocchio*, *Ide dec Hi*, dan *Ide Regular*. Metode lain yang dapat digunakan adalah metode atau algoritma klasifikasi yang sudah ada. *Support Vector Machine* merupakan salah satunya.

Support Vector Machine (SVM) merupakan salah satu algoritma untuk *pattern recognition*, dimana memiliki salah satu ciri khusus yaitu, mengklasifikasikan 2

buah *class*. Ini merupakan pertimbangan utama mengapa digunakannya *SVM*. Pada permasalahan ini *SVM* digunakan untuk mengklasifikasi antara dokumen yang relevan dan tidak relevan. Dengan klasifikasi ini diharapkan informasi atau dokumen yang dikembalikan kepada user menjadi lebih relevan, dan menjauhkan dokumen atau informasi yang tidak relevan. Salah satu juga hal yang menjadi pertimbangan kenapa dipilih metode *SVM* ini dikarenakan parameter yang nantinya akan digunakan dalam proses klasifikasi adalah term, dimana term ini memang unsur utama yang mewakili setiap dokumen. Dan pada banyak riset yang dilakukan untuk kasus klasifikasi atau dalam pattern recognition, *SVM* memperlihatkan performansi yang sangat baik.

1.2 Perumusan masalah

Permasalahan yang dijadikan sebagai objek penelitian pada tugas akhir ini adalah:

1. Bagaimana menerapkan *relevance feedback* dengan *Support Vector Machine* pada *information retrieval*.
2. Bagaimana pengaruh penerapan *relevance feedback* dengan *Support Vector Machine* pada pencarian informasi yang relevan dengan keinginan user.
3. Bagaimana menganalisis performansi dari *relevance feedback* dengan *Support Vector Machine* pada *information retrieval*.

Batasan masalah dari tugas akhir ini adalah:

1. Koleksi dokumen dan *query* yang digunakan adalah dokumen berbentuk teks dalam bahasa Inggris.
2. Koleksi dokumen yang menjadi objek pencarian ialah koleksi dokumen yang diambil dari <ftp://ftp.cs.cornel.co.uk/smart>.
3. Sistem temu kembali yang dibangun berbasis *desktop*.
4. Mekanisme penggunaan *relevance feedback* yang digunakan adalah secara manual atau secara *explicit feedback* dimana user akan secara nyata melakukan *marking* mana dokumen yang relevan mana yang tidak, terhadap dokumen hasil pencarian.

1.3 Tujuan

Tujuan pembuatan tugas akhir ini adalah:

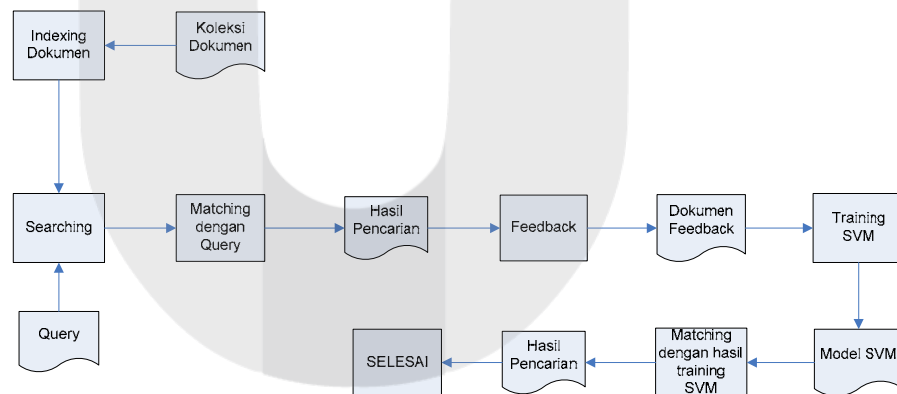
1. Merancang dan membangun perangkat lunak untuk menerapkan *relevance feedback* dengan *Support Vector Machine* pada *information retrieval*.
2. Menganalisis performansi terhadap hasil dari penerapan *relevance feedback* dengan *Support Vector Machine*. Performansi dihitung berdasarkan nilai *precision*, *recall*, *Interpolated Average Precision (IAP)*.

1.4 Metodologi penyelesaian masalah

Metodologi penyelesaian masalah yang digunakan dalam menyelesaikan penelitian ini adalah:

1. Identifikasi Masalah
Pada langkah ini akan dilakukan pengidentifikasian permasalahan apa saja yang timbul. Hal apa saja yang akan dilakukan dan mungkin terjadi pada penerapan *Support Vector Machine* pada *relevance feedback*.

2. Studi literatur
Langkah ini bertujuan untuk memahami dasar teori mengenai *information retrieval*, *relevance feedback*, dan *Support Vector Machine* serta hal lain yang mendukung penyelesaian tugas akhir ini. Sumber dasar teori dapat berupa buku, *paper*, maupun halaman web.
3. Analisis kebutuhan perangkat lunak
Yaitu melakukan analisis kebutuhan perangkat lunak yang akan dibangun, agar didapatkan gambaran umum seperti apa perangkat lunak yang ingin dibangun.
4. Perancangan perangkat lunak
Berdasarkan kebutuhan yang telah diidentifikasi, maka dapat dirancang perangkat lunak yang sesuai untuk memenuhi kebutuhan. Rancangan perangkat lunak dapat menjadi panduan saat implementasi perangkat lunak.
5. Implementasi
Pada tahapan ini dilakukan pembangunan perangkat lunak yang telah dirancang dengan menggunakan teknik pemrograman tertentu. Pada tahapan ini dibangun sistem yang menerapkan *Support Vector Machine*. Dimana akan dilakukan penerapan SVM yang dapat mengakomodir proses klasifikasi dokumen. Berikut merupakan gambaran proses yang akan diimplementasikan.



Gambar 1-1 Gambaran proses yang akan diimplementasikan.

6. Pengujian dan analisis hasil.
Pengujian dilakukan untuk memperhatikan pengaruh *relevance feedback* terhadap *information retrieval*. Pada proses ini akan dibandingkan nilai *recall*, *precision* dan *Interpolated Average Precision (IAP)*.
7. Penyusunan laporan tugas akhir.

5. PENUTUP

Pada bab ini akan diuraikan hal yang dapat disimpulkan dari pelaksanaan Tugas Akhir ini. Selain itu diuraikan pula beberapa saran yang dapat digunakan dalam pengembangan Tugas Akhir ini.

5.1 Kesimpulan

Berdasarkan hasil analisis dan pengujian perangkat lunak yang dilakukan dalam Tugas Akhir ini dapat diambil beberapa kesimpulan yaitu :

1. *Support Vector Machine* dapat diterapkan untuk *relevance feedback* pada *information retrieval system*.
2. Penggunaan *relevance feedback* dengan metode *Support Vector Machine* tidak selalu bisa dapat meningkatkan performansi *information retrieval system*.
3. Penambahan jumlah dokumen yang dijadikan sebagai *feedback* akan memberikan performansi yang berbeda dibandingkan dengan pencarian awal.
4. Untuk nilai *recall* dari kedua koleksi dokumen dapat diambil pola bahwa dengan penerapan *Support Vector Machine* pada *relevance feedback* ini akan memberikan penurunan nilai *recall* pada *feedback* tertentu tetapi akan kembali naik lagi pada dokumen *feedback* tertentu. Ini dikarenakan *query* baru yang dihasilkan pada *feedback* memberikan term tambahan yang menyebabkan dokumen relevan yang terambil oleh sistem semakin kecil.
5. Nilai *precision* yang dihasilkan dari penerapan *Support Vector Machine* ini cenderung meningkat walaupun ada kondisi dimana nilai *precision* akan turun. Nilai *precision* yang didapat juga tidak mendekati nilai maksimal, yaitu 1 dikarenakan jumlah dokumen pada koleksi dokumen sangat banyak sementara dokumen relevannya hanya sedikit. Misalnya saja untuk *query* pertama pada dokumen ADI, jumlah dokumen relevannya hanya 3 dokumen, sementara dokumen yang terambil oleh sistem dari koleksi dokumen adalah bisa mencapai 40 sampai 50 dokumen.
6. Nilai *IAP* dari hasil penerapan *relevance feedback* dengan *Support Vector Machine* ini mengalami kenaikan. Dapat disimpulkan bahwa semakin banyak dokumen yang kita *feedback*, maka akan semakin naik juga nilai *IAP* nya, berarti dokumen relevan yang terambil oleh sistem urutannya menjadi lebih baik dibandingkan dengan pencarian awalnya. Pembentukan *query* baru merupakan salah satu faktor terjadinya kenaikan *IAP*, dengan *query* baru yang dibangun maka akan memberi peluang dokumen yang relevan akan muncul terurut. Walaupun tidak semua dokumen bisa berada terurut pada rangking teratas.

5.2 Saran

Untuk pengembangan Tugas Akhir di masa mendatang, penulis menyarankan hal-hal sebagai berikut:

1. Penambahan jenis *file* dokumen yang dapat dicari agar tidak hanya (.txt) saja.
2. Perangkat lunak sebaiknya mampu menangani jika pengguna melakukan beberapa kali *feedback*.



Daftar Pustaka

| | |
|------|---|
| [1] | Berry, Michael W. and Murray Browne. Understanding Search Engine : <i>Mathematical Modeling and Text Retrieval</i> . SIAM. Philadelphia. 1991 |
| [2] | Manning, Christopher D. (et.al.), 2008, <i>Introduction to Information Retrieval</i> , USA:Cambridge University Press. http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf Didownload pada tanggal 08 Maret 2009 |
| [3] | Ziegler, W., 2008, <i>Information Retrieval: Relevance Feedback system for text documents</i> . http://more.zites.net/morefiles/public%20system%20documentation.pdf Didownload pada tanggal 08 Maret 2009. |
| [4] | Zhao Xu, Xiaowei Xu, Kai Yu and Volker Tresp. Journal : A Hybrid Relevance Feedback Approach to Text Retrieval. http://ifsc.ualr.edu/xwxu/publications/ecir03_hybrid.pdf Di download pada tanggal 08 Maret 2009 |
| [5] | Nugroho, Anto. S.,Witarto, Arif. B., Handoko, Dwi., 2003, Support Vector Machine-Teori dan aplikasinya dalam Bioinformatika, http://ilmokomputer.com/ikcsvm.pdf Di download pada tanggal 08 Maret 2009 |
| [6] | Karls, E., 2007, <i>Information Retrieval : Relevance Feedback and Query Expansion</i> . Seminar International Studies in Computational Linguistics, Tuebingen University. http://www.sfs.uni-tuebingen.de/~parmenti/slides/slides9-1x4.pdf Didownload pada tanggal 08 Maret 2009. |
| [7] | Rocchio J. J., 1971, Relevance Feedback in Information Retrieval, In Salton, G. (Ed.): <i>The SMART Retrieval System. Experiments in Automatic Document Processing</i> ,313-323, New Jersey: Prentice Hall. |
| [8] | Drucker, Harris.,Shahrany, Behzad., Gibbon, David., Relevance Feedback using Support Vector Machines. http://www.svcl.ucsd.edu/~nuno/10.1.128.7440.pdf Didownload pada tanggal 08 Maret 2009 |
| [9] | Handouts : Relevance Feedback http://www.ir.iit.edu/~dagr/cs529/files/handouts/07Feedback.pdf Didownload pada tanggal 08 Maret 2009. |
| [10] | Baeza-Yates, R., Ribeiro-Neto, B., 1999, <i>Modern Information Retrieval</i> . Addison Wesley |
| [11] | Relevance Feedback http://en.wikipedia.org/wiki/Relevance_feedback Didownload pada tanggal 08 Maret 2009. |
| [12] | Salton, Gerald and Chris Buckley, <i>Improving Retrieval Performance by Relevance Feedback</i> , In <i>Journal of The American Society for Information Science</i> , 41 (4), 288-297,1990. http://users.cs.fiu.edu/~vagelis/classes/COP6776/publications/jasistSalton1990.pdf Didownload pada tanggal 18 Maret 2009. |
| [13] | Sembiring, Krisantus., Tutorial SVM Bahasa Indonesia. http://xsanctus.blog.wordpress.com/tutorial-svm-bahasa-indonesia-oleh-krisantus1.pdf Didownload tanggal 25 Maret 2009. |