

INTISARI

PENDEKATAN HYBRID UNTUK SISTEM REKOMENDASI FILM DENGAN COLLABORATIVE FILTERING DAN CONTENT-BASED

Oleh

Taufik Eko Cahyono

11/313643/PA/13735

Pertumbuhan jumlah data dan keanekaragaman informasi memunculkan dilema bagi masyarakat untuk memilih informasi yang relevan. Banyak teknik pembentuk rekomendasi yang dapat digunakan untuk merekomendasikan informasi yang sesuai dengan keinginan pengguna. *Collaborative filtering* (CF) dan *content-based* (CB) merupakan contoh teknik dasar sistem rekomendasi yang populer hingga saat ini. Pada kasus pencarian film yang ingin ditonton, CF menggunakan *rating* dari *user* lain untuk membentuk rekomendasi sedangkan CB menggunakan *content* dari *item* berupa atribut film (sinopsis, *keyword*, *genre* dan sebagainya). Namun, teknik CB dan CF memiliki kelebihan dan kekurangan masing-masing. Salah satu cara untuk menangani kelemahan masing-masing teknik adalah dengan menggabungkan keduanya (*hybrid*). Penggabungan teknik dasar ada yang dapat dilakukan secara *parallel* dan ada yang hanya bisa dilakukan secara *sequential*. Dalam prakteknya ada kondisi tertentu yang mana menyebabkan suatu rekomendasi berkurang kualitasnya, misalnya pada sistem rekomendasi film dengan jumlah *rating* yang terlalu sedikit atau informasi *content* dari film kurang lengkap.

Dalam penelitian ini dilakukan berbagai macam skenario percobaan untuk mengevaluasi teknik *hybrid*. Teknik *hybrid* yang dievaluasi meliputi *hybrid weighted*, *switching* dan *feature augmentation* dengan kondisi di mana *sparsity level* dari *dataset* yang tersedia lebih besar dan informasi *content* yang diperlukan lebih lengkap. *Dataset rating* film yang digunakan berasal dari MovieLens, sedangkan *content item* (atribut film) diperoleh dari situs IMDB.

Hasil akhir dari penelitian ini memperlihatkan bahwa teknik *hybrid weighted* memiliki nilai terbaik dalam hal waktu eksekusi dengan lama 1,145 detik maupun akurasi dengan MAE sebesar 0,743. Sedangkan dalam hal penanganan *sparsity problem* dan juga efektifitas, teknik *feature augmentation* memiliki hasil terbaik dengan nilai *conversion rate* 100%. Selain itu, penambahan *content* film dengan atribut lain tidak terlalu memberikan perubahan yang signifikan.

Kata kunci: *sistem rekomendasi, hybrid, collaborative filtering, content-based, neighborhood, sparsity, content*

ABSTRACT

HYBRID APPROACH FOR MOVIE RECOMMENDATION SYSTEM WITH COLLABORATIVE FILTERING AND CONTENT-BASED

By

Taufik Eko Cahyono
11/313643/PA/13735

The growing number of data and the diversity of information create a dilemma for people to choose relevant information. Many recommendation-forming techniques can be used to recommend information that suits the user's wishes. Collaborative filtering (CF) and content-based (CB) are examples of the basic recommendation techniques which is popular until now. In the case of movie that want to watch, CF uses ratings from other users to generate recommendation while CB uses movie's (item's) content as it's attributes (synopsis, keyword, genre and so on). However, CF and CB techniques have their respective advantages and disadvantages. One way to deal the weaknesses of each technique is by combining both technique (hybrid). Combining basic techniques can be done in parallel and some can only be done in sequentially. Practically, there are certain conditions which cause the quality of recommendations to decrease such as for example on movie recommendation system with small number of rating or the content information of the item is not enough.

In this study, various experimental scenarios were conducted to evaluate hybrid techniques. Hybrid techniques that were evaluated are hybrid weighted, switching and feature augmentation with conditions where the sparsity level of the available dataset is greater and the required content information is more complete. The movie rating dataset used comes from MovieLens, while the content item (movie attribute) is obtained from the IMDB site.

The end result of this study shows that hybrid weighted technique has the best value in terms of execution time with 1,145 seconds long and accuracy with 0,743 of MAE score. Meanwhile in terms of sparsity problem handling and effectiveness, feature augmentation technique get the best result with 100% of conversion rate. In addition, the addition of movie content with other attributes did not give any significant changes.

Keywords: *recommendation system, hybrid, collaborative filtering, content-based, neighborhood, sparsity, content*