

CLUSTERING APPROACH BASED ON FEATURE WEIGHTING FOR
RECOMMENDATION SYSTEM IN MOVIE DOMAIN

AMIR HOSSEIN NABIZADEH RAFSANJANI

A thesis submitted in fulfillment of the
requirements for the award of the degree of
Master of Science (Information Technology - Management)

Faculty of Computing
Universiti Teknologi Malaysia

JULY 2013

To my beloved Father and Mother

ACKNOWLEDGEMENT

I am heartily expressing my greater gratefulness to Allah s.w.t for His blessing and strength that He blessed to me during the completion of this research.

In preparing this dissertation, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main dissertation supervisor, Professor Dr. Naomie Salim, for encouragement, guidance, critics and friendship. Without her continued support and interest, this dissertation would not have been the same as presented here. Furthermore, I would like to thank my dear friend Mr. Amin Minouei who helped me during application development phase of my research.

I appreciate my family especially my father and mother because of their support to finish this study.

ABSTRACT

The advancement of the Internet has brought us into a world that represents a huge amount of information items such as movies, web pages, etc. with fluctuating quality. As a result of this massive world of items, people get confused and the question “Which one should I select?” arises in their minds. Recommendation Systems address the problem of getting confused about items to choose, and filter a specific type of information with a specific information filtering technique that attempts to present information items that are likely of interest to the user. A variety of information filtering techniques have been proposed for performing recommendations, including content-based and collaborative techniques which are the most commonly used approaches in recommendation systems. This dissertation introduces a new recommendation model, a feature weighting technique to cluster the user for recommendation top-n movies to avoid new user cold start and scalability problem. The distinctive point of this study lies in the methodology used to cluster the user and the methodology which is utilized to recommend movies to new users. The model makes it possible for the new users to define a weight for every feature of movie based on its importance to the new user in scale of one (with an increment of 0.1). By using these weights, it finds nearest cluster of users to the new user and suggests him the top-n movies (with the highest rate and most frequency) which are reviewed by users that are in the targeted cluster. Rating and Movie dataset were used during this study. Firstly, purity and entropy are applied to evaluate the clusters and then precision, recall and F1 metrics are used to assess the recommendation system. Eventually, the results of accuracy testing of proposed model are compared with two traditional models (OPENMORE and Movie Magician Hybrid) and based on the evaluation the level of preciseness of the proposed model is more better than Movie Magician Hybrid but worse than OPENMORE.

ABSTRAK

Kemajuan internet telah membawa kita ke dalam dunia yang mewakili sejumlah besar barangan maklumat seperti filem dengan kualiti yang berubah. Hasil daripada dunia barangan ini secara besar-besaran, orang keliru dan persoalan "Mana satu yang perlu saya pilih?" timbul dalam fikiran mereka. RS menangani masalah kekeliruan dalam memilih maklumat dan menapis jenis maklumat tertentu dengan teknik khusus penapisan maklumat yang cuba untuk membentangkan maklumat yang mungkin menarik minat pengguna. Pelbagai teknik penapisan maklumat telah dicadangkan untuk melaksanakan cadangan, termasuk teknik berasaskan kandungan dan kerjasama yang merupakan pendekatan yang paling biasa digunakan dalam RSs. Disertasi ini memperkenalkan, satu ciri teknik pemberat untuk kelompok pengguna bagi mencadangkan filem top-n untuk menghindar permulaan yang dingin bagi pengguna baru dan masalah berskala daripada sistem penentu. Perkara tersendiri kajian ini terletak pada kaedah yang digunakan untuk kelompok pengguna dan kaedah yang digunakan untuk mengesyorkan filem kepada pengguna baru. Model ini membolehkan pengguna baru untuk menentukan berat bagi setiap ciri filem berdasarkan kepentingan untuk pengguna baru dalam skala satu (dengan peningkatan sebanyak 0.1). Dengan menggunakan berat, ia mendapati kelompok terdekat dari pengguna kepada pengguna yang baru dan mencadangkan top-n filem (dengan kadar yang paling tinggi dan kekerapan yang paling). Dataset penilaian dan filem telah digunakan semasa kajian ini. Pertama, purity dan entropy digunakan untuk menilai kelompok dan kemudian precision, recall dan F1 metrics digunakan untuk menilai RS. Akhirnya, keputusan ujian ketepatan model yang dicadangkan berbanding dengan dua model tradisional (OPENMORE dan Movie Magician hibrid) dan berdasarkan penilaian tahap ketepatan model yang dicadangkan adalah lebih baik daripada Movie Magician Hibrid tetapi lebih teruk daripada OPENMORE.