คำบรรยายลักษณะผู้ใช้ที่มีหลายปัจจัยแบบใหม่บนพื้นฐานของการจัดลำดับความสำคัญของ
ปัจจัยสำหรับระบบแนะนำภาพยนตร์

นายจิรัชย์ ดวงจำปา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และสารสนเทศ ภาควิชาคณิตศาสตร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2553

A NOVEL MULTI-CRITERIA USER PROFILE BASED ON CRITERIA-RANKING FOR MOVIE RECOMMENDER

Mr. Jirach Duangjumpa

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Computer Science and Information

Department of Mathematics

Faculty of Science

Chulalongkorn University

Academic Year 2010

Thesis Title	A NOVEL MULTI-CRITERIA USER PROFILE BASED ON CRITERIA-RANKING FOR MOVIE RECOMMENDER
By	Mr. Jirach Duangjumpa
Field of Study	Computer Science and Information
Thesis Advisor	Assistant Professor Saranya Maneeroj, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

.... Dean of the Faculty of Science

(Professor Supot Hannongbua, Dr.rer.nat.)

THESIS COMMITTEE

.... Chairman

(Suphakant Phimoltares, Ph.D.)

.... Thesis Advisor

(Assistant Professor Saranya Maneeroj, Ph.D.)

.... External Examiner

(Atchara Mahaweerawat, Ph.D.)

จิรัชย์ ดวงจำปา : คำบรรยายลักษณะผู้ใช้ที่มีหลายปัจจัยแบบใหม่บนพื้นฐานของการจัดลำดับความสำคัญของปัจจัยสำหรับระบบแนะนำภาพยนตร์. (A NOVEL MULTI-CRITERIA USER PROFILE BASED ON CRITERIA-RANKING FOR MOVIE RECOMMENDER) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ.ดร.ศรันญา มณีโรจน์, 57 หน้า.

วิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์ในการจัดขึ้นเพื่อที่จะเสนอวิธีการแบบใหม่สำหรับระบบแนะนำภาพยนตร์บนพื้นฐานข้อมูลความคิดเห็นที่มีหลากหลายปัจจัย เหล่าผู้ค้นคว้าได้พัฒนาข้อมูลความคิดเห็นที่มีหลากหลายปัจจัยไปในหลายๆวิธี ส่วนใหญ่จะใช้ค่าของคำบรรยายลักษณะผู้ใช้มาใช้โดยตรงในการหาความสัมพันธ์ระหว่างผู้ใช้ ซึ่งอาจเป็นสาเหตุของปัญหาและนำไปสู่การจัดหาเพื่อนผู้ซึ่งมีความคิดเห็นคล้ายกัน ได้คุณภาพแย่ ดังนั้น ผมจึงเสนอ การเรียงลำดับปัจจัย กับ คะแนนความใกล้เคียง เพื่อบ่งบอกความสำคัญของปัจจัยต่อลักษณะนิสัยของผู้ใช้แต่ละคน ซึ่งจะทำให้คุณภาพของคำบรรยายลักษณะผู้ใช้ดีขึ้น จากผลการทดลอง การเรียงลำดับปัจจัย กับ คะแนนความใกล้เคียง ให้ผลการแนะนำที่แม่นยำกว่าวิธีเทคนิคแบบดั้งเดิมของระบบแนะนำบนข้อมูลที่มีหลากหลายปัจจัย

ภาควิชา คณิตศาสตร์ ลายมือชื่อนิสิต จิรัชย์ ดวงจำปา

สาขาวิชา วิทยาการคอมพิวเตอร์และสารสนเทศ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก

ปีการศึกษา 2553

5173614923: MAJOR COMPUTER SCIENCE AND INFORMATION
KEYWORDS :KNOWLEGDE MANAGEMENT / CRITERIA-RANKING / MULTI-CRITERIA RECOMMENDER SYSTEM / HYBRID RECOMMENDER SYSTEM

JIRACH DUANGJUMPA: A NOVEL MULTI-CRITERIA USER PROFILE BASED ON CRITERIA-RANKING FOR MOVIE RECOMMENDER. ADVISOR: ASST.PROF SARANYA MANEEROJ, Ph.D., 57 pp.

This thesis aims at proposing a novel methodology for Movie Recommender System based on multi-criteria ratings. Many researchers have been developed the multiple criteria ratings on various ways. Most of them directly use the values of user profiles to find the relation among users, which may cause problems and tend to provide poor quality neighbors. Therefore, I propose Criteria-Ranking with Closeness Score to indicate the significance of criteria toward each user characteristic that will improve the quality of user profiles. According to the experimental evaluation, Criteria-Ranking with Closeness Score provides more accurate recommendation results than other traditional multi-criteria recommender techniques.

Department : Mathematics	Student's Signature Jirach Duangjumpa
Field of Study : Computer Science and Information	Advisor's Signature
Academic Year : 2010	

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Assistant Professor Saranya Maneeroj, Ph.D. and Sasipa Panthuwadeethorn, Ph.D. for all her great encouragement, suggestions and assistance. Without her support and attention, this thesis would have never been completed.

I am also grateful to the thesis committee, Professor Suphakant Phimoltares, Ph.D. and Professor Atchara Mahaweerawat, Ph.D. for their constructive criticisms and invaluable advises.

Finally, I also express the deeply thanks to my family and relative to their encouragements since the beginning of my study and practically, when I feel to break down.

CONTENTS

CHAPTER Page

Abstract in Thai.. iv

Abstract in English.. v

Acknowledgements.. vi

Contents.. vii

List of Tables... Ix

List of Figures.. x

CHAPTERS

I. INTRODUCTION... 1

1.1 Motivation and problem description.. 1

1.2 Scope of study 10

1.3 The Objectives of the Research... 10

1.4 The benefit 10

II. THEORETICAL BACKGROUND 11

2.1 Recommendation techniques... 12

2.1.1 Collaborative Filtering.. 12

2.1.2 Content-Based Filtering.. 14

2.1.3 Hybrid... 15

2.2 The step of recommendation.. 16

2.2.1 Information Preparation... 17

2.2.2 Similarity Measurement... 17

2.2.3 Prediction.. 20

2.3 Single rating and Multi-criteria ratings.. 20

2.3.1 Single rating.. 20

2.3.2 Multi-criteria ratings... 21

2.3.3 Multi-criteria ratings recommender technique... 22

CHAPTER Page

2.3.3.1 Aggregating traditional similarities that are based on each individual rating........ 24

2.3.3.2 Calculating similarity using multidimensional distance metrics.... 24

2.4 Multidimensional distance metrics........ 27

2.4.1 Manhattan Distance........ 27

2.4.2 Euclidean distance........ 28

III. METHODOLOGY 29

3.1 User Profile........ 31

3.2 Ranking Profile........ 33

3.3 Similarity Measurement........ 33

3.3.1 Closeness Score Measurement........ 34

3.3.2 Top N Neighbor's Selection........ 37

3.3.3 Similarity Value Calculation........ 38

3.4 Prediction Generation........ 39

IV. EXPERIMENTAL RESULTS AND DISSCUSION........ 41

4.1 Data........ 41

4.1.1 Accumulated the dataset from Yahoo! Movie Site 43

4.1.2 Transforming Data 47

4.2 Evaluation Metric........ 48

4.3 Experimental Result........ 48

4.4 Discussions........ 50

V. CONCLUSION AND FUTURE WORKS........ 51

5.1 Conclusions........ 51

5.3 Future works........ 51

REFERENCES........ 53

VITAE........ 57

LIST OF TABLES

Table		Page
1.1	The comparison of Single rating and Multi-criteria ratings…………………………	3
1.2	The user profile of multi-criteria ratings…………………….………………………	4
1.3	The difference between target user and user A ………………………………………	5
1.4	The difference between target user and user B ………………………………………	5
1.5	The difference between target user and user A on Criteria-Ranking profile ………	6
1.6	The difference between target user and user B on Criteria-Ranking profile ………	7
1.7	The example of SDDS problem on criteria-ranking profile …………………………..	7
1.8	The difference between target user and user A on the example of SDDS problem on criteria-ranking profile ……………………………………………………………	8
1.9	The difference between target user and user B on the example of SDDS problem on criteria-ranking profile ……………………………………………………………	8
1.10	The example of a result using the closeness score with criteria-ranking profile …	9
2.1	The example “co-rated” items..…………………………………………………………	13
2.2	The example single rating and multi-criteria ratings on a user ……….……………..	21
2.3	The example single rating and multi-criteria ratings on multi-users…………………	22
3.1	The example of data of a user………………………………………………………...	32
3.2	The example of a user profile…………………………………………………………	32
3.3	The criteria-ranking profile…………………………………………………………….	33
3.4	The difference of rank order…………………………………………………………..	35
3.5	The closeness value table……………………………………………………………..	36
3.6	The example of a pair of criteria-ranking profiles……………………………………..	36
3.7	The example list of neighbor with their closeness score (unordered)………………	37
3.8	The example list of neighbor with their closeness score (ordered)…………………	38
3.9	The example of the calculation of similarity value……………………………………..	39
4.1	The format of real database …………………………………………………………..	42
4.2	The example of real rating information………………………………………………..	42
4.3	The mapping table from character rating into numerical rating……………………..	47

Table		Page
4.4	The experimental result……………………………………………………………………..	49

LIST OF FIGURES

Figure		Page
1.1	The example of transformation of Criteria-Ranking …………………………………..	6
2.1	The difference between Manhattan distance and Euclidean distance…………….	28
3.1	Recommendation Process ……………………………...………………………………	31
4.1	Yahoo! Movie main page ………….…………………………………………………..	43
4.2	The box office page …...…………………………………………………………………	44
4.3	The page of Movie "Unknown"…..	45
4.4	User Reviews page of "Unknown"………………………………………………………	46

CHAPTER I

INTRODUCTION

1.1 Motivation and Problem Description

The proliferation of the internet provides an enormous variety of choices to obtain more information for customers in selecting the product(s) that meets his/her desires. Recommender systems were created for helping user to make the decision(s) when they have to select their favorite item(s) from the huge database. The system collects history of user's preference toward the items. Then, their past behaviors and preferences are used to make the prediction by matching them with a set of available product in the database to generate a ranking of products the most interesting to each user. In the present, recommender systems become base method for E-commerce, which is widely used in many Internet websites.

The recommendation process starts with the initial set of ratings that are either explicitly provided by the users or implicitly inferred by the system. The goal of a typical recommendation system is to predict the rating of unrated items using the rated items' ratings and also find the items that should be utilized to the user. Generally, there are three steps in the recommendation process. The first one is to create the user profiles which collect users' preferences toward items that they have been rated. Next step is the similarity measurement which uses the metric to measure the correlation between users by using their user profiles. The Last one is the prediction step which uses data from both previous steps to generate the recommendation.

There are three reliable techniques of the recommendation system based on their approach: Content-Based Filtering, Collaborative Filtering and hybrid techniques. Content-Based Filtering technique generate the recommendations based on the similarity of items to the ones that user preferred in the past. Collaborative Filtering technique generates the recommendations based on the opinions of the group of similar users toward the target user. Hybrid technique is designed by combining the other two or more techniques together. This combination can be done on many different ways in order to improve recommendation quality such as Maneeroj [2], Balabanovic & Shoham [9], and Adomavicius & Tuzhilin [8]. They integrate Content-Based Filtering into Collaborative Filtering to reduce disadvantage of each other.

These three main techniques use the overall rating to predict the preference of users for unseen items. In another word, the vast majority of current recommendation systems typically used single rating to represent the utility of an item to users in the two-dimensional Users × Items space.

Next generation of the recommendation system techniques is Multi-criteria ratings recommender technique, which provides additional information about user preferences regarding several important aspects of an item. This will potentially benefit for recommender system because they can define the characteristic of users more clearly. Therefore, it can increase the accuracy of the recommendations.

Normally, the user profile is created by using rating of user with the aspects of items or items feature. This user profile will represent the characteristic of user on purchasing or choosing items. Only user profiles of single criteria may not be able to denote the suitable characteristic because it can just express how much users prefer the items. However, user profiles of Multi-criteria ratings, which obtain more information about users' preferences, will be more efficient to define the characteristic of users since it can indicate the reason why users prefer the items.

For example, if we consider single rating (overall rating) from table1 user U_1 and U_2 seem to be the nearest neighbors for the target user, since they got exactly the same overall rating score on the similar items. Nevertheless, if we examine in the multiple criteria ratings view, the suitable neighbors for the target user should be U_3 and U_4. Since multi-criteria provide the feedback about the specific aspect of items. Therefore, we can clearly see that U_1 and U_2 have the difference in their tastes toward the target user even though their overall ratings for the items match perfectly. This comparison as shown in table 1.1 will indicate the reason why Multi-criteria ratings provide more appropriate characteristics than single rating.

Table1.1: The comparison of Single rating and Multi-criteria ratings

User	Item1	Item2	Item3	Item4
Target	6 (5,7,7,5)	5 (7,3,3,7)	6 (5,7,7,5)	5 (7,3,3,7)
U_1	6 (8,4,4,8)	5 (2,8,7,3)	6 (8,4,3,9)	5 (2,8,8,2)
U_2	6 (8,4,3,9)	5 (2,8,8,2)	6 (8,4,4,8)	5 (2,8,7,3)
U_3	5 (4,6,6,4)	6 (8,4,4,8)	5 (4,6,6,4)	6 (8,4,4,8)
U_4	5 (4,6,6,4)	6 (8,4,4,8)	5 (4,6,6,4)	6 (8,4,4,8)

Multi-criteria ratings represent the utility of an item *i* to a user *u* in the multi-dimensional Users × Items × Criteria space. When the recommender system has to handle with *n* criteria data, then methodically definition of the Multi-criteria ratings can be defined as the following:

$$R(u, i) = (r_{c1}(u, i) \ldots r_{cn}(u, i)) \tag{1}$$

Where $r_c(u, i)$ denotes the rated score of user *u* for item *i* on the criterion *c*. $R(u, i)$ stands for the predicted rating value of user *u* toward item *i*.

The Multi-criteria ratings recommender systems assess the correlation among users on Multi-criteria ratings in order to find the neighborhood. In the calculation process, several literatures use the difference between each pair of users' profiles, which directly uses the Multi-criteria ratings to perform the calculation and then convert the derived difference values into the similarity. The lowest difference is the most similar and also the good neighbor.

However, the exact meaning of the good neighbor is the user who has similar tastes with the target user. In another word, the user who thinks the alike as the target user should prefer the same things. Therefore, is it suitable to use the real rating values in the similarity calculation? Since, system might have the problem to select the better neighbor when the different among them are equal.

For example, if the difference values between the target user and the other two users are identically that means the other two users should have the same level of the similarity values toward the target user. However, one of them got high difference on one criterion while another one got the low difference on several criteria. In this case, their overall different values are equal but the level of the similarity values of them should be different. Since one of them got the difference on only one criterion while another one got the difference on several criteria. We call this problem as “SDDS problem” (Same Distance but Different Similar)

For example, suppose that the target user, user A, and user B had their own profile that can be display as table 1.2.

Table1.2: The user profile of Multi-criteria ratings

User ID	Criteria1	Criteria 2	Criteria 3	Criteria 4
Target user	9	8	7	7
User A	9	8	7	5
User B	9	7	7	6

In order to find the relation among users, system will measure the distance or difference between the user A with the target user as shown in table 1.3 and also user B with the target user as displayed in table 1.4. For these two tables, the Manhattan Distance is used as the metric to find the difference values between a pair of users.

Table1.3: The difference between target user and user A

User ID	Criteria1	Criteria 2	Criteria 3	Criteria 4	Sum
Target User	9	8	7	7	
User A	9	8	7	5	
Difference of Target user, User A	0	0	0	2	2

Table 1.4: The difference between target user and user B

User ID	Criteria1	Criteria 2	Criteria 3	Criteria 4	Sum
Target User	9	8	7	7	
User B	9	7	7	6	
Difference of Target user, User B	0	1	0	1	2

According to the evaluation result from table 1.3 and table 1.4, the overall difference values of user A and user B toward the target user are equal. However, if we thoroughly consider about the values of each criterion between the target user and the other two users, user A should be the better neighbor. In another word, user A should have more similarity toward the target than user B, since the user A got same value on three criteria, which are criteria 1, criteria 2, and criteria 3 while user B got only two criteria, which are criteria 1 and criteria 3.

The cause of the SDDS problem is on the calculation process, since the different pair of numbers can produce the same value. In order to reduce this SDDS problem, we would like to focus on the significance of criteria as the optimal solution. The significance of criteria is considered to be a technical assistance for the system that not only certifies the importance of each criterion to users but also better represent the characteristic of user than real criteria values.

To represent the significance of criteria, we transform the values on each criterion into the level of importance of each criterion. We name this technique as Criteria-Ranking. For example, in the case of table1.2, we use the Criteria-Ranking to transform each criteria rating value into the Criteria-Ranking value. So the Criteria-Ranking profiles can be shown as the figure 1.1.

User ID	Criteria1	Criteria 2	Criteria 3	Criteria 4
Target user	9	8	7	7
User A	9	8	7	5
User B	9	7	7	6

User ID	Criteria1	Criteria 2	Criteria 3	Criteria 4
Target user	Rank1	Rank2	Rank3	Rank3
User A	Rank1	Rank2	Rank3	Rank4
User B	Rank1	Rank2	Rank2	Rank4

Figure 1.1: The example of transformation of Criteria-Ranking

For every criterion value of target user, user A's profile and user B's profile are transforming to be Criteria-Ranking profiles as the lower table from figure 1. Then Manhattan Distance is used to evaluate the difference between them. As the result from table 1.5 and table 1.6, because the user A has lower distance of Criteria-Ranking than user B, user A is a better neighbor than user B on another word user A is more similar to the target user than user B. Consequently, the SDDS problem from table 1.3 and table 1.4 is eliminated.

Table1.5: The difference between target user and user A on Criteria-Ranking profile

User ID	Criteria1	Criteria 2	Criteria 3	Criteria 4	Sum
Target User	Rank1	Rank2	Rank3	Rank3	
User A	Rank1	Rank2	Rank3	Rank4	
Difference of Target user, User A	0	0	0	1	1

Table 1.6: The difference between target user and user B on Criteria-Ranking profile

User ID	Criteria1	Criteria 2	Criteria 3	Criteria 4	Sum
Target User	Rank1	Rank2	Rank3	Rank3	
User B	Rank1	Rank2	Rank2	Rank4	
Difference of Target user, User A	0	0	1	1	2

Although Criteria-Ranking seems to play a role in helping the system to solve the SDDS problem, but even we transform the criteria rating value into criteria ranking, we still face the SDDS problem in some situation. For example, suppose that system face with the user profile that looks like the table 1.7.

Table1.7: The example of SDDS problem on criteria-ranking profile

User ID	Criteria1	Criteria 2	Criteria 3	Criteria 4
Target user	Rank1	Rank1	Rank1	Rank4
User A	Rank1	Rank1	Rank4	Rank1
User B	Rank1	Rank3	Rank3	Rank2

After the system evaluated the Manhattan Distance metric to measure the difference between the target user and the other user, the result of measurement can be displayed as table 1.8 and table 1.9. This can prove that SDDS problem can also happen with the Criteria-Ranking technique too.

Table 1.8: The difference between target user and user A on the example of SDDS problem on criteria-ranking profile

User ID	Criteria1	Criteria 2	Criteria 3	Criteria 4	Sum
Target User	Rank1	Rank1	Rank1	Rank4	
User A	Rank1	Rank1	Rank4	Rank1	
Difference of Target user, User A	0	0	3	3	6

Table 1.9: The difference between target user and user B on the example of SDDS problem on criteria-ranking profile

User ID	Criteria1	Criteria 2	Criteria 3	Criteria 4	Sum
Target User	Rank1	Rank1	Rank1	Rank4	
User B	Rank1	Rank3	Rank3	Rank2	
Difference of Target user, User B	0	2	2	2	6

In order to reduce the SDDS problem in Criteria-Ranking profile, system needs more technique to handle with this situation. From the fact that, the first rank of the Criteria-Ranking profile should play the more crucial role than the other ranks, that means it is the most effective criteria toward the target user. Therefore, we construct the Closeness Score in order to indicate the importance of rank in different level. This Closeness Score will give more value for higher rank and less value for lower rank when the system has to face with the same difference between a pair of user. For example, if we used the Closeness Score with the table 1.9. Then, the result of evaluation by using the Closeness Score will be shown as the table 1.10.

Table 1.10: The example of a result using the closeness score with criteria-ranking profile

User ID	Criteria1	Criteria 2	Criteria 3	Criteria 4
Target User	Rank1	Rank1	Rank1	Rank4
User B	Rank1	Rank3	Rank3	Rank2
Difference of Target user, User A	0	2	2	2
Closeness Score	Highest	Higher	Higher	Lower

In the criteria 1, the target user and also user B have the same rank which is the first rank. That means the Closeness Score of this criterion should get the highest score. For the criteria 2, criteria 3, and criteria 4, even these three criteria have the same different value but criteria 2 and criteria 3 are more important for the target user. Since, the Criteria-Ranking values of the criteria 2 and also criteria 3 are the first rank of the target user while the Criteria-Ranking value of criteria 4 is the fourth rank of the target user. Therefore, criteria 2 and criteria 3 should get the higher score than the criteria 4.

We use this Closeness Score to support the Criteria-Ranking technique in order to cope with the SDDS problem. The detail of our methodology will be described in chapter 3.

In this thesis, the idea of Criteria-Ranking and Closeness Score are proposed to give the significance of each criterion toward each user on the user profiles. This will improve the quality of the user profile. After the suitable user profile obtained, the group of good quality neighbors is formed and more the accurate recommendations will be achieved.

The next chapter described about the theoretical background which explains the detail of related work. We then describe more about our methodology in chapter 3. In chapter 4, we discuss about the experimental results. Finally, we give some conclusion, discussion and future work in the last chapter.

1.2 The Scope of study

1. The domain in this thesis is only Movie.
2. Data is obtained from Yahoo Movie System, which consists of 200 users, 1358 movies, and 2550 rating.
3. The considered Criteria of the movie are the overall, story, acting, directions, and visual.
4. The prediction result is in the term of overall rating.

1.3 The Objectives of the Research

1) To enhance the accuracy of recommendation based on Multi-criteria ratings.
2) Propose the Criteria-Ranking technique to transform the user profile into the novel user profile.
3) Propose the Closeness Score to measure the similarity among users in the term of the novel user profile.

1.4 The benefit

The proposed methodology will help the Multi-criteria ratings recommender system to produce the higher quality of the recommendation by transforming the preference data into the user profile in the more suitable way.

CHAPTER II

THEORETICAL BACKGROUND

In this chapter, we will briefly describe about theories that is related to this thesis. Details of the theories comprise of the derivation, procedure, and also the example of the related work.

Naturally, when people have to make up their mind in selecting something or make a choice on the things that they do not have any idea, they usually rely on the experience and opinions of the others. Of course we can find out the suggestions from the people who are expert at those things or familiar with the choices we face. Sometimes we cannot make any decision on the available choices by ourselves, unless the topic of interest is the talk of the town.

In the past, the recommendation occasionally lacks of credibility because they are considered as a means of hidden advertising. Therefore, a new generation of the recommender system site has merged to provide a user with more comfortable when a user wants to access the information on the site. The recommender system will accumulate the history behaviors and also the preferences of users' experiences. This experience feedback from web users will be gathered to form the content of recommendation. Finally, the recommender system will analyze the experience of users to perform the appropriated recommendation. The recommendation from the recommender system will be substitute for the suggestion from the expert.

In the present, the internet has been widely used for exploring the information which has various categories and also huge information. Therefore, the increasing numbers of people are turning to the computational recommender system as Akharraz [11] to support, mediate, or automate the process of sharing the suggested information as Terveen [13]. This is the beginning of the recommender system.

2.1 Recommendation techniques

The recommender system is another useful system that can help user to handle with the large information. The technique of recommendation is the methodology that uses information about user as the helpful data by analyzing the factor of user's behavior, the attribute of the items, and also the contextual information that might affect the decision of users and then classify the appropriate information to the user. This method will support user to access information with more comfortableness and satisfaction, especially when users have to face with the large item spaces by providing user with the interesting suggestion. The recommendation techniques are usually classified into three categories based on their approach to recommendation, which are Collaborative Filtering, Content-Based Filtering and Hybrid.

In addition, the recommender system can also be classified based on the nature of their algorithm technique into Memory-based and Model-based approach as Breese [10]. For the memory based approach, it usually evaluates the recommendation based directly on the prior activities of user. This approach represents the study of how people use their experience in order to improve performance heuristics. In contrast, model-based approach uses previous user activities to make a recommendation by learning a predictive model with prior behavior of user (typically using some statistical or machine-learning methods).

2.1.1 Collaborative Filtering

The Collaborative Filtering is the most widespread technique as Breese [10], and Herlocker [12]. From the fact that, people usually prefer to ask their friends who have the same taste for the suggestion of the things that they do not have their own experience with. This can explain the nature of this kind of recommendation technique.

The Collaborative Filtering technique generated recommendations by providing the items that were interested by other users who have the similar preference with the target user. This technique will accumulate the user's preference data in order to find the relation among users by measuring the similarity between a pair of users. In the similarity measurement process, system accumulated the set of preferred items that both target user and another user have rated, which is called co-rated items as shown in table 2.1.

Table 2.1: The example co-rated items

User	*Item1*	*Item2*	*Item3*	*Item4*
A	5	-	3	5
B	4	5	-	5

co-rated items

After the measurement is over, the group of similar users is obtained and then the prediction is generated. After the prediction has finished, the items which are predicted with the higher score will be recommended to target user.

The Collaborative Filtering technique does not depend on the components of the items on the consideration. Instead, the recommendation process automatically based on user's opinion only.

Unfortunately, the main disadvantage of this technique is that people might give their own opinions on the different items so that the co-rated items between the target user with the other users are less or none. This means that system may not be able to find the neighbor of user or system can only provide the user with the poor quality neighbor. Since, the finding neighborhood process needs a lot the co-rated items in order to generate the good quality neighbor. Therefore, it cannot be sure that they really have the same taste or not.

2.1.2 Content-Based Filtering

The Content-Based Filtering technique analyzed the user preference data to determine which item is relevant to the item that user liked in the past. The Content-Based Filtering technique derived from the idea that user should satisfy the items similar to the items the user had his/her experiences with. Most of the system tried to summarize the past item contents which user has been rated on purchased or consumes items in terms of user profile. Usually, the user profile is indicated by a vector whose elements explain the aspect of the evaluated items. When the user profile is completely created, the system will select the items that should be interesting by the user using similarity metric to measure the similarity between content in a user profile and an item based on user's prior feedback in the characteristics and properties of items. Finally, the items which have a high similarity will be recommended to the user.

Unfortunately, the weakness of this technique is the list of predicted items will be the static kind of items. Since, the system will estimate the rating of unrated item base on the given rating. That means the user will be recommended with only a set of the same old things. Therefore, a user will have no opportunity to be recommended to the items that have different attribute or feature from the items he/she familiar with.

2.1.3 Hybrid

Hybrid is the technique that combines two or more techniques such as Collaborative Filtering and Content-Based Filtering to decrease the limitation or weakness of each technique in order to gain the higher recommendation quality result. There are many ways to combine techniques together such as in the case of Lakiotaki [14]. Uta-rec was a utility-based recommender that worked using the preference disaggregation principle. In order to model a user's preference in terms of a set of additive utility functions, this system implemented by using UTA algorithm. Claypool [7] introduced the p-tango method, which makes use of both Content-Based and Collaborative Filtering through a linear combination for online newspaper domain. Bichler [15] combined decision analysis techniques and multi-attribute auction mechanisms in order to procure goods and services. Ricci [16] make the recommendation rely on the combination of factors as: appropriate destination modeling; data retrieval and filtering with both sharp and approximate matching; scoring using personal preferences that can be derived from a base of previous cases. Green [17] used conjoint analysis as their methodology. They use various models to infer buyers' partworths for attribute level then use it predict how buyers will choose among products and services. Balabanovic & Shoham [9] they created the Fab system which is the combination of Content-Based Filtering with Collaborative Filtering. It created user profiles instead of used user's ratings used based on Content-Based Filtering technique. The similarity for prediction is only based on the user profiles which mean that the quality of predictions is fully dependent on the Content-Based Filtering technique. Cho & Kim [26] proposes a recommendation methodology based on Web usage mining that populates the rating database by tracking customers' shopping behaviors on the Web, and product taxonomy, which is used to improve the performance of searching for nearest neighbors through dimensionality reduction of the database. Li [18] they evaluated a new algorithm, which is Collaborative Filtering based on item and user (CF-UI), for help E-Commerce in the recommendation. Lee [19] tries to personalize each user in order to know their buying behaviors and accordingly develop more appropriate marketing strategies for each user and provide the suitable information and products/services to serve user needed. The customer's satisfaction and loyalty can thus be enhanced, and the increase in each user's visiting frequency can further create more transaction opportunities and benefit the Internet businesses. Maneeroj [2] performed the hybrid recommendation system based on Content-Based Filtering to integrate into Collaborative Filtering in order to improve recommendation quality. Instead of using only overall rating value, they use user's opinions on features to deal with a poor neighbor problem. In the finding high quality neighbor, they used two filtering processes in order to form a set neighbor. They used Content-Based as the first filtering in order to obtain the rough neighbors set and then used it into co-rated item method with the "user's opinion on various features" as the second filtering to obtain the final neighbors set that have good quality.

Nowadays, the way of combination is composed in order to measure similarity between a pair of users. This way follows the idea of Collaborative Filtering approach. It differs from the traditional one in the sense that the similarity measurement between a pair of users avoids using the co-rated item set by using Content-Based user profiles, and measures the similarity among users based on their profiles. After the list of neighbors is formed, the suggestion will be done when the system finished the prediction of the unevaluated item.

This technique is quite useful and effective, since it can increase the opportunity that users are recommended the serendipitous item. Moreover, it does not use co-rated item set in similarity measurement. Therefore, the system can know the similarity between a pair of users even though the users have not rated on the same set of items.

2.2 The step of recommendation

Generally, there are three steps to generate recommendations in the recommender system, which are information preparation, similarity measurement and prediction.

2.2.1 Information Preparation

The first step is information preparation. In this step, all useful information such as the user preference toward the items (overall rating,…), the attribute of items (movie genre,…), and the contextual information (date, time, place,…) are incorporated and transformed by the system into the term that their algorithm need such as the user profile. This profile will be comprised of the information that represents the characteristic of each user.

2.2.2 Similarity Measurement

The second step is similarity measurement. After the system got all needed information, the system will find the users or the items that related to the target user by measuring the similarity between the target user information with the other users' information or the information about the items of user with the available items. This step will provide the set of the nearest neighbor or the set of the similar items to be the component of the prediction step. The similarity can be performing by using the correlation coefficient to find out the relation among users such as Pearson's Correlation coefficient, and Cosine coefficient, which will be described as the following.

- Pearson's Correlation coefficient

Pearson's Correlation coefficient is a kind of correlation coefficient that denotes the relationship between two variables that are measured on the same interval or ratio scale.

Numerically, the Pearson's Correlation coefficient is indicated as same as a correlation coefficient that is used in linear regression; ranging from -1 to +1. A value of +1 is the result of a perfect positive relationship between two or more variables. On the other hand, a value of -1 determines a perfect negative relationship. When it is used with a non-linear equation, the Pearson coefficient can be misleadingly small.

$$r(S) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (2)$$

- **Cosine coefficient**

Cosine similarity measures the cosine of the angle between two vectors in order to measure the similarity between them. When the angle is 0, the result of the Cosine function is equal to 1. When the angle is of unequal to 0, the result of the Cosine function is less than 1. Calculating the cosine of the angle between two vectors thus determines whether two vectors are pointing the same direction.

Usually, this Cosine coefficient is used to compare the documents in text mining. Furthermore, in the field of Data Mining, it is used to measure cohesion within clusters.

Cosine of two vectors can be easily evaluated by using the Euclidean Dot Product formula as the following equation:

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \, \|\mathbf{b}\| \cos\theta \qquad (3)$$

The cosine similarity is represented using a dot product and magnitude as.

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \quad (4)$$

For text matching, the attribute vectors *A* and *B* are usually the term frequency vectors of the documents. The cosine similarity can be seen as a method of normalizing document length during the comparison.

The result of similarity ranges from -1 means it was exactly opposite, to 1 means exactly the same, with 0 usually indicates independence, and in-between values indicate intermediate similarity or dissimilarity.

The cosine similarity of two documents will range from 0 to 1 in the case of information retrieval, since the term frequencies (tf-idf weights) cannot be negative. The angle between two term frequency vectors will be less than 90°.

This cosine similarity metric may be extended such that it yields the Jaccard coefficient in the case of binary attributes. This is the **Tanimoto coefficient**, *T* (*A*, *B*), represented as.

$$T(A,B) = \frac{A \cdot B}{\|A\|^2 + \|B\|^2 - A \cdot B} \quad (5)$$

2.2.3 Prediction

The final step is the prediction. In this step, system used all related information in the numerical format such as rating and similarity to calculate predictions that indicate user feeling toward the items. After the prediction finished, the system will generate the recommendation result based on the list of the predicted items. The items the target user has not rated were predicted in the term of single rating by using neighbor's rating given on such an item weighted by similarity values between the target user and their neighbors.

$$Pr_i = \sum_{j=1}^{n} (r_{ij} * s_j) / \sum_{j=1}^{n} (s_j) \qquad (6)$$

Where Pr_i represents predicted rating for the movie i^{th} of target user, r_{ij} denotes overall rating score of neighbor j^{th} for movie i^{th} and s_j indicates the similarity value between neighbor j^{th} and the target user.

2.3 Single rating and Multi-criteria ratings

2.3.1 Single rating

In the past, the recommender system used single rating to calculate and evaluate the recommendation. Single rating denotes the overall preference of the user toward the items.

Above three techniques Collaborative Filtering, Content-Based Filtering, and Hybrid have been developed on the single rating (overall rating) to represent the utility of an item to a user in the two-dimensional *User* $\times$ *Item* space. In order to improve recommendation, the developer needs to accumulate more information.

2.3.2 Multi-criteria ratings

This addition data will help system to understand user better, since it can provide more data that will represent the user characteristic in more reasons.

Table2.2 The example single rating and Multi-criteria ratings on a user

User\Movie	Movie A	Movie B	Movie C	Movie D	Movie E
User A (Single Criterion)	9	8	9	7	3
User A (Multi-criteria)	9(13,13,5,5)	8(11,11,5,5)	9(11,11,7,7)	7(10,10,4,4)	3(5,5,1,1)

Single rating can only define how much users prefer toward the items, but Multi-criteria ratings can represent the preference of users toward the items and also indicate the reason of users' feeling toward each criterion. For example, as you can see from table 2.2 User A is concerned on each criterion with the specific pattern which is (high, high, low, low). This can be explained that the criterion effect to a user in different level. This should help the recommendation performance because the additional information will provide more understandability than the previous one.

After obtaining rating from users, most of the recommender algorithm aim to analyze the common trend of preference among users. Suppose that there are three users as table 2.3. All users were rated the same movie, which is the movie "A". If we consider the value in single rating column, which can describe only the feeling of users toward the items, system will understand that user A and user C is the better neighbor than user A and user B. However, if we consider the values in Multi-criteria ratings column, the values of each criterion can be summarized that user A and user B are certainly adjudged to be the high quality neighbor because they have the similarity on each criterion more than user C.

Table 2.3: The example single rating and Multi-criteria ratings on multi-users

Movie / User	Movie A	
	Single Rating	Multi-criteria ratings
User A	9	9(13,13,5,5)
User B	8	8(11,11,5,5)
User C	9	9(11,7,7,11)

Multi-criteria ratings would allow the recommender systems to respond to users' individual dynamic needs in a more personalized manner and also adjust the recommendations accordingly.

Moreover, the Multi-criteria ratings obtained more information about user preference in several interesting components of items. Leveraging this additional information in the recommender system should be more advantage for users, since it can remarkably increase the accuracy of the recommendation quality.

The main goal of Multi-criteria ratings is to provide more information that helped the system to maximize user's justification. The difference between single rating and Multi-criteria ratings is the latter have more information about user and items, which can be useful for the recommendation process. In addition, the usage of Multi-criteria ratings in recommender systems can provide more benefits to their users. In order to using Multi-criteria ratings in the recommendation system the appropriate technique is required.

2.3.3 Multi-criteria ratings recommender technique

Multi-criteria ratings recommender technique is a new generation of hybrid recommendation technique based on Multi-criteria ratings, which provide additional information about user's behavior, preference, knowledge or the things that can help system to classify a user more clearly. This information is the data that affect the user's opinion based on the items' information.

Multi-criteria ratings have an important impact on many applications. Such systems, which refer as the multi-criteria recommender, were developed in numerous application domains. These include the movie recommendation [28], [29], [31], [32], the restaurant recommendation [6], the product recommendation [8], [22]-[27], [30], [34]-[36], and others [33]. For example, in food industry for E-commerce, service, cuisine and distance are three significant criteria for restaurant rating. In fact, Multi-criteria ratings for an item can provide us more precise approximations to the similarity between two users than the overall rating, since they give a good insight into why users like the item, whereas the latter can only tell us how much users like it.

Recently, Multi-criteria ratings recommender technique has been developed in various approaches based on the data information of the considered domains. For instance, in order to leverage and incorporate the Multi-criteria ratings in the recommendation system on movie domain, Adomavicius [3] proposed two approaches: (i) the aggregation function-based approach and (ii) the similarity based approach.

For (i) the aggregation function-based approach, they generate pseudo ratings for unrated items in each criterion by using the traditional recommender technique. After that, they use machine learning to generate the aggregation function base on the real Multi-criteria ratings. Finally, the predicted overall ratings were performed by using the pseudo Multi-criteria ratings as the input into the aggregation function.

About (ii) the similarity based approach; they applied the traditional Collaborative Filtering on the Multi-criteria ratings. About the similarity measurement, there are usually two different approaches to leverage Multi-criteria ratings in the similarity calculations which are "aggregating traditional similarities that are based on each individual rating", and "Calculating similarity using multidimensional distance metrics".

2.3.3.1 Aggregating traditional similarities that are based on each individual rating

This approach can use any standard similarity metric such as cosine-based (4) and calculates the similarity between users (or item based on each individual criterion)

The overall similarity can be calculated by aggregating the individual similarities in several ways such as average similarity and worst-case similarity.

- Average similarity is performed by averaging all individual similarity using the following equation:

$$sim_{avg}(u,u') = \frac{1}{k+1}\sum_{i=0}^{k} sim_i(u,u') \tag{7}$$

- Worst-case similarity is performed by using the smallest of similarity as the following equation:

$$sim_{min}(u,u') = \min_{i=0,\dots,k} sim_i(u,u') \tag{8}$$

2.3.3.2 Calculating similarity using multidimensional distance metrics

In Multi-criteria ratings, multidimensional distance metrics is one natural approach that is used to compute similarity among users. Such metrics is simple to understand and to implement. It is noted that the metric of distance and similarity are inversely related: the smaller the distance between two users, the higher the similarity. There are three steps to calculate the similarity between two users.

First, it is essential to calculate the distance between two users' rating for the same item. For this purpose, any of the standard multidimensional distance metrics can be used such as Manhattan Distance and Euclidean distance. The details of them are described in the last part of this chapter.

Second, the overall distance between two users *u* and *u'* is calculated using the following equation:

$$d_{\text{user}}(u,u') = \frac{1}{|I(u,u')|} \sum_{i \in I(u,u')} d_{\text{rating}}\left(R(u,i), R(u',i)\right) \tag{9}$$

Where *I (u, u')* denotes the set of items that both *u* and *u'* have rated. In other words, the overall distance between two users *u* and *u'* is the average distance between their ratings for all their co-rated items.

Finally, the similarity is calculated by using the following equation:

$$sim(u,u') = \frac{1}{1 + d_{\text{user}}(u,u')} \tag{10}$$

From the above two approaches, it can be concluded that co-rated items are necessary in the similarity measurement process.

Although users give their own preferences on different items, sometimes co-rated items are just a few or none. The consequence of this fact will lead to the sparsity effect problem, because they used co-rated items to perform the similarity measurement. Thus, they cannot be sure that which ones are the suitable neighbors.

Another literature, Le Roux [1] constructed a recommendation base on amalgamating the multiple criteria decision-making model as Plantié [20], Park [21] with the Collaborative Filtering technique to suggest the relative courses for graduate student(s). This should utilize especially for the non-native ones by using student(s)'s background and interested career(s).

Instead of using co-rated items, many literatures were created user profiles to reduce the sparsity effect problem such as Chapphannarungsri [4] and Rattanajitbanjong [5]. They both also focus on movie domain by using the hybrid concept to aggregate multi-criteria with multidimensional and used movies'(items) feature vector integrate with users' preference ratings. Then the system normalized it to generate the multi-criteria user vectors (profiles).

Chapphannarungsri [4] change the way of weighting by weighting all the component of each feature instead of weight only the biggest component of a feature and also focus on the frequency of selection when the users are searching for the movie. They separate characteristic of user into three vectors: (i) User Preference Vector (UPV), (ii) Selection on Movie feature Vector (SMV), (iii) Multi-Dimensional Vector (MDV). (i) The UPV represents users' opinion on a feature of movie characteristic. In order to construct the UPV, the movie feature vector needs to transform by multiplying normalized rating toward each movie. Whenever the user rates the movie, the UPV will be automatically updated for that user. (ii) SMV denotes the behavior of user when they search for the movie by accumulating the frequency of feature selection. The SMV was constructed to reduce the unsuitable weight. (iii) MDV defines the factor that might affect user preference toward the items by collecting more information about their contextual information such as date, time, and place. The MDV use multiple linear regression analysis to perform the multidimensional instead of using reduction-based.

They use UPV, SMV, and MDV in the finding neighborhood process using the distance metric to measure the different of users' characteristic among users.

Rattanajitbanjong [5] use pseudo ratings based on multi-criteria by applying Naïve Bayes. In order to find a neighbor, the pseudo CF table that derived from user profile vector with movie profile vector and also contextual information are needed.

Actually, user profiles were performed by accumulating data of user behaviors and preferences such as aspects of items, which can represent the characteristic of users.

Even though user profiles can denote the characteristic of users, but to measure the similarity among users the multi-criteria values in user profiles may not be efficient, since Multi-criteria ratings values can be the cause of SDDS problem as mentioned in the previous section.

Therefore, in order to help system to meet the appropriate data information that indicates the characteristic of users more clearly, we have to focus on applying the user profile by using Criteria-Ranking and then use the Closeness Score for the similarity measurement. Both Criteria-Ranking and Closeness Score can completely denote the significant of criteria, which is useful and capable of coping with the cause of SDDS problem, in the suitable way.

2.4 Multidimensional distance metrics

2.4.1 Manhattan Distance

The functionality of the Manhattan Distance is to calculate the distance that would be traveled to get from one data point to the other if a grid-like path is followed. The Manhattan distance is the sum of the differences of the corresponding components between two items.

The formula for this distance between a point ***X*** = (***X****1*, ***X****2*, etc.) and a point ***Y*** = (***Y****1*, ***Y****2*, etc.) is:

$$d=\sum_{i=1}^{n} |\mathbf{x}_i - \mathbf{y}_i| \tag{11}$$

Where n is the number of variables, and X_i and Y_i are the values of the i^{th} variable, at points X and Y respectively.

The difference between Manhattan distance and Euclidean distance are illustrated as the following figure:

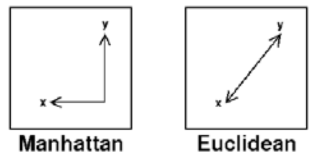

Figure2.1: The difference between Manhattan distance and Euclidean distance

2.4.2 Euclidean distance

In mathematics, the Euclidean distance or Euclidean metric is the "ordinary" distance between two points that one would measure with a ruler, and is given by the Pythagorean formula. The Euclidean space (or even any inner product space) becomes a metric space by using the formula as distance. The associated norm is called the Euclidean norm. The **Euclidean distance** between point **p** and **q** is the length of the line segment connecting them ($\overline{\mathbf{pq}}$).

In Cartesian coordinates, if **p** = (p_1, p_2,..., p_n) and **q** = (q_1, q_2,..., q_n) are two points in Euclidean *n*-space, then the distance from **p** to **q**, or from **q** to **p** is given by the following equation:

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2} = \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}. \quad (12)$$

CHAPTER III

METHODOLOGY

This chapter talks about our proposed methodology in deeply detail. All principals and also reasons are described on this chapter too.

The main idea of our methodology is that people who are thinking in the same way should be considered as the good quality neighbor. Therefore, we should separate users by using their personal feedback toward items. The similar feedback should indicate that they have the same idea.

Many researchers directly used the Multi-criteria ratings to calculate the similarity among users to classify users based on the multi-criteria recommendation system. This might not be the suitable solution to group the similar users since it may be occurred the case of “Same Distance but Different Similar” as mentioned in chapter 1. Therefore, we propose Criteria-Ranking and Closeness Score to be the appropriate solution for the process of finding the group of users who are thinking in the same way.

In order to enhance the accuracy of recommendation based on Multi-criteria ratings, we use the Criteria-Ranking to transform the user profile into Criteria-Ranking profile, and we use the Closeness Score to measure the similarity among users in terms of Criteria-Ranking profile.

To use the Criteria-Ranking and the Closeness Score technique, the system should process as the following steps. First, the user’s ratings are summarized and transformed into the user profiles. Then, each criterion on user profiles is ranked by comparing each criterion’s value with each other to transform user profile into Criteria-Ranking profile. After that, in order to measure the similarity among users in terms of Criteria-Ranking profile, we use Closeness Score to indicate the different importance of rank to find the correlation among users without the SDDS problem. Finally, we select Top-N neighbors who have higher Closeness Score and then transform Closeness Score of them into the similarity value in order to perform the recommendation.

The process of this work can be divided into these following steps:

3.1 **User Profile:** This will describe about how to create the user profile based on the multi-criteria rating.

3.2 **Ranking Profile:** This will represent how to obtain our novel user profile based on multi-criteria rating.

3.3 **Similarity Measurement:** This process is about how to measure the distance or difference among users on our novel user profile.

3.4 **Prediction Generation:** This explains the prediction process which uses the rating and also similarity to evaluate by applying on the weight average technique in order to generate the prediction.

A Recommendation Process is designed for the whole process of the recommendation system. The Recommendation Process in the experiment study is illustrated in Figure 3.1.

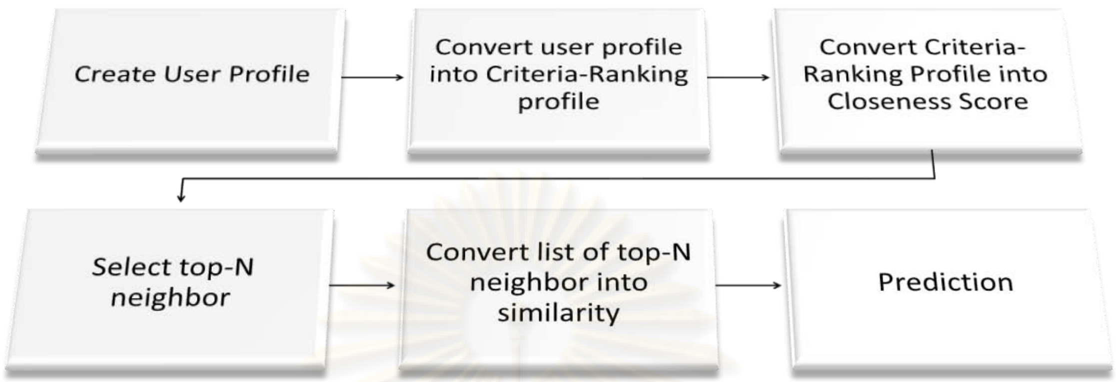

Figure 3.1: Recommendation Process

3.1 User Profile

The user profile is created by collecting the preference values of a user toward rated items and items' aspects. It comprises of overall rate, story rate, acting rate, direction rate and visual rate according to the Yahoo Movie System. The scale of rating score starts from 1 to 13, which represents the preference level from the least to the most.

To meet the goal of recommendation, the rating scores which have values more than 7 are utilized to perform the user profiles $(p_{AC_j}, \dots, p_{AC_n})$, where p_{AC_j} denotes the summation of the satisfaction's frequency for user *A* in criteria C_j The p_{AC} is calculated by using the following equation.

$$p_{ac_j} = \sum_{i \in I_c} y_i \begin{cases} y_i = 1\,; \text{if } r_i > 7 \\ y_i = 0;\ \text{others} \end{cases} \tag{13}$$

Where I_{c_j} denotes the set of user's rated movies on that criteria c_j. r_i indicate the rated value of user for the movie *i* on criteria c_j.

For example, suppose that the system obtained the multi-criteria preference data of a target user as shown in table 3.1, which displayed the rating information of a user who has user id equal to 1. This example shows his rating on each criterion for five movies. In order to compute the user profiles, the system will be processed as the following steps. First, any criteria ratings for any movie that has value greater than seven are counted as the frequency of user's criteria. Then, these frequency values are divided by using the total number of rated items of that user in order to normalize the frequency of user's criteria values in every element of the user profile. Table 3.2 displayed the result of the calculated user profile.

Table 3.1: The example of data of a user

User ID	Movie ID	Story	Acting	Direction	Visual
1	3	9	8	9	7
1	4	11	13	10	12
1	17	7	6	8	6
1	25	10	11	11	10
1	68	7	8	9	7

Table 3.2: The example of a user profile

User ID	Story >7	Acting >7	Direction >7	Visual >7
1	3/5	4/5	5/5	2/5

3.2 Ranking Profile

It should be potentially advantage for users if the system can understand more clearly about user purpose on giving their preference values to the items. In order to do that, only general user profiles may not be sufficient enough, since the SDDS problem occurs on it as mention in section 1. Thus, we propose Criteria-Ranking that can define the appropriate characteristics of users by demonstrating the level of criteria's significance. Rank order starts from one, which is the most to the least important.

For example, as the result of the table 3.2, now the system compares the values of each criterion with other criteria of that user to determine the significance level of criteria. Then the result of comparison is displayed as table 3.3.

Table 3.3: The criteria-ranking profile

User ID	Story's rank	Acting's rank	Direction's rank	Visual's rank
1	R3	R2	R1	R4

3.3 Similarity Measurement

The goal of calculating users' similarity is to form a group of good neighbors in order to generate the better recommendation.

As mentioned in chapter 1, Criteria-Ranking still needs a technique to cope with the SDDS problem instead of directly using Criteria-Ranking value to perform the measurement. Therefore, we intend to propose using the Closeness Score.

The Closeness Score is derived from the assumption that higher rank should get the more impact than the lower one. Therefore, the difference on rank 1 should be more important than rank2, rank3 and rank4 respectively.

In order to give importance for higher rank, the Closeness Value table is created to transform difference values of Criteria-Ranking on each criterion between target user and other users into Closeness Value according to the importance of rank.

After that, Closeness Values of all criteria are summarized to be the Closeness Score. This Closeness Score will represent the similarity between target user and other users. Then, system will order the Closeness Score descending to get a list of the nearest neighbors.

After the system obtains the list of neighbor completely, system will select top-N rank and then convert them into the similarity values.

In order to have clearly understandability, the process of similarity measurement is divided into three parts: Closeness Score measurement, top-N neighbors' selection and similarity value calculation

3.3.1 Closeness Score Measurement

The Closeness Value table is created based on the assumption that the similarity of higher rank is more important than the similarity of lower rank. The result of this should help us to reduce the effect of SDDS problem because the score will show the different of significant on each criterion.

Since our experimental data have four criteria, the possibility of rank order will be started from rank1, rank2, rank3, and rank4. Therefore, the difference between each rank can be only from 0 to 3 as the result on table 3.4.

Table 3.4: The difference of rank order

	Rank1	Rank2	Rank3	Rank4
Rank1	0	1	2	3
Rank2	1	0	1	2
Rank3	2	1	0	1
Rank4	3	2	1	0

This table shows the possible difference value among the ranks of our criteria.

Since the rank order is 1 to 4 and the difference of rank is 0-3, so Closeness Value table must have four columns and four rows in order to fulfill all possible Closeness Value. The column heading of Closeness Value table will represent the level of rank of target user. In contrast, row heading of Closeness Value table denotes the level of difference value of rank between target user and other users.

The value of each cell on the table 3.5 is The Closeness Value for each level of difference value (Di) between rank of other users and rank of target user (Rj) on every rank level. At cell (1, 1) in which value is 4 determines that rank of another user (R1) equals to the rank of target user (R1), so difference level is D0.

At the first place, cause our experiment have four criteria, so we set the highest values of the Closeness Score table to 4 at the most importance cell (1, 1) of the table and then continuously minus it with one for the lower rank and also the higher different values. So, Closeness Value table is displayed like table 3.5.

Table 3.5: The closeness value table

Rank of Criteria / Different Value	*R1*	*R2*	*R3*	*R4*
D0	4	3	2	1
D1	3	2	1	0
D2	2	1	0	-1
D3	1	0	-1	-2

Where *Ri* denotes rank i^{th} of target user, and *Dn* indicates the different value between the rank of target user and the rank of another user on each criterion with the difference value *n.*

The Closeness Value table is used as the lookup table. For example, target user got his/her profile as {R3 for criteria 1, R2 for criteria 2, R1 for criteria 3, and R4 for criteria 4} and another user has a profile as {R3 for criteria 1, R2 for criteria 2, R4 for criteria 3, and R1 for criteria 4}, which are displayed in table 3.6.

Table 3.6: The example of a pair of criteria-ranking profiles

User	Criteria1	Criteria2	Criteria3	Criteria4
Target User	R3	R2	R1	R4
User A	R3	R2	R4	R1

First, we concentrate on rank 1 (criteria3) of the target user. After that, difference value between target user and user A of such a criterion that contains rank1 (R1) of target user is calculated by using the following equation.

$$Dc_i = |R_t c_i - R_u c_i| \tag{14}$$

Where Dc_i denotes the different value on the criterion i, R_tc_i indicates the rank of target user on the criterion I, and R_uc_i defines as the rank of another user on the criterion i.

So, the difference value on criteria3 equal to absolute of R1 subtract with R4 that is equal to 3. After that, the closeness value from the closeness value table (table 3.5) at the cell (R1, D3) is retrieved to be the component of the closeness score. Then, the Closeness Value of rank 2, rank 3, and rank 4 are retrieved respectively as same as the rank 1. Finally, the Closeness Value of every criterion summed to be the closeness score between the user A and the target user.

3.3.2 Top N Neighbors' Selection

After finishing Closeness Score calculation, system will retrieve the list of the neighbor using the Closeness Score to order the other users respectively. Then, system will select Top-N users to be the list of the nearest neighbor for target user while Top1 of the list denoted the user who gets the highest Closeness Score among users.

Table 3.7: The example list of neighbor with their closeness score (unordered)

User ID	Closeness Score
8	7
12	10
25	6
33	10
45	8
66	6

For example, as displayed on table 3.7 user 12 and user 33 got the highest Closeness Score that is 10. So, user 12 and user 33 will be the first order of the list. Then user 45 who got Closeness Score equal to 8 will be the third order respectively as shown in table 3.8.

Table 3.8: The example list of neighbor with their closeness score (ordered)

User ID	Closeness Score	Order number
33	10	1
12	10	1
45	8	3
8	7	4
66	6	5
25	6	5

3.3.3 Similarity Value Calculation

In order to calculate the similarity between target user and other users, the list of top-N users will use to perform the similarity while the highest value(s) is top1 on the list. Their top-N positions will be converting into the similarity values using the following algorithm.

```
For (x = 1) to n
{
        Similarity value = (n – (x-1)) /n
}
```

Where x denotes the position of top-N neighbors, n represents the number of nearest neighbors.

For example, as table 3.9 this represents the converting process from the nearest neighbor list order into the similarity values.

Table 3.9: The Example of the calculation of similarity value

Neighbor name	list order	Similarity value
D	1	(5-(1-1))/5 = 1
B	1	(5-(1-1))/5 = 1
E	3	(5-(3-1))/5=0.6
C	3	(5-(3-1))/5=0.6
A	5	(5-(5-1))/5=0.2

The range of similarity is 0 to 1. While 1 denotes the most similarity that means the user and the target user have the same taste, 0 define the least similarity that means the user and the target user have the opposite taste.

3.4 Prediction Generation

The prediction is performed by using the weight average based on real rating of neighbors as the following equation.

$$Pr_i = \sum_{j=1}^{n} (r_{ij} * s_j) \,/ \sum_{j=1}^{n} (s_j) \qquad (15)$$

Where Pr_i represents predicted rating for the movie i^{th} of target user, r_{ij} denotes overall rating score of neighbor j^{th} for movie i^{th} and s_j indicates the similarity value between neighbor j^{th} and the target user.

After prediction has finished, the recommendation can be performed by suggesting the items with the high values from the list of predictable items.

CHAPTER IV

EXPERIMENTAL RESULTS AND DISSCUSION

The experiment aims to evaluate the performance of our proposed methodology by comparing with other two methods on application based implementation using Visual Basic.Net programming language and SQL Server database management system. The first one is the typical similarity-based on Multi-Criteria recommendation technique and another one is a technique which directly applied Euclidean Distance on the multi-criteria user profiles.

Before the experiment is performed, the system needed two things to support the experiment. The first one is data, and the second one is the evaluation metric.

4.1 Data

We gather data from http://movies.yahoo.com. The data consists of 200 users and 1358 movies with produce 2550 ratings. The ratings split into two different sets, first one is training set (70% of ratings) and another one is test set (remaining 30% of rating) in the experiment. Normally, Yahoo Movie System will request users to give their feedback for each movie on the overall rating and four criteria, which comprise of story, acting, direction and visual. Each user will perform rate by giving their opinion on each criterion of each movie. The meaning of each criterion can be described as followed.

- Story: this criterion is about movie story, plot, or scenario of each movie.
- Acting: this will define user preference in the actor / actress.
- Direction: this depends on the satisfaction of user to the performance of the movie director.
- Visual: this corresponds to what user saw in the movie. Such as costume of actor/actress, location, view, or place where the movie was created.

So, there are four criteria on this experiment. The possible rating values start from F to A+. After collecting data, we converted them to numerical numbers in the ways that A+ and F stand for the most and the least. The preferable values are ranging from 1 to 13.

Table 4.1 show the real database format that used to accumulate the information of user. This format comprises of user id, movie id, overall rating, story rating, acting rating, direction rating, and visual rating.

User id	Movie id	Overall rating	Story rating	Acting rating	Direction rating	Visual rating
-	-	-	-	-	-	-

Table 4.1: the format of real database

I supposed that the user who has user id "1" rated the movie that has movie id "115". He gave his opinion by rating seven for overall rating, eight for story rating, nine for acting rating, five for direction rating, and six for visual rating. Therefore, the result of this example is displayed as the table 4.2.

User id	Movie id	Overall rating	Story rating	Acting rating	Direction rating	Visual rating
1	115	7	8	9	5	6

Table 4.2: The example of real rating information

We research by using various combinations of parameters for our experiments: number of the nearest neighbors (3, 5 and 10) and number of users (100,200).

The obtaining data from the yahoo movie is not straightforward and automatic. Some manual processes are required because the purpose of the website is for the commercial only. Therefore, they not provide any service for the researcher to obtain rating information. Next, we will describe the step for obtaining the data from the site.

4.1.1 Accumulated the dataset from Yahoo! Movies Site

Figure4.1: Yahoo! Movie main page

As shown in figure 4.1, the Yahoo Movie is the site that displayed movies' information. For every movie, it will have its own page that described about the information of the movie in detail, show times & tickets, trailers & clips, cast & credits, critics review, user review … etc.

The user review is the function that the register user can give his/her opinion by rating on four criteria for each movie and also give some comment for the movie that was interested by him/her. The data can be obtaining by following these steps.

1. After we reach the main page of the site, we can be looking for the interesting movie by many ways. In this case, we are looking from the box office by click on the box office. Then the box office page will be shown as the figure 4.2.

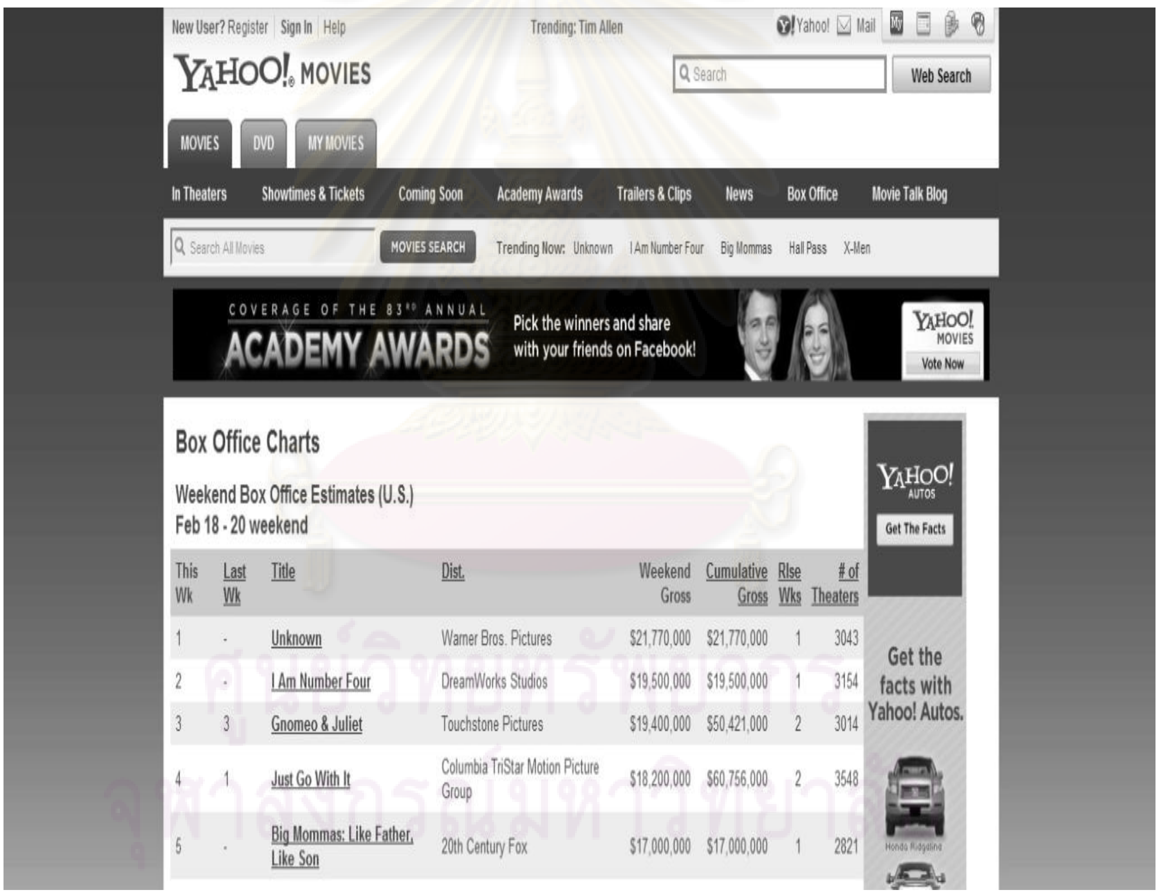

Figure 4.2: The box office page

2. As shown in figure4.2, the list of box office charts will be displayed. In this case, we choose the top of the list which has titled "Unknown". Then the page of movie "Unknown" will be displayed as figure 4.3.

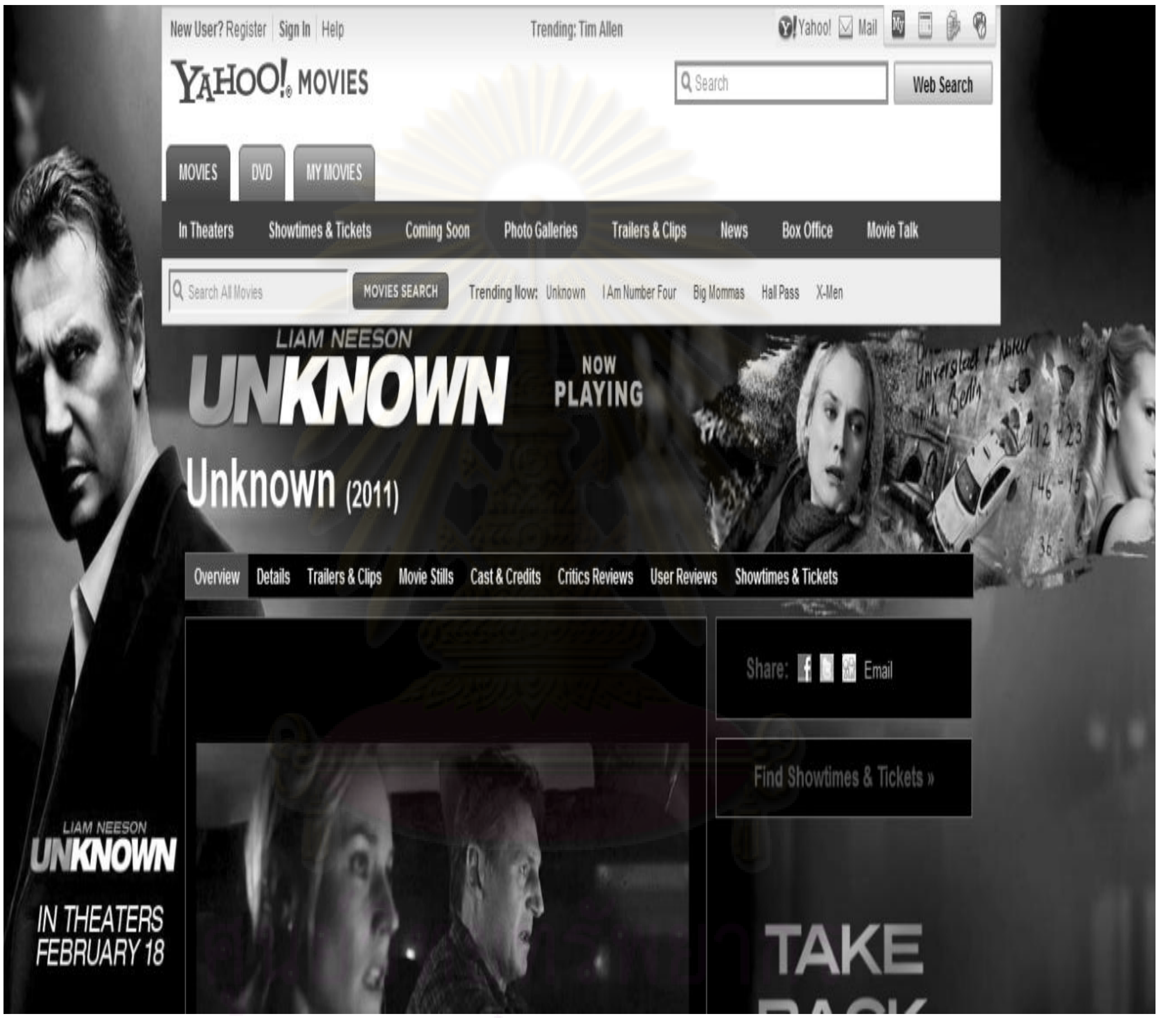

Figure 4.3: The page of Movie "Unknown"

3. This page will provide a user with the detail of such a movie. To see the rating information of register user, the simple way is clicked on the "User Reviews" link which on the middle of the page.

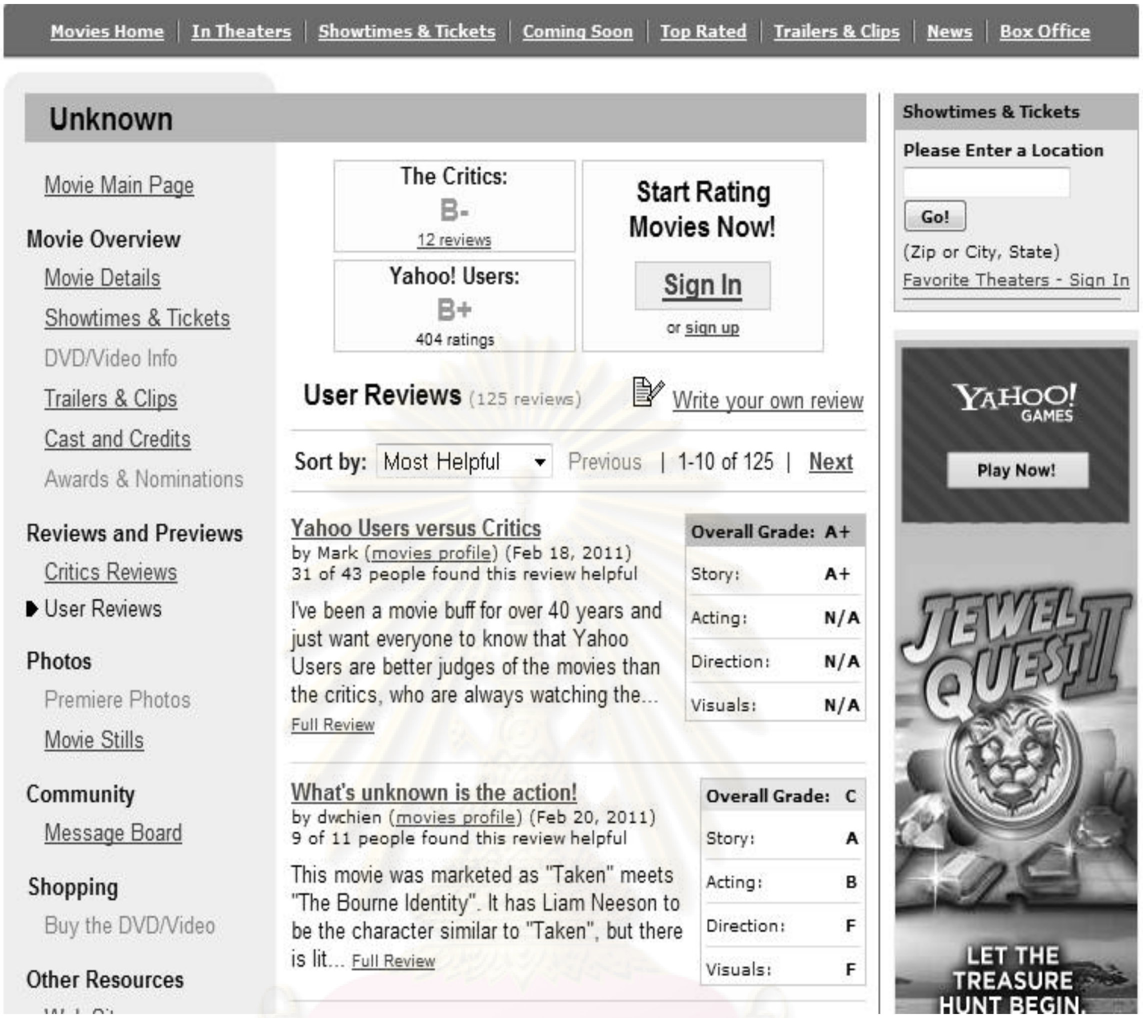

Figure 4.4: User Reviews page of “Unknown”

4. After clicking the user review link, page that has been rating information from the register user will be displayed as figure 4.4. The overall rating, story rating, acting rating, direction rating, and also visuals rating of the register user are shown as same as their comments for such a movie in this page. Therefore, we can manually collect the data from this page.

4.1.2 Transforming Data

After we finish obtaining the Multi-criteria ratings, now the Multi-criteria ratings that are the character format must be transformed into the numerical format by using the mapping table as shown in table 4.3.

Table4.3: The mapping table from character rating into numerical rating

Character rating	Numerical rating
A+ ⇨	13
A ⇨	12
A- ⇨	11
B+ ⇨	10
B ⇨	9
B- ⇨	8
C+ ⇨	7
C ⇨	6
C- ⇨	5
D+ ⇨	4
D ⇨	3
D- ⇨	2
F ⇨	1

After we finish the transformation of the numerical format, the ratings' information ready to use in the calculation of our proposed methodology.

4.2 Evaluation Metric

MAE (MEAN ABSOLUTE ERROR)

The mean absolute error is an average of the absolute errors. In statistics, the mean absolute error indicated the quality of outcomes that derived from measure how close forecasts or predictions are to the final result.

The mean absolute error is a typical measure of the predicted error in the time series analysis, where the terms "mean absolute deviation" is sometimes used in confusion with the more standard definition of mean absolute deviation. The same confusion exists more generally.

Each method is evaluated by using Mean Absolute Error evaluation metric, which can be displayed as the following equation:

$$MAE = \sum_{i \in I} |Rc_i - Rp_i| / |I| \tag{16}$$

The Rc_i indicates the actual overall rating that user had given on movie i^{th} in the test set .The Rp_i is the predicted overall rating on the test set for movie i^{th} by using our method, and I represent the set of the movie items in test set.

4.3 Experimental Result

In order to illustrate the performance of our proposed methodology on real-life data, we performed the empirical analysis on four methodologies. These four methods were implemented on the same environment to determine the accuracy and quality of each methodology.

- Rank: Our Criteria-Ranking and Closeness Score technique with original Closeness Value table (Table 3.5)
- Multi-CF: The typical multi-criteria recommendation proposed by Adomavicius [3]. It calculates the similarity between a pair of users by obtaining their overlapped rating information.
- Euclidean: A technique, which directly applies Euclidian distance on multi-criteria user, profiles.

For fairly comparison, every method above generates the prediction using weight average approach described in sub section D of section 3. After the predictions were completely done, we use MAE (Mean Absolute Error) to determine the accuracy of the predicted results. The lower MAE value represents more accuracy the result.

Table 4.4: The experimental result

Method	Neighbor =3		Neighbor =5		Neighbor =10	
	Number of user		*Number of user*		*Number of user*	
	100	*200*	*100*	*200*	*100*	*200*
Rank	1.83***	1.89***	1.95***	1.91***	2.13***	1.99***
Multi CF	2.25	2.42	2.23	2.25	2.22	2.20
Euclidian	2.98	2.37	2.72	2.46	2.28	2.43

As the empirical result from the table 4.4, rank method yield lower MAE values on every combination than the other two methods that mean rank method will produce better recommendation results than the other method.

4.4 Discussions

According to the evaluation results, ranking technique can provide higher quality of recommendations than both multi-criteria CF and Euclidian distance on multi-criteria user profile.

The reasons are the traditional multiple criteria CF need co-rated items to perform the similarity measurement that means if co-rated items is quite low or none this technique may produce the unsuitable neighbors, And Euclidian Distance use real criteria frequent values to perform the similarity measurement. Therefore, they both tend to produce poor neighbors, because the traditional multiple criteria CF faces the sparsity effect problem and the Euclidian Distance based method faces the SDDS problem.

In contrast, rank and rank up methods that focus more on the significance of criteria provide higher quality of recommendation results, since rank used similarity on user profiles instead of co-rated items to reduce the sparsity effect problem and use Criteria-Ranking profiles with Closeness Value table to reduce the SDDS problem. Therefore, the high quality neighbors and better recommendation quality will be obtained accordingly.

CHAPTER V

CONCLUSION AND FUTURE WORKS

5.1 Conclusion

The major purpose of this thesis is to propose a novel method for the movie recommender system based on multiple criteria rating.

The SDDS problem occurs when the system directly used value of user profile elements to perform the similarity measurement. To cope with such problem, Criteria-Ranking technique is proposed in this work to specify the order of criteria's significant level toward the user characteristic. Moreover, Closeness Score is used to convert rank to be the value yielding to find the similarity value between a pair of users. Consequently, the improved user profile and good quality neighbors are achieved respectively.

As the experimental results, it can be concluded that the significance of criteria is useful to help system to understand users more clearly. Since our Criteria-Ranking and Closeness Score technique can overcome the Same Distance but Different Similarity problem and provide more appropriate neighbors. Finally, the higher accuracy recommendations were achieved.

5.2 Future works

For the future work, the contextual information should be additional information. This data will represent the user's characteristic as well. How can we apply this data in order to gain better performance? Can we apply the novel methodology to improve our concept? How can we combine the new methodology with the current one to support the integration of multi-criteria with multidimensional? This combination might produce some problem, but it should increase the prediction performance. Since, the system might understand the user more clearly.

REFERENCES

[1] Le Roux, F., Ranjeet, E., Ghai, V., Gao, Y. and Lu, J., 2007. A Course Recommender System Using Multiple Criteria Decision Making Method. *ISKE-2007 Proceedings.* Atlantis Press, Chengdu, China, pp. 346-350.

[2] Maneeroj, S. and Bhattarakosol, P., 2006. Hybrid System Based on Intelligent Neighbor Formation Algorithm. *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006)*. Hong Kong, China, pp. 765-768.

[3] Adomavicius, G. and Kwon, Y., 2007. New Recommendation Techniques for Multicriteria Rating System. *IEEE Intelligent Systems.* New Jersey, USA, pp. 48-55.

[4] Chapphannarungsri, K. and Maneeroj, S., 2009. Combining Multiple Criteria and Multidimension for Movie Recommender System. *Proceedings of the International MultiConference of Engineers and Computer Scientists 2009*, Hong Kong China, pp.698-703.

[5] Rattanajitbanjong, N. and Maneeroj, S., 2009. Multi criteria Pseudo rating and Multidimensional User profile for Movie Recommender System. *Proceedings of the2009 2th IEEE International Conference on Computer Science and Information Technology*, Beijing China: pp.596-601.

[6] Z. Huang, H. Chen and D. Zeng, January 2004. Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering," *ACM Transaction on Information Systems vol.22 (1)* pp. 116-142.

[7] M. Claypool, A. Gokhale and T. Miranda, 1999. Combining Content-Based and Collaborative Filters in an Online Newspaper," *In Proceedings of the ACM SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation*, Berkeley CA, USA.

[8] G. Adomavicius and A. Tuzhilin, Jun. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, *IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6*, pp. 734-749.

[9] M. Balabanovic and Y. Shoham, 1997. Fab: Content-Based, Collaborative Recommendation, _Communications of the ACM, vol. 40, no. 3_, pp. 66-72.

[10] J. S. Breese, D. Heckerman, and C. Kadie, July 1998. Empirical Analysis of Predictive Algorithms for Collaborative Filtering, _Proceedings 14th Conference on Uncertainty in Artificial Intelligence (UAI)_.

[11] Akharraz, A., Montmain, J., Mauris, G, 2004. Elucidation and risk expressions of a movie recommendation based on a multi-criteria aggregation with a Choquet integral. IPMU 2004, _10th International Conference on Information processing and Management of uncertainty in Knowledge-Based Systems_, Genova, Italy.

[12] Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J. An algorithmic framework for performing Collaborative Filtering. _Proceedings of the Conference on Research and Development in Information Retrieval_, Berkeley, California, USA.

[13] Terveen, L., Hill, W., 2001. Beyond recommender systems: helping people help each other. _In HCI in the millennium_, J. Caroll, ed., Addison-Wesley, pp. 1-21, 2001.

[14] Lakiotaki, K., Tsafarakis, S. and Matsatsinis N., 2008. Uta-rec: A Recommender System based on Multiple Criteria Analysis. _Proceedings of the 2008 ACM Conference on Recommender Systems._ New York, USA, pp. 219–226.

[15] M. Bichler, 2000. An experimental analysis of multi-attribute auctions, _Decision Support Systems, vol. 29, no. 10_, pp. 249-268.

[16] F. Ricci and H. Werthner, 2002. Case-Based Querying for Travel Planning Recommendation, _Information Technology and Tourism, vol. 4, nos. 3–4_, pp. 215–226.

[17] P. E. Green, A. M. Krieger, and Y.Wind, 2001. Thirty years of conjoint analysis: Reflections and Prospects, _Interfaces, vol. 31, no. 3_, pp. 56-73.

[18] Y. Li, L. Lu and L. Xuefeng, 2005. A hybrid Collaborative Filtering method for multiple-interests and multiple-content recommendation in E-Commerce, _Expert Systems with Applications vol. 28_ pp. 67-77.

[19] W. Lee, C. Liu, and C. Lu, May 2002. Intelligent agent-based systems for personalized recommendations in Internet commerce, Expert Systems with Applications, vol. 22, no. 4, pp. 275-184.

[20] Plantié, M., Montmain, J. and Dray, G., 2005. Movies Recommenders Systems: Automation of the Information and Evaluation Phases in a Multi-criteria Decision-Making Process. Lecture Notes in Computer Science. Vol. 3588/2005, pp. 633-644.

[21] Park, M. H., Park, H. S. and Cho, S. B., 2008. Restaurant Recommendation for Group of People in Mobile Environments Using Probabilistic Multi-criteria Decision Making. In Proceedings of the 8th Asia-Pacific conference on Computer-Human Interaction. Seoul, Korea, pp. 114-122.

[22] N.J. Belkin, 2000. Helping People Find What They Don't Know, Communications of the ACM vol. 43(8) pp. 58-61.

[23] P. Bharati and A. Chaudhury, 2003. An empirical investigation of decision-making satisfaction in web-based decision support systems, Decision Support Systems, vol. 1051 pp. 1-11.

[24] R. Burke, 2002. Hybrid Recommender Systems: Survey and Experiments, User Modeling and User Adapted Interaction vol. 12 pp. 331-370.

[25] W. Cheetham, December 2003.Global Grade Selector: A Recommender System for Supporting the Sale of Plastic Resin, Technical Information Series, GE Global Research, TR 2003GRC261.

[26] Y.H. Cho and J.K. Kim, 2004. Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce, Expert Systems with Applications vol. 26 pp. 233-246.

[27] Y.H. Cho, J.K. Kim and S.H. Kim, 2002. A personalized recommender system based on web usage mining and decision tree induction, Expert Systems with Applications vol. 23 pp. 329-342.

[28] M. Deshpande and G. Karypis, 2004. Item-based Top-N Recommendation Algorithms," ACM Transaction on Information Systems vol. 22(1) pp. 143-177.

[29] DIMACS and LAMSADE, 2004. Computer Science and Decision Theory: Applications of Notions of Consensus, *DIMACS/LAMSADE Partnership*, (http://dimacs.rutgers.edu/Workshops/Lamsade/).

[30] W. Falle, D. Stoefler, C. Russ, M. Zanker and A. Felfernig, 2004. Using Knowledge-based Advisor Technology for Improved Customer Satisfaction in the Shoe Industry, *In Proceedings International Conference on Economic, Technical and Organizational Aspect of Product Configuration Systems*, Technical University of Denmark, Copenhagen.

[31] D. Goldberg, D. Nichols, B.M. Oki and D. Terry, 1992. Using Collaborative Filtering to Weave an Information Tapestry, *Communications of the ACM vol. 35(12)* pp. 61-70.

[32] K. Goldberg, T. Roeder, D. Gupta and C. Perkins, 2001. Eigentaste: A Constant Time Collaborative Filtering Algorithm, *Information Retrieval Journal vol. 4(2)* pp. 133-151.

[33] S. Guan, C.S. Ngoo and F. Zhu, 2002. Handy broker: an intelligent product-brokering agent for m-commerce applications with user preference tracking, *Electronic Commerce Research and Applications vol.1* pp. 314-330.

[34] U. Hanani, B. Shapira and P. Shoval, 2001. Information Filtering: Overview of Issues, Research and Systems," *User Modeling and User Adapted Interaction vol. 11* pp. 203-259.

[35] J.L. Herlocker, J.A. Konstan, L.G. Terveen and J.T. Riedl, 2004. Evaluating Collaborative Filtering Recommender Systems, *ACM Transaction on Information Systems vol. 22(1)* pp. 5-53.

[36] E. Herrera-Viedma, G. Pasi and A.G. Lopez-Herrera, 2004. Evaluating the Information Quality of Web Sites: A Qualitative Methodology Based on Fuzzy Computing with Words," *Research Group Soft Computing and Intelligent Information Systems*, University of Granada, TR. #SCI2S-2004-10.

VITAE

Jirach Duangjumpa was born on September 16th, 1984, in Bangkok Province. He obtained his bachelor’s degree in Computer Science from the Faculty of Information and Technology, Mae Fah Luang University in 2006.