A Project report on

Movie Recommendation System Using Sentiment Analysis From Micro blogging

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

Bachelor of Technology

in

Computer Science and Engineering

Submitted by

P. HEMANTH
(20H55A0517)

R. AKSHAYA
(20H55A0518)

V. SHYAMALA
(20H55A0523)

Under the esteemed guidance of

Ms.E.Krishnaveni
(Assistant Professor)

Department of Computer Science and Engineering

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

(An Autonomous Institution under UGC & JNTUH, Approved by AICTE, Permanently Affiliated to JNTUH, Accredited by NBA.)

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

2019- 2023

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Major Project Phase-1 report entitled **"Movie Recommendation System Using Sentiment Analysis From Micro blogging"** being submitted by P.HEMANTH (20H55A0517), R.AKSHAYA (20H55A0518), V.SHYAMALA (20H55A0523) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

Ms.E.Krishnaveni
Assistant Professor
Dept. of CSE

Dr. Siva Skandha Sanagala
Associate Professor and HOD
Dept. of CSE

ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Ms.E.Krishnaveni, Assistant Professor** , Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academic, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the Teaching & Non- teaching staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Mr. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

P. Hemanth 20H55A0517
R. Akshaya 20H55A0518
V.Shyamala 20H55A0523

TABLE OF CONTENTS

List of Figures

List of Tables

ABSTRACT

Recommendation systems (RSs) have garnered immense interest for applications in e-commerce and digital media. Traditional approaches in RSs include such as collaborative filtering (CF) and content-based filtering (CBF) through these approaches that have certain limitations, such as the necessity of prior user history and habits for performing the task of recommendation. To minimize the effect of such limitation, this article proposes a hybrid RS for the movies that leverage the best of concepts used from CF and CBF along with sentiment analysis of tweets from microblogging sites. The purpose to use movie tweets is to understand the current trends, public sentiment, and user response of the movie. Experiments conducted on the public database have yielded promising results.

CHAPTER 1
INTRODUCTION

CHAPTER 1

INTRODUCTION

A movie's popularity is based on the type of reviews it gets from the audience .These reviews are also responsible for affecting the choice of other users. Users are more likely to choose a movie that was preferred by most people rather than a movie that was largely disliked.

Analysing these reviews, ignoring the reviews that contain misleading information also adds to the difficulty of decision-making . Sentiment Analysis provides a solution to this problem. Sentiment Analysis facilitates a way to use NLP (natural language processing) to extract information from a textual source and classify the statement or word or document as positive or negative . It is very useful to understand the opinion of the author and indicate the user experience.

Opinion mining uses the concepts of data mining to extract and classify the opinions expressed in various online forums or platforms. This enables better understanding of the user's sentiment or feeling towards a particular subject matter.

1.1 Problem Statement

Users often face the problem of excessive available information. Recommendation systems (RSs) are deployed to help users cope up with the information explosion. RS is mostly used in digital entertainment, such as Netflix, Prime Video, and IMDB, and e-commerce portals such as Amazon, Flipkart, and eBay. In this article, we focus on RS for movies, which is an important source of recreation and entertainment in our life. Movie suggestions for users depend on Web-based portals. Movies can be easily differentiated through their genres, such as comedy, thriller, animation, and action. Another possible way to categorize the movies based on its metadata, such as release year, language, director, or cast. Most online video-streaming services , provide personalized user experience by utilizing the user's historical data, such as previously viewed or rated history.

1.2 Research Objective

The purpose to use movie tweets is to understand the current trends, public sentiment, and user response of the movie. Experiments conducted on the public database have yielded promising results.

1.3 Project Scope and Limitations

The number of choices for anything on internet is very high and it's tedious to refine most wanted data by self while searching. The scope of this proposal system includes working within numerous data, with ease.

CHAPTER 2
BACKGROUND WORK

CHAPTER 2

BACKGROUND WORK

There are three existing systems that deal with the Movie Recommendation system. A Twitter based user modeling for news recommendations: A Machine Learning and microblogging and Hybrid recommender system based on content feature relationship

2.1 Twitter-based user modeling for news recommendations

2.1.1 INTRODUCTION

framework for user modeling on Twitter which enriches the semantics of Twitter messages (tweets) and identifies topics and entities (e.g. persons, events, products) mentioned in tweets. We analyze how strategies for constructing hashtag-based, entity-based or topic-based user profiles benefit from semantic enrichment and explore the temporal dynamics of those profiles. We further measure and compare the performance of the user modeling strategies in context of a personalized news recommendation system. Our results reveal how semantic enrichment enhances the variety and quality of the generated user profiles. Further, we see how the different user modeling strategies impact personalization and discover that the consideration of temporal profile patterns can improve recommendation quality.

2.1.2 MERITS AND DEMERITS

Merits:

- Gives an efficient information regarding our desires
- User friendly

Demerits:

- The existing users not only receive information according to their social links but also gain access to other user-generated information.
- The necessity of prior user history and habits for performing the task of recommendation..

2.1.3 IMPLEMENTATION OF TWITTER BASED USER MODELING FOR NEWS RECOMMENDATION

design dimension	design alternatives
topic modeling	(i) hashtag-based, (ii) category-based (iii) entity-based
enrichment	(i) tweet-only-based enrichment or (ii) exploitation of external Web resources
temporal constraints	(i) specific time period(s), (ii) temporal patterns (*weekend*, *night*, etc.) or (iii) no constraints
weighting scheme	(i) TF, (ii) TFxIDF, or (iii) time-sensitive weighting schemes

Fig 1. Twitter based modeling framework

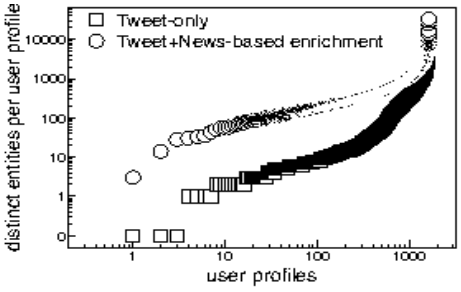
Methodology:

Data collection step

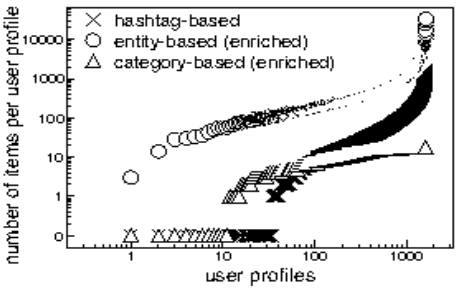
We processed each Twitter message and each news article via the semantic enrichment component of our user modeling framework to identify topics and entities mentioned in the the tweets and articles (see Section 3.1). Further, we applied two different linking strategies and connected 458,566 Twitter messages with news articles of which 98,189 relations were explicitly given in the tweets by URLs that pointed to the corresponding news article. The remaining 360,377 relations were obtained by comparing the entities that were mentioned in both news articles and tweets as well as by comparing the timestamps. In previous work we showed that this method correlates news and tweets with an accuracy of more than 70% [12]. Our hypothesis is that – regardless whether this enrichment method might introduce a certain degree of noise – it impacts the quality of user modeling and personalization positively

To validate our hypothesis and explore how the exploitation of linked external sources influences the characteristics of the profiles generated by the different user modeling strategies, we analyzed the corresponding profiles of the 1619 users from our sample. In Figure 1 we plot the number of distinct (types of) concepts in the topic- and entity-based profiles and show how this number is influenced by the additional news-based enrichment.

For both types of profiles the enrichment with entities and topics obtained from linked news articles results in a higher number of distinct concepts per profile (see Fig. 1(a) and 1(b)). Topic-based profiles abstract much stronger from the concrete Twitter activities than entity-based profiles. In our analysis we utilized the OpenCalais taxonomy consisting of 18 topics such as politics, entertainment or culture. The tweet-only-based user modeling strategy, which exploits merely the semantics attached to tweets, fails to create profiles for nearly 100 users (6.2%, topic-based) as for these users none of the tweets can be categorized into a topic. By enriching the tweets with topics inferred from the linked news articles we better understand the semantics of Twitter messages and succeed in creating more valuable topic-based profiles for 99.4% of the users.



(a) Entity-based profiles



(b) Comparison of different types of profiles

Category-based profiles abstract much stronger from the concrete Twitter activities than entity-based profiles. In our analysis, we utilized the OpenCalais [Reuters, 2008] taxonomy consisting of 18 category such as politics, entertainment or culture. The tweet-only-based user modeling strategy, which exploits merely the semantics attached to tweets, fails to create profiles for nearly 100 users (6.2%, category-based) as for these users none of the tweets can be categorized into a category. By enriching the tweets with categories inferred from the linked news articles, we better understand the semantics of Twitter messages and succeed in creating more valuable category-based profiles for 99.4% of the users. A comparison of the entity- and category-based user modeling strategies with the hashtag-based strategy (see Fig. 1(b)) shows that the variety of entity-based profiles is much higher than the one of hashtag-based profiles. While the entity-based strategy succeeds to create profiles for all users in our dataset, the hashtag-based approach fails for approximately 90 users (5.5%) as the corresponding people neither made use of hashtags nor re-tweeted messages that contain hashtags. Entitybased as well as category-based profiles moreover make the semantics more explicit than hashtag-based profiles. Each entity and category has a URI which defines the meaning of the entity and category respectively. The advantages of well-defined semantics as exposed by the category- and entity-based profiles also depend on the application context, in which these profiles are used. The results of the quantitative analysis depicted in Fig. 1 show that entity- and category-based strategies allow for higher coverage regarding the number of users, for whom profiles can be generated, than the hashtag-based strategy. Further, semantic enrichment by exploiting news articles which are (implicitly) linked with tweets increases the number of entities and categories available in the profiles significantly and improves the variety.

2.2 Hybrid recommender systems based on content feature relationship

2.2.1 INTRODUCTION

This relationship is embedded into the hybrid recommenders to improve their accuracy. We first introduce a novel method to extract the content feature relationship matrix, and then the collaborative filtering recommender is modified such that this relationship matrix can be effectively integrated within the algorithm. The proposed algorithm can better deal with the cold-start problem than the state-of-art algorithms. We also propose a novel content-based hybrid recommender system. Our experiments on a benchmark movie dataset show that the proposed approach significantly improves the accuracy of the system, while resulting in satisfactory performance in terms of novelty and diversity of the recommendation

2.2.2 MERITS AND DEMERITS

Merits:

- The experiment is more effiecient than content based and collaborative filltering
- Quick in predicting

Demerits:

- If a new user rates few or no items, the system cannot find like-minded users and therefore cannot provide recommendations.

2.2.3 IMPLEMENTATION OF DETECTION OF CYBER-AGGRESSIVE COMMENTS ON SOCIAL MEDIA NETWORKS: A MACHINE LEARNING AND TEXT MINING APPROACH

All steps of the Existing system framework are presented in Fig 2 and discussed in .

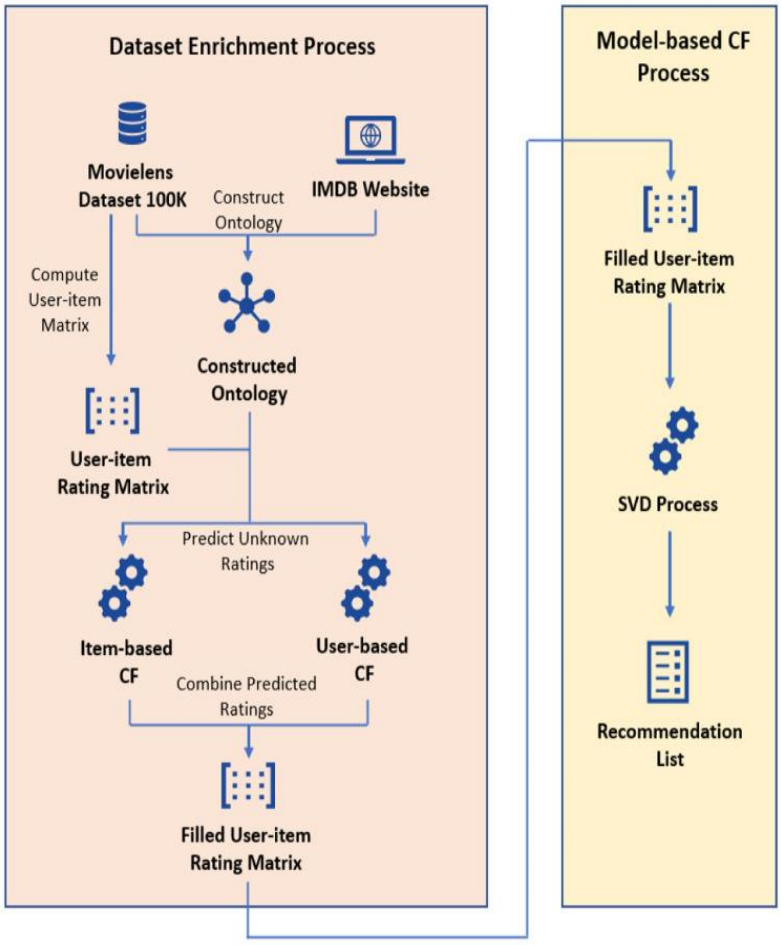
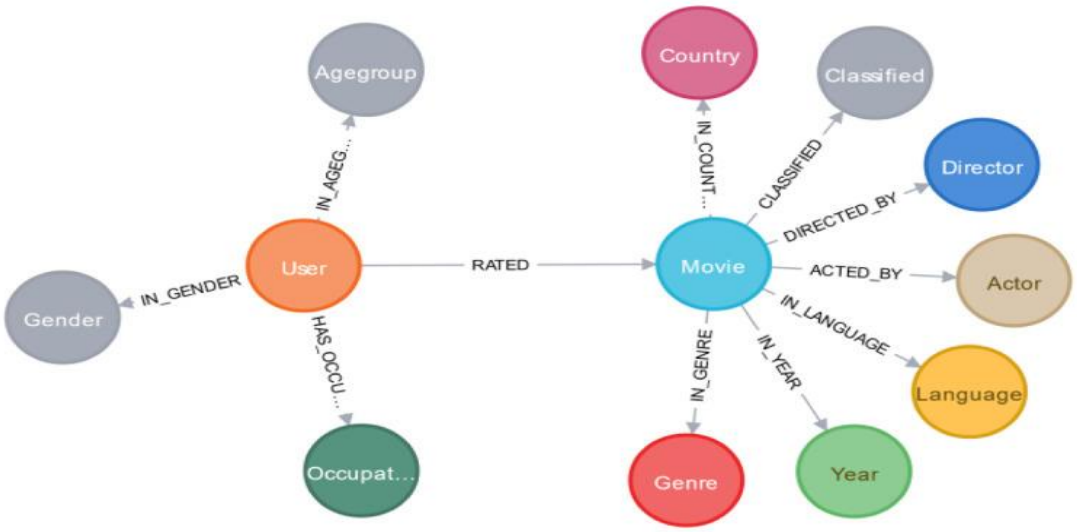


Fig 2. Hybrid recommendaor system Framework

Methodology:

Data collection step

To undertake the experiments we have collected social media text data of twitter comments from Data world website (https://data.world/crowdflower/hate-speech-identification). The original data set contains 2 attributes and 3000 instances. Attributes are 'label' and 'twitter text'. There are three types of labeled data, like hate speech, offensive and neither. We use only 1000 text documents from total data set with the ratio of offensive– 35%, neither -30%, and hate speech – 35%. We have sampled our experimental data into two sets; training data and testing data. The training dataset has contained 70% of total data and test dataset has contained rest of the 30%. Training documents contain 701-labeled examples and test.



CHAPTER 3
RESULTS AND DISCUSSION

CHAPTER 3

RESULTS AND DISCUSSION

3.1 COMPARISON OF EXISTING SOLUTIONS

Many RSs have been developed over the past decades. These systems use different approaches, such as CF, CBF, hybrid, and sentiment analysis to recommend the preferred items. These approaches are discussed as follows. A. Collaborative, Content-Based, and Hybrid Filtering Various RS approaches have been proposed in the literature for recommending items. The primordial use of CF was introduced in, which proposed a search system based on document contents and responses collected from other users. Yang et al. inferred implicit ratings from the number of pages the users read. The more pages read by the users, the more they are assumed to like the documents. This concept is helpful to overcome the cold start problem in CF. Optimizing the RS is an ill-posed problem. Researchers have proposed several optimization algorithms, such as gray wolf optimization, artificial bee colony, particle swarm optimization, and genetic algorithms [. Katarya et al. and Verma developed a collaborative movie RS based on gray wolf optimizer and fuzzy c-mean clustering techniques. Both techniques are applied to the Movielens data set and predicted a better RS. They improved the existing framework in proposing an artificial bee colony and k-mean cluster (ABC-KM) framework for a collaborative movie RS to reduce the scalability and cold start complication.

3.2 DATA COLLECTION AND PERFORMANCE METRICS

. Table 1: Details of MovieTweetings database.

Metric	Value
Ratings	646410
Unique Users	51081
Unique Movies	29228
Start Year	1894
End Year	2017

Table 2: Example of a movie entry in the modified MovieTweetings database.

Attribute	Value
MovieID	0451279
Title	Wonder Woman
Runtime	141 min
Genre	Action,Adventure,Fantasy
Director	Patty Jenkins
Writer	Allan Heinberg
Actors	Gal Gadot',Chris Pine
Rating	7.6 Massachusetts Institute of Technology in 1996.
Production Companies	DC Films,Tencent Pictures
Popularity	524.772
Language	en
Production Countries	United States of America
Budget	816303142

Table 3: Details of the modified MovieTweetings database.

Metric	Value
Ratings	292863
Unique Users	51081
Unique Movies	4515
Start Year	2014
End Year	2017

Types of noise	Example
Stop words	a, and, the, after, am
Lemma	serve, served and serving
Web links	www.tripadvisior.com
Filtering of repeating words	happyyyy, heloooo
Special Characters	!, @, #, $, %, and _

Table 4: Correlation measures between sentiment and movie ratings.

Correlation coefficient	Definition	Value
PLCC	$\sum Ni=1(xi-\bar{}x)(yi-\bar{}y)\sqrt{\sum Ni=1(xi-\bar{}x)2\sqrt{\sum Ni=1(yi-\bar{}y)2}}$	0.76
SROCC	$1-6N(N2-1)N\sum i=1d2i$	0.72
KRCC	$2(Nc-Nd)N(N-1)$	0.51

IMDb	TMDb	Recommendations from the proposed system
Justice League	**Guardians of the Galaxy Vol. 2**	**Batman v Superman: Dawn of Justice**
Batman v Superman: Dawn of Justice	Spider-Man: Homecoming	**Suicide Squad**
Suicide Squad	Logan	**Thor: Ragnarok**
Thor: Ragnarok	**Thor: Ragnarok**	**Justice League**
Spider-Man: Homecoming	**Justice League**	Warcraft
Deadpool	Pirates of the Caribbean: Dead Men Tell No Tales	**Doctor Strange**
Logan	**Doctor Strange**	**Guardians of the Galaxy Vol. 2**
Captain America: Civil War	Baby Driver	**Kong: Skull Island**
Doctor Strange	**Kong: Skull Island**	The LEGO Batman Movie
Guardians of the Galaxy Vol. 2	Life	Batman and Harley Quinn

IMDb	TMDb	Recommendations from the proposed system
Airlift	**Airlift**	**Simran**
Pink	Pink	**Fan**
Kapoor & Sons	Rustom	**Raabta**
Udta Punjab	Ghayal Once Again	**Udta Punjab**
Drishyam	Mary Kom	**Rocky Handsome**
Rustom	**Udta Punjab**	**Rangoon**
M.S. Dhoni: The Untold Story	**Force 2**	**Raabta**
Raabta	**Fan**	**Force 2**
Dear Zindagi	**Rocky Handsome**	Te3n
Rangoon	**Simran**	**Airlift**

CHAPTER 4
CONCLUSION

CHAPTER 4

CONCLUSION

RSs are an important medium of information filtering systems in the modern age, where the enormous amount of data is readily available. In this article, we have proposed a movie RS that uses sentiment analysis data from Twitter, along with movie metadata and a social graph to recommend movies. Sentiment analysis provides information about how the audience is respond to a particular movie and how this information is observed to be useful. The proposed system used weighted score fusion to improve the recommendations. Based on our experiments, the average precision in Top-5 and Top-10 for sentiment similarity, hybrid, and proposed model are 0.54 and 1.04, 1.86 and 3.31, and 2.54 and 4.97, respectively. We found that the proposed model recommends more precisely than the other models. In the future, we plan to consider more information about the emotional tone of the user from different social media platforms and non-English languages to further improve the RS.

CHAPTER 5
REFERENCES

CHAPTER 5

REFERENCES

1. F. Abel, Q. Gao, G.-J. Houben, and K. Tao, “Analyzing user modeling on Twitter for personalized news recommendations,” in Proc. 19th Int. Conf. Modeling, Adaption, Pers. (UMAP). Berlin, Germany: SpringerVerlag, 2011, pp. 1–12.

2.F. Abel, Q. Gao, G.-J. Houben, and K. Tao, “Twitter-based user modeling for news recommendations,” in Proc. Int. Joint Conf. Artif. Intell., vol. 13, 2013, pp. 2962–2966.

3.G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” IEEE Trans. Knowl. Data Eng., vol. 17, no. 6, pp. 734–749, Jun. 2005.

4.O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, “Enhancing deep learning sentiment analysis with ensemble techniques in social applications,” Expert Syst. Appl., vol. 77, pp. 236–246, Jul. 2017.

5.E. Aslanian, M. Radmanesh, and M. Jalili, “Hybrid recommender systems based on content feature relationship,” IEEE Trans. Ind. Informat., early access, Nov. 21, 2016, doi: 10.1109/TII.2016.2631138.

6.J. Bobadilla, F. Ortega, A. Hernando, and J. Alcalá, “Improving collaborative filtering recommender system results and performance using genetic algorithms,” Knowl.-Based Syst., vol. 24, no. 8, pp. 1310–1316, Dec. 2011.

7.R. Burke, “Hybrid recommender systems: Survey and experiments,” User Model. User-Adapted Interact., vol. 12, no. 4, pp. 331–370, 2002.

8. E. Cambria, “Affective computing and sentiment analysis,” IEEE Intell. Syst., vol. 31, no. 2, pp. 102–107, Mar./Apr. 2016.

9. I. Cantador, A. Bellogín, and D. Vallet, “Content-based recommendation in social tagging systems,” in Proc. 4th ACM Conf. Rec. Syst. (RecSys), 2010, pp. 237–240.

10.P. Cremonesi, Y. Koren, and R. Turrin, “Performance of recommender algorithms on top-N recommendation tasks,” in Proc. 4th ACM Conf. Rec. Syst. (RecSys), 2010, pp. 39–46.