

A hybrid approach for User profiling in Movie Recommendation

R.M. Devanshani Rasanjika, Thushari Silva

Faculty of Information Technology

University of Moratuwa

Sri Lanka

{devanshanirasanjika.uom@gmail.com, thusharip@uom.lk}

Abstract— Profiling plays an important role in personalized recommendation in research social networking platforms and variety of big data sources. Dynamic changes of user interests posted in social media and trillions of data generated through big data sources lead to significant challenges on comprehensive user profiling. Existing approaches mostly rely on content-based analysis or item-based rating mechanisms. Most of these approaches rely on single sources and largely overlook the power of integration of multiple data sources especially social media. In light of this deficiency, this research proposes an integrated novel approach to automatically capture dynamic user interests. The proposed method facilitates automated learning of user profiles based on the proposed approach which integrates preprocessing and electronic symbol extraction model, opinion-confidence mining hybrid model, feature extraction model and profile data aggregation weighted k-means clustering model and analysis of social data. According to our experimental analysis the proposed method outperforms existing benchmark methods.

Keywords—Comprehensive profiling, big data, opinion mining, Feature extraction

I. INTRODUCTION

Movie recommendation one of important application which comes under personalized recommendation. Generation of effective user profiles controls effectiveness of movie recommendation. Thus, in this research we propose a hybrid approach for generation of user profiles in the context of movie recommendation. Therefore, mining social media contents to derive opinions as well as mining interests, the movie related content is focused.

Personalized recommendation is crucial for effective recommendation application which involve users. Dynamic changes in user interest have posed significant challenges in user profiling as they are not clearly visible in personal websites or institution level databases. Thus, research on personalization and hence user profiling has gained significant attention recently.

With the proliferation of social data as a result of high level user involvement in social media, has demanded intelligent way of user modeling which are hard to achieve via the analysis of one single data sources. Social data in social networks consist of large volume of information about user interests which helps to identify the data about user interest domain. Hence use of a single data source marks limitation

¹<https://www.facebook.com/>

²<https://www.imdb.com/>

in content-based user profiling as well as collaborative filtering approaches.

Profiling using social network-based methods aim to model users by analyzing the user's behavior in social networks such as comments, threads from group of friends, likes posted to social networks and demographic information. Users in social networks upload contents, update status, share ideas by generating billions of data in every second which provides a methodology for more comprehensive profiling in which each user is represented with a social network profile.

The proposed method of user profiling integrates preprocessing and electronic symbol extraction model, opinion-confidence mining hybrid model, feature extraction model and profile data aggregation weighted K-means clustering model with novel algorithms to construct comprehensive user profiles. The proposed method integrates the data from Facebook¹ and IMDB² big data sources in which IMDB² aggregate the representative features of movies to construct user profiles and for the personalized recommendation.

The rest of the paper is organized as follows. Literature review reports the review on existing methods in profiling. Following the literature review reports the proposed method for comprehensive user profiling including the design, proposed models, algorithms and experimental results and analysis, benchmark analysis and finally the research offers concluding remarks.

II. LITERATURE REVIEW

Profiling is a set of information representing an individual entity via related rules, features, interests, behaviors and preferences [2,3]. A profile can consist of static data (e.g. location) or dynamic data (e.g. interests) which is vary depending on the application [4]. The accuracy of a profile which represents an individual character depends on how the information is gathered and organized to create the summarization and description which reflects the character [5,6].

There are two ways to profile a character fundamentally which is called as explicit or implicit information gathering. In the explicit profiling, information and interests regarding the entity is provided explicitly to the system. Explicit profiling is a static method which is applicable until the entity changes their interest and preference parameters [7,14]. Implicit profiling analyzing the behavior pattern of the entity to determine the interests in which the accuracy depends on the volume of generated data through user interaction with the system [8]. Profiling approaches can be categorized based on content-based and collaborative methods.

A. Content based method

In Content based method the current behavior of the entity is predicted from his past behavior [2,3,9]. In this approach system selects the item which has high content correlation with the profile.

1) Vector space model

Vector space model is a statistical term-based technique which represent the content as vectors of weighted terms [3]. The profile of the entity is represented as weighted keywords which summarize the interests and preferences [3]. Weights of the keywords indicate the importance of the term which identifies the interest significantly which is the frequency of the term appears in the content [9]. The dimensions of the vector are equivalent to the number of terms that reflects the entity's interests and preferences significantly [10]. Boolean, Term-Frequency Inverse Document Frequency (TF-IDF) and Term-Frequency (TF) are the methods that used to derive the weighted representation of the terms.

2) Latent Semantic Indexing (LSI)

LSI is a statistical term-based technique which examines the patterns and relationships between terms within the content [10,14]. It retrieves the relevant documents even though the profiles do not have any common terms. In this technique the document is considered as a word by document matrix which is computed by using Term Frequency Inverse Document Frequency method (TF-IDF).

B. Collaborative based method

Collaborative based filtering method is based by assuming that the entities who belong to the same group (e.g. same age, gender, location) has a similar behavior and have similar profiles [2,3,9]. Collaborative filtering method ignores the item content like content-based filtering method and it does recommendation based on item ratings of similar entities [3]. The main drawbacks of collaborative filtering are the lack of available ratings which is called as sparsity and lack of number of ratings from a new user which is called as first rater problem [8].

1) Memory based and Model based techniques

In this technique the profile of the entity consists of ratings of the users for a set of items. When the ratings for item set increases it increase the accuracy of the profiling. Memory based technique predict the item's rating given from a particular user based on the past ratings of similar users. In model-based methods clustering and classification techniques are used to make item rating predictions [10,15].

C. Hybrid method

Hybrid method combines both content and collaborative based methods by aggregating the advantages of both methods. In Hybrid profiling a new user is assigned to a group by using collaborative filtering and enhances the profile using content-based filtering with a more accurate description of interests and preferences [8,9].

D. Ontology based user profiling

Most existing ontology-based models focus on only consider the capturing of interested information of the entity [11]. Ontology based techniques initiate the semantic web approach which consist of well-defined meanings for the information in the predefined ontology. Predefined ontology has term descriptions and their interrelationships with other terms which is able to retrieve more accurate and relevant information by searching based on keywords [12]. Ontology based modelling identifies the correct meaning of a term with respect to its context and match it with a semantic web resource which provides more accurate definition of the specified meaning and it form a profile ontology from set of keywords [13].

III THE PROPOSED HYBRID APPROACH FOR USER PROFILING

Mining of social media data is a methodology for user profiling in the context of movie recommendation which model users by analyzing the user's dynamic behavior in Facebook¹ such as comments, watched movies, threads from group of friends, likes and heart reactions posted to social networks and demographic information. Furthermore, the proposed methodology aggregates the movie genres, actors, director, movie reviews, movie plot, IMDB rating and IMDB votes from IMDB² to characterize a user. Initially the proposed approach crawl social media data from big data sources such as Facebook¹ using Facebook¹ graph API and aggregates the representative features of user watched movies from big data sources namely IMDB² movie database using IMDB² API. Large volume of unstructured and structured data is extracted and processed through preprocessing and electronic symbol extraction model, feature extraction model, opinion-confidence mining hybrid model and profile data aggregation weighted K-means clustering model to generate comprehensive user profiles. The following figure 1 illustrates the proposed hybrid approach for user profiling in movie recommendation.

¹<https://www.facebook.com/>

²<https://www.imdb.com/>

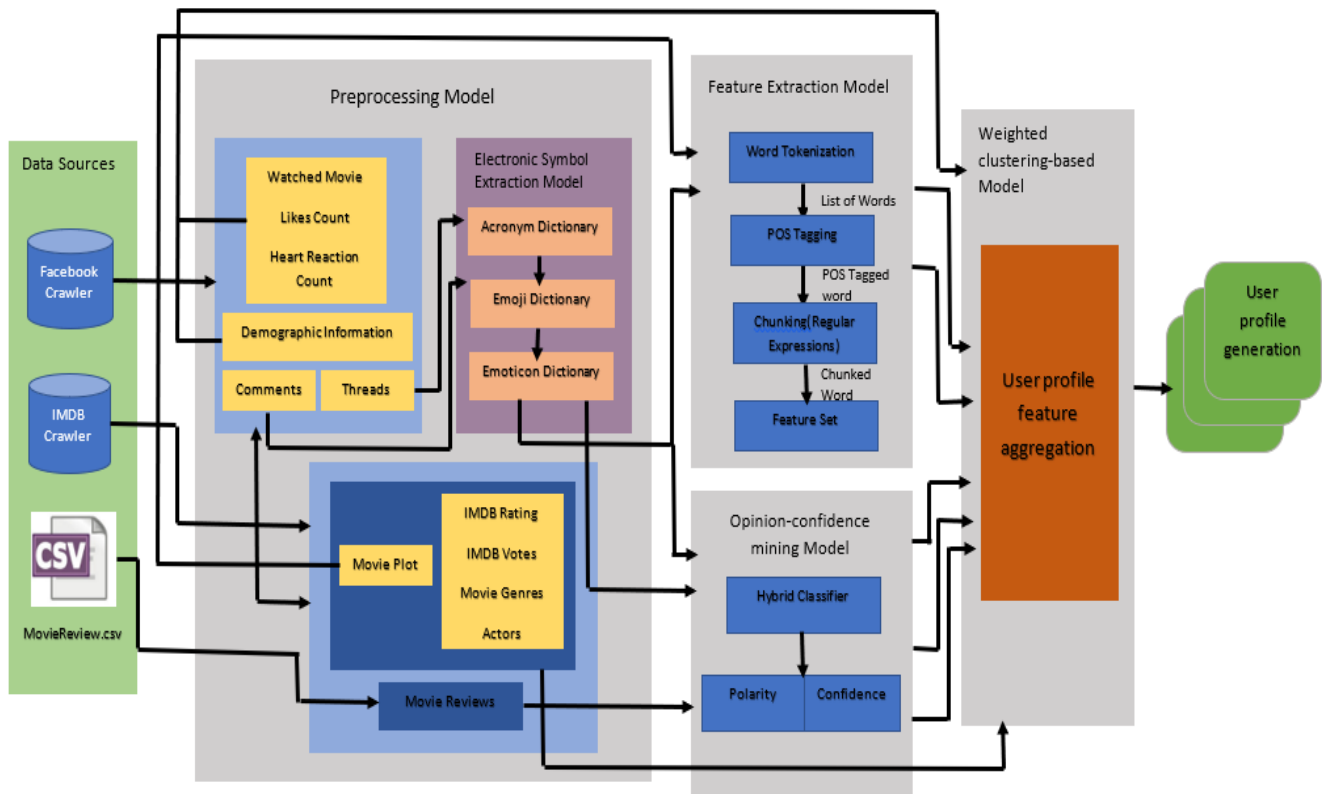


Figure1. Proposed hybrid approach for user profiling in movie recommendation

A) Preprocessing model for user profiling

The major tasks included in preprocessing model are processing of Facebook¹ and IMDB² data to extract representational features of the user profiles. The likes, heart reactions posted in Facebook¹, demographic information of users, watched movies and representative features of user watched movies namely movie plot, IMDB rating, IMDB votes, movie genre, actors and movie reviews are processed through preprocessing model. Furthermore, Facebook comments and threads from group of friends are extracted and processed through electronic symbol extraction model.

1) User profile feature generation with electronic symbol extraction model

Comments and threads extracted from Facebook¹ consist of emojis, emoticons and acronyms. The proposed Electronic symbol extraction model includes three sub processes to replace emojis, emoticons and acronyms by defining dictionary-based algorithms with natural language. Dictionary based algorithm for electronic symbol extraction model is given below.

Algorithm

¹<https://www.facebook.com/>

²<https://www.imdb.com/>

Method Tokenize (Sentence s) {

Create $E_i(x, y)$, $i=1,2,3$

where $E_1(S_j, d_j) = \text{emoticon_dictionary}$,

$E_2(S_j, d_j) = \text{emoji_dictionary}$,

$E_3(S_j, d_j) = \text{acronym_dictionary}$

For all word $w_1 \in \text{Sentence } s$ do {

$i=1$

Select $E_1(S_j = w_1)$ where symbol $S_j = S_1, S_2, \dots, S_n$

And dictionary $d_j = d_1, d_2, \dots, d_n$

Replace (S_j, d_j)

}

For all word $w_2 \in \text{emoticon_replaced_sentence } s$ do {

$i=2$

Select $E_2(S_j = w_2)$ where symbol $S_j = S_1, S_2, \dots, S_n$

And dictionary $d_j = d_1, d_2, \dots, d_n$

```

    Replace (Sj, dj)
    }

    For all word w3 ∈ emoji_replaced_sentence s do {
        i=3
        Select E3(Sj=w3) where symbol Sj=S1, S2,..... Sn

        And dictionary dj=d1, d2,..... dn

        Replace (Sj, dj)
    }
}

```

B) Feature extraction model for user profiling

Filtered comments from Facebook¹ and preprocessed movie plots from IMDB² for user watched movies are passed to the feature extraction model to apply part of speech tagging and chunking techniques in natural language processing to extract representative features. Part of speech tagged words go through the process of chunking by grouping words in to syntactically correlated phrases(chunks). Part of speech tagged words are processed by defining rules using regular expressions and extracted features are obtained. Therefore, it increases level of characterizing a given entity than existing algorithms because it measures the semantic correlation of words. The proposed Feature extraction algorithm for both filtered comments and movie plots is given below.

Algorithm

```

Method POS_tagged(sentence) {

Grammar:NP: {

Rule 1: (<JJ?><NN?>(<NN?>?) (<VB?>?))

Rule 2: (<NN?><VB?><JJ?><NN?>)

Rule 3: (<JJ?><NN?>)

Rule 4: (<NN?><VB?>)

Rule 5: (<VB?><NN?><VB?>)

Rule 6: (<NN?><NN?>(<NN?>?)

Rule 7: (<RB?><JJ?><NN?>)

Rule 8: (<JJ?><VB?>)

}

Chunkparser=Regexparse(grammar)

Tree=chunkparser(pos_tagged_sentence)

    For leaf in leaves(tree) {

```

¹<https://www.facebook.com/>

²<https://www.imdb.com/>

L=length of leaf

```

    }

    For n in 0 to L {

        Sentence=sentence+ leaf[n][0]

    }

}

```

C)Opinion-Confidence Mining Hybrid model

Filtered data namely filtered comments and threads from group of friends crawled from Facebook¹ and preprocessed movie reviews from IMDB² for user watched movies are go through the opinion-confidence mining hybrid model which classifies the filtered sentiments based on its polarity and obtained the confidence of the polarity through hybrid model which is implemented by combining the predictions of the Naïve Bayes(NB), Support Vector machine(SVM), Multinomial Naïve Bayes(MNB), Bernoulli Naïve Bayes(BNB) and Logistic Regression(LR) classifiers. Bag of bigrams model is used as the methodology for classification where frequency of word bigrams is used as the feature vectors for training the hybrid model. The proposed algorithm for opinion-confidence mining hybrid model is given below.

Algorithm

Prediction of class label y via majority voting of each classifier C_j

$$y = \max \sum_{j=1}^m (c_j(x) = i)$$

Confidence C via combining voting classifiers C_j

$$C = \frac{\max \sum_{j=1}^m (c_j(x) = i)}{\sum_{j=1}^m (c_j(x))}$$

Where C_j=NB, SVM, BNB, MNB, LR

D)User profile data aggregation-Weighted K-means Clustering model

Feature selection is done by setting weights for each feature and the selected features of user profiles namely likes and heart reactions posted in Facebook¹, demographic information (e.g. age, gender), filtered comments and thread polarity confidence are considered as a combined feature which has equal weighted value for the features. The selected features of user watched movies namely

actors, director, IMDB rating, IMDB votes are considered as a combined feature which has equal weighted value for the features. Then extracted features of comments and movie plot from feature extraction model, watched movies and movie genres are considered as individual features by setting different weights. Priorities for the features are set based on the importance to the user profiling algorithm.

Algorithm

Method load (Feature Set S) {

For all features f in Feature Set S {

Range (x=1, x=6)

$y = \text{Aggregate}(f(x))$
}

Range (x=7, x=9)

$V(x) = \text{Numeric_FeatureVector}(y, f(x))$

Compute feature vectors by associating a weight W_j with the clustering model M_j to predict the cluster C

Range (0, i) {

$$C = \sum_{j=1}^m (W_j * V(x))$$

}

Where i = 0, 1,, n

$C_{\text{optimum}}, \text{Max}(\text{silhouette-coefficient}) = \text{Silhouette method}(C)$

$C_{\text{optimum}}, \text{Min}(\text{cluster-error}) = \text{elbow method}(C)$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A) Electronic symbol extraction model

Filtered comments from Facebook¹ are processed through electronic symbol extraction model to generate appropriate meanings for emojis, emoticons and acronyms. The following example illustrates the experimental results of the proposed algorithm.

e.g.

Sentence:

“another wonderful movie to watch <3”

Emoticon replaced sentence:

¹<https://www.facebook.com/>

²<https://www.imdb.com/>

another wonderful movie watch love

Sentence:

“gr8 acting gr8 story”

Acronym replaced sentence:

great acting great story

B) Feature extraction model for user profiling

1) NLP based feature extraction algorithm

Representative features of movie plot from IMDB² and comments from Facebook¹ is extracted by using part of speech tagging and chunking techniques as an experimented approach. It measures the semantic correlation of words by defined rules using regular expressions. The following example illustrates the experimental result of the tested approach.

e.g.

Sentence:

“Good Story with lots of mysteries love the story but disappointing about the production Boring start by the way acting is good”

POS tagged sentence:

[('Good', 'JJ'), ('Story', 'NNP'), ('with', 'IN'), ('lots', 'NNS'), ('of', 'IN'), ('mysteries', 'NNS'), ('love', 'VBP'), ('the', 'DT'), ('story', 'NN'), ('But', 'CC'), ('disappointing', 'VBG'), ('about', 'IN'), ('the', 'DT'), ('production', 'NN'), ('Boring', 'NNP'), ('start', 'NN'), ('by', 'IN'), ('the', 'DT'), ('way', 'NN'), ('acting', 'NN'), ('is', 'VBZ'), ('good', 'JJ')]

Chunked phrases (extracted features):

Good Story

mysteries love

disappointing production

Boring start

acting is good

2) Feature extraction using Rapid Automatic Keyword Extraction Algorithm(RAKE)

RAKE extract the keywords by generating a RAKE score for each keyword. The following example illustrates the experimental result of the tested approach.

e.g.

Sentence:

“Good Story with lots of mysteries love the story but disappointing about the production Boring start by the way acting is good”

Keywords:

[('boring start', 4.0), ('good story', 3.0), ('story', 1.5), ('good', 1.5), ('lots', 1.0), ('mysteries', 1.0), ('love', 1.0), ('disappointing', 1.0), ('production', 1.0), ('acting', 1.0)]

According to the results obtained from the above-mentioned experiments RAKE does not compute semantic correlation between words of the extracted features. Therefore, the proposed NLP based feature extraction algorithm address these shortcomings by defining rules using regular expressions.

C) opinion-confidence mining hybrid model

Natural language processing techniques are used to create bigram and trigram feature vectors in which bag of bigram model and bag of trigram model are used as the methodology for classification. Experimental results are obtained by using bag of bigram model and bag of trigram model to show that the accuracy of the proposed hybrid model. Experimental results are compared with the Naïve Bayes(NB), Support Vector Machine(SVM), Bernoulli naïve Bayes(BNB), Multinomial Naïve Bayes(MNB) and Logistic Regression(LR) single algorithms and ensemble classifiers namely Decision Tree classifier(DT) and Random Forest classifier(RF) with Hybrid Model(HM) in terms of accuracy level. Comparison between above mentioned tested approaches is shown in the table1 given below.

TABLE 1. comparison of experimental result

Classifiers	Bigram Model(Accuracy %)	Trigram Model(Accuracy %)
HM	71.70971709717097	71.61746617466174
NB	66.60516605166052	71.27921279212792
SVM	71.64821648216481	71.58671586715867
MNB	70.84870848708486	71.37146371463714
BNB	71.61746617466174	71.61746617466174
LR	71.69421894218943	71.61746617466174
DT	71.49446494464945	71.40221402214021
RF	71.69421894218943	71.52521525215252

According to the results obtained from the bag of bigram model and bag of trigram model accuracy level is maximized in the hybrid model when tested with bag of bigram model. Therefore, the proposed opinion-confidence mining hybrid model used bigram feature vectors for classification.

d) User profile data aggregation-Weighted K-means Clustering model

Elbow method and average silhouette method is experimented to determine optimum number of clusters by changing the K value ranges from (2,30).

1) Elbow method

¹<https://www.facebook.com/>

²<https://www.imdb.com/>

Elbow method is experimented to determine the optimum number of clusters for user profile data by increasing K value. Sum of Squared Errors(SSE) is goes down rapidly when increasing the K value and then it reaches the elbow after that distortion goes down slowly [16]. The K value which is at the elbow of the curve is chosen as the optimum number of clusters where the elbow represents the diminishing return [16]. The following figure 2 shows that the graph of the elbow method for user profile features is optimized at when K=8 where cluster error is minimized.

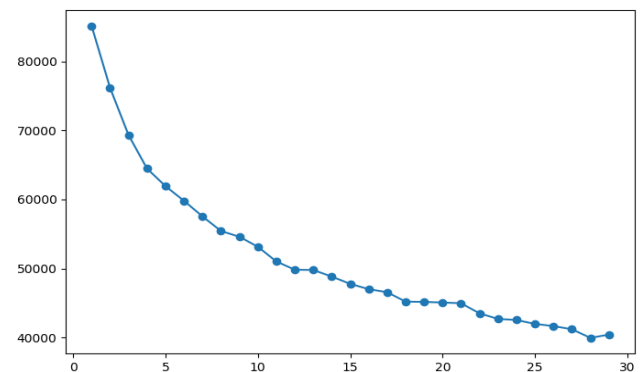


Figure2. Elbow method graph for user profiling

2) Average silhouette method

Average silhouette method is experimented to determine optimum number of clusters when increase K value. The cluster which has the highest Silhouette coefficient is chosen as the optimum K value to define the number of clusters in the profiling algorithm [16]. Experimental results from silhouette method is shown in the table2 given below.

TABLE 2. experimental result from silhouette method

No. of clusters	User profiling (Silhouette Coefficient)	Movie profiling (Silhouette Coefficient)
2	0.3584944927367878	0.1196832210731034
3	0.36663272429311	0.13212724101066775
4	0.2521660881011146	0.1605299776140723
5	0.2701484325338426	0.17482185992674484
6	0.04568216681816124	0.1329604151541994
7	0.020904709032240187	0.1702109982382697
8	0.03049309697963994	0.1640209647197902
9	0.03205222998816212	0.18483336261870634
10	0.008522490830782053	0.1687250534252433
11	0.01889617645008163	0.16401342154246548
12	0.020197611699678653	0.17018625600246792
13	0.03099162357080225	0.15960525651882304
14	0.03179073205031598	0.16734918297483758
15	0.03003265748888705	0.16283942975700894
16	0.02666462261003181	0.11503798490050504
17	0.032464422945104536	0.17028069320557718

18	0.032976073891183455	0.18751973381376202
19	0.03283846183924377	0.17452444667078668
20	0.03041842385058277	0.1817488251579319
21	0.0360282968086087	0.18967544882124274
22	0.034299657453298577	0.1891285955129887
23	0.03709483263629965	0.18110345907877995
24	0.03970894215322266	0.1829696242397622
25	0.041915692215075476	0.18420079073590667
26	0.04489029265659179	0.19902940664637866
27	0.048031671795858086	0.18711412276407013
28	0.0461778614919711	0.19392658804887228
29	0.03817673123675763	0.1920919121426588

Average silhouette method illustrates that K=9 is the optimum value when finding the optimum cluster centers from user profile features where silhouette coefficient is maximized.

According to the experimented results the above mentioned two tested approaches can use to determine optimum number of clusters accurately.

V. BENCHMARK ANALYSIS

According to existing literature content-based method recommend the items which has high correlation with the content. The major drawback of content-based method is items are selected based on the frequency of the term appears in the document.

Collaborative filtering method ignores the content and it does recommendation based on item ratings of similar entities. Therefore, past item ratings of similar users are used to make predictions of the ratings of new user. The major drawback of collaborative filtering method is item ratings of similar entities are used to predict the rating of new user due to first rater problem.

Hybrid method aggregates the advantages of both content-based method and collaborative filtering method in which new user is assigned to a group using item ratings of similar entities and the item which has high content correlation enhance the modelling of user interests and preferences. The major drawback of hybrid method is features including demographic information, numerical data and derived data are not used for recommendation and it does not compute semantic correlation of words.

Ontology based profiling forms a profile ontology from set of keywords which identifies the correct meaning of the keywords and match it with a semantic web resource. Unlike content based, collaborative filtering and hybrid method, ontology-based profiling identifies the correct meaning of keywords to construct the profile ontology. Furthermore, major drawback of ontology-based profiling is features other than keywords such as demographic information, numerical data and derived data are not used for recommendation.

In the proposed approach the experiments on User profiling used a Facebook¹ dataset of 40 users for 75

¹<https://www.facebook.com/>

watched movies by users and IMDB² movie dataset of 1930 movies and 15610 movie reviews for corresponding movies including user watched movies for comprehensive user profiling and personalized movie recommendation and the dataset covers 17 movie genres. In the user profile dataset each user is represented with 15 attributes that are preferred movie genres, watched movies, gender, age, comments, likes count, heart reactions count, director, actors, IMDB votes, IMDB ratings, movie plot, movie review polarity confidence of watched movies, thread polarity confidence and cluster of weighted K-means clustering model.

Therefore, proposed hybrid approach for user profiling in movie recommendation overcome the limitations of existing approaches by analyzing above mentioned attributes. All the algorithms trained and tested using same user profile dataset.

The proposed user profile feature generation electronic symbol extraction model creates emoji dictionary covering 40 Unicode characters, emoticon dictionary covering 25 electronic symbols and acronym dictionary covering 60 informal language meanings to generate appropriate natural language meanings. The proposed electronic symbol extraction model overcome the limitations of existing approaches such as existing approaches eliminate the emojis, emoticons and acronyms through preprocessing. Therefore, the proposed model help to enhance the opinions of users in social networks.

The proposed feature extraction model extracts the representative features which has semantic correlation related to context when compared with other tested approaches such as rapid automatic keyword extraction algorithm(RAKE). Therefore, the NLP based proposed feature extraction model compute 8 rules to create the feature set. The proposed model is intelligent in feature generation than existing frequency-based approaches such as RAKE.

The proposed opinion confidence mining hybrid model combines support vector machine, Naïve Bayes, Bernoulli Naïve Bayes, multi nominal naïve Bayes and logistic regression 5 classifiers which computes maximum accuracy level (71.70%) when compared with single algorithms. Therefore, the proposed model improves the intelligence of classification and computes the confidence. The corpus of 9300 positive feats and 8600 negative feats are used and 3/4 dataset is used as training dataset and 1/4 is used as testing dataset. Furthermore, the proposed model computes the confidence of reviews which is the degree of polarity level. So, we can conclude that proposed hybrid model overcome the shortcomings of existing approaches because existing approaches has the ability to classify a given review either positive or negative but intelligent profiling needs to identify the confidence of the polarity when characterizing a given entity.

The proposed profile data aggregation weighted K-means clustering model computes weights for each attribute in

user profile dataset. Weighted clustering-based model enhance the cluster generation mechanism because the weights are assigned to each attribute based on priority and feature aggregation is done based on weighted values. Therefore, the proposed model overcome the limitations of existing approaches such as clustering based on item ratings.

Therefore, the proposed hybrid approach integrates the above mentioned four models namely preprocessing and electronic symbol extraction model, feature extraction model, opinion-confidence mining hybrid model and profile data aggregation weighted K-means clustering model which provides a hybrid approach for user profiling in movie recommendation.

VI. SUMMARY AND CONCLUSION

Characterizing dynamic behavior of a user is a challenging issue in recommendation domain. This paper proposed a novel approach for characterizing dynamic behavior and interests of the user for the generation of user profiles in the context of movie recommendation. I have developed a system for user profiling which aims to discover personalized user interests by aggregating big data sources namely Facebook¹ and characterizing movies including user watched movies by their features from big data sources namely IMDB² covering demographic information, numerical data and social information.

Accordingly, the proposed hybrid method constructs semantically rich user profiles through preprocessing and electronic symbol extraction model, feature extraction model, opinion-confidence mining hybrid model and profile data aggregation weighted k-means clustering model which is capable of handling dynamic interests of the user.

The business use of this research is that commercial organizations can apply this approach to model user's dynamic behavior and characterize product features for the recommendation and it provides personalized user experience for the customers which increase the revenue of the organization.

REFERENCES

- [1] Chia-Chuan Hung, Yi-Ching Huang, Jane Yung-jen Hsu, David Kuan-Chun Wu, Tag-Based User Profiling for Social Media Recommendation
- [2] G. Araniti, P. D. Meo, A. Iera and D. Ursino (2003). Adaptive controlling the QoS of multimedia wireless applications through user profiling techniques, *IEEE Journal on selected areas in communication*, 21(10), pp. 1546-1556.
- [3] T. Kuflik and P. Shoval (2000). Generation of user profiles for information filtering-research agenda, *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 313-315.
- [4] M. J. Martin-Bautista, D. H. Kraft, M. A. Vila, J. Chen and J. Cruz (2002). User profiles and fuzzy logic for web retrieval issues, *Soft Computing (Focus)*, 15(3-4), pp. 365-372.

- [5] European Telecommunications Standards Institute (ETSI) (2005). *Human Factors (HF); User Profile Management*, pp.1-100.
- [6] S. Henczel (2004). Creating user profiles to improve information quality, *Factiva*, 28(3), p. 30.
- [7] C. Gena (2005). Methods and techniques for the evaluation of user-adaptive systems, *The Knowledge Engineering*, 20(1), pp.1-37.
- [8] D. Poo, B. Chng and J. M. Goh (2003). A hybrid approach for user profiling, *Annual Hawaii International Conference on System Sciences*, 4(4), pp. 1-9.
- [9] D. Godoy and A. Amandi (2005). User profiling in personal information agents: a survey, *The Knowledge Engineering Review Journal*, 20(4), pp. 329-361.
- [10] S. Steward and J. Davies (1997). User profiling techniques: a critical review, *British Computer Society, BCS-IRSG Annual Colloquium on IR Research*, pp. 1-22.
- [11] Susan Gauch (2007). *Ontology-Based User Profiles for Personalized Search*.
- [12] Mohammed Abdellah Alimam, Yasyn Elyusufi, Hamid Seghioer (2017). *Ontology-Based User Profiling for Personalized Access to Information within Collaborative Learning System*.
- [13] Anett Hoppe, Automatic ontology-based User Profile Learning from heterogeneous Web Resources in a Big Data Context.
- [14] Manoj Kumar, D.K. Yadav, Ankur Singh, Vijay Kr. Gupta, A Movie Recommender System: MOVREC, *International Journal of Computer Applications* (0975 – 8887).
- [15] Bhumika Bhatt, Premal J Patel, Hetal Gaudani, A Review Paper on Machine Learning Based Recommendation System.
- [16] Purnima Bholowalia, Arvind Kumar, Review on determining number of Cluster in K-Means Clustering, *International Journal of Advance Research in Computer Science and Management Studies*.

¹<https://www.facebook.com/>

²<https://www.imdb.com/>