Context-aware Social Recommendation in Sharing Community

Dong Qin

Supervised by: Dr. Xiangmin Zhou
A/Prof. Xiuzhen Zhang

Abstract. Social recommendation has become a hot research topic because of its important applications in the domains such as entertainment, online news broadcasting etc. Recommendation systems analyze users' interests and deliver the potentially interested information to them. Although a lot of effort has been put on this research, it is still challenging to address the problems of effective and efficient social recommendation over sharing communities due to the huge data volume and extremely complex social data content and contexts. In this proposal, we aim to address three recommendation problems for different application scenarios: context-aware individual recommendation, context-aware group recommendation and context-aware online media data stream recommendation. This proposal focuses on our research background on this topic, problem formulation and possible solutions.

1 Introduction

The creation of online media data sharing communities has resulted in the astonishing increasing of digital videos, and their wide applications in the domains such as entertainment, advertisement, online new broadcasting etc. A sharing community is a media sharing service, like Youtube, Flickr and Instagram etc, in which social users can upload, view, comment, rate and download the media data. Users access to these communities, operate and comment on videos for various purposes such as entertainment, online news broadcasting and advertisement. In sharing communities, the behaviour of people is highly affected by the recommendations from the system. According to a white paper released in February 2012 by Unruly Media [1], viewers enjoy online videos they discover from a recommendation more than ones they discover through browsing, which reports 65% of viewers who watched a recommended video enjoyed it, while only 57% of viewers enjoyed a video found through browsing. recommendation system also brings unavoidable commercial profits. 30% profits of Amazon is from recommendation [15]. Providing high quality recommendation service is beneficial to both users and commercial activities.

Recommendation systems have been developed for applications in different domains, such as digital libraries, online marketing, and online video sharing etc. The proposed techniques include content-based filtering [11, 17, 25], collaborative filtering [5, 8, 33] and hybrid filtering. These work mainly exploit the content information and user activity patterns. In addition, these recommendation systems only work well with small datasets. These approaches can not be effectively applied to social media data recommendation.

Social media data have rich contexts, such as time, location, user connections etc, which can be useful for enhancing the effectiveness of recommendation. For example, people from different areas may have different preferences to media data or advertised products. Peoples interest may also change over time, and be affected by their viewing history. The popularity of a social media data may be different in different time periods, among users in different location areas, and in different social user communities. It is important to conduct context aware recommendation. However, context aware recommendation in sharing communities is very challenging because of the high complexity of media content and contexts, ambiguity in media content, complex interaction between content and contexts, and huge volume size of media data. To the best of our knowledge, there is no prior work that fully addresses context aware social recommendation problem.

The reset of the proposal is organised as follows: Section 2 gives the definitions on three research problems and their challenges. Section 3 presents the background and related work. In Section 4,our methodology for three research questions and Section 5 gives the overall Gantt chart of the proposal.

2 Research Questions

In this proposal, we will address three research questions: individual recommendation, group recommendation and streaming media data recommendation from effectiveness and efficiency aspects. This section will present our research motivation, challenges and definition of each problem, followed by the sub-problems to be solved.

1. **How to recommend media data to individual users based on content and context features?**
   Existing media data recommendation approaches use only content features or simply take certain contexts as additional information for better recommendation quality. They do not systematically investigate what features should be selected from a large number of social media data contexts. In addition, the interactions between different features are not kept in relevance identification. So in our first research question, we investigate the problem of individual personalized recommendation by fully exploiting the content and context information of media data. We formulate individual recommendation as:

   Definition 1. *Given a social user u, a social item relevance function f_I, our context-aware media data recommendation algorithm automatically constructs a user profile $p(u)$, and detects a list of most relevant data, S_v, such that for any media data $v_i \in S_v$ and $v_j \notin S_v$, the following condition holds:*

   $$f_I(p(u), v_i) \geq f_I(p(u), v_j). \tag{1}$$

   The input of our context-aware recommendation algorithm is a social user described by a user profile, and its output is a list of media data with highest relevance scores to this profile. Social media data recommendation based on user browsing history faces three challenges: (1) It is hard to decide what contexts should be used in recommendation; (2) It is difficult to capture the interaction between these contexts for effective recommendation; (3) Big social media dataset size. To solve these identified challenges, we divide our first research question into three sub-problems:

(a) Select effective ones among many context and content features by designing novel feature selection algorithms.
(b) Propose a graph-based social media model to capture the similarity between media data and user profiles.
(c) Design a multi-level index structure to enhance recommendation efficiency.

2. **How to recommend media data to a social group users?**
In social communities, users tend to join a group to share, study, and organise events. Recommending social items to a group helps alleviate its user's effort of identifying interested information and enhance data click rate. Thus, our second research question aims to recommend a list of media data to a user group in sharing community. We formulate our group social recommendation problem as follows.

Definition 2. *Given a user group U, a relevance function f_G, our group recommendation automatically constructs a number of user profiles $\{p_i\} \in P$ for all $\{u_i\} \in U$, and detects a list of most relevant media data S_v such that for any item $v_i \in S_v$ and $v_j \notin S_v$ the following condition holds:*

$$f_G(P, v_i) \geq f_G(P, v_j). \tag{2}$$

To solve this research problem, we need to address two challenges. For one thing, users in a group may have different interest points. We need to balance the interest points to maximize the number of interested users in recommendation. For another, compared with individual recommendation, multi users in a group should be taken into account, which suffers from high time cost. It is vital to improve the efficiency of group recommendation. We will address these challenges by solving sub-problems:

(a) How to construct group user profiles to reflect different interests and then aggregate all the recommendation lists considering all users' interests and the novelty of returned items?
(b) How to conduct group recommendation efficiently?

3. **How to continuously recommend incoming streaming media data to users?**
In sharing communities, huge media data are uploaded per minute. These media data can be digitial commercials, news etc. It is important to deliver them to interested users continuously to improve user experience, enhance the effect of media data in real applications. We will address our third research question: How to continuously recommend incoming streaming media data to users in sharing communities? This research problem is formulated as follows.

Definition 3. *Given a stream consisting of incoming media data $S = \{v_1, v_2, ...\}$, a relevance function f_S, and a threshold τ, our recommendation system continuously detects a list of suitable users $U = \{u_1, u_2, ..., u_k\}$ for each incoming media items such that if v_i is recommended to user $u_i \in U$, the following condition holds:*

$$f_S(u_i, v_i) \geq \tau \tag{3}$$

Since incoming media data are unknown to users in sharing communities, no social information attached to them can be used directly for effective recommendation. Only content and limited contexts about them are available to us. In addition, this research problem is typically time-critical, which demands the development of efficient online recommendation approaches. To address these challenges, we will answer our third research question by solving two sub-problems.

(a) How to construct a profile by exploiting its content, contexts and uploader's information and then match media data profile to user profile effectively.
(b) How to efficiently conduct online streaming media recommendation continuously?

3 Background and Related Work

Effective and efficient recommendation in sharing community plays a key role in information delivering, commercial products promotion, sharing knowledge and so on. In this section, we will review the recent related work about recommendation in sharing communities, followed by the evaluation metrics on recommendation.

3.1 Individual Recommendation

Individual recommendation aims to recommend items to one single user in sharing community. We illustrate the current research from technique review, feature illustration, and profile construction respectively.

Technique review Research on conventional recommendation has proceeded along three major dimensions, that is, content-based, collaborative, and hybrid recommendation. They all based on filtering technique which removes unwanted data from large volume of data using automated methods prior to recommend the data to the user.

– Content-based Filtering. Solely based on individual users' history preference. This method exacts the similar items to user's behaviour history based on content information.
– Collaborative Filtering. This technique predicts the opinion of the user and recommends items based on the users opinions and the opinions of the other like minded users. Collaborative filtering can be divided into two categories: memory-based and model-based. The most analyzed exmaples of memory-based collaborative filtering include user-based approaches and item-based approaches. User-based approaches predict the ratings of active users based on the rating of similar users found, and

item-based approaches predict the ratings of active users based on the computed information of items similar to those chosen by the active user. In the model-based approaches, training datasets are used to train a predefined model. including the clustering model, aspect models and the latent factor model.
– Hybrid Filtering. This recommendation method combines content-based and collaborative filtering together to generate recommendation list for users. In fact, it takes not only content information but also user connection information. It has been shown that always achieve precise recommendation performance.

One important aspect of recommendation is determining how interested a user will be in a certain piece of information. Determining how interesting information is to a user is basically a form of predicting. Recent researchers believe that combining multiple techniques in a more dynamic and intelligent way can provide more accurate and stable predictions. Predicting the interest of a user in information is an important process in personalized information system. Current user, other users, the information of item itself may also contribute the final recommendation outcome. Van Setten et al. [27] calculates how interested a certain user will be in a piece of information, using prediction algorithm such as social filtering, case-based reasoning, information filtering, item-item filtering, and genre Least Mean Square. The resulting prediction is a numerical value representing the amount of expected interest for the user. [12] proposes a hybrid movie recommender system based on neural networks. It combines the result of content-based and Collaborative Filtering. The MovieLens data set was used to test the proposed hybrid system. The recommendation process is separated in two stage: the collection of items in order to create a data base, and then the selection of the suitable element from the base for the individuals. The evaluation it used is Precision and Recall. Luo et al. [13] utilize multi-modal information sources(audio, video and closed captions) for news videos to achieve more reliable news topic detection. A personalized news recommendation is proposed in [35] to explore the relations between newly-published news articles and the user's profile. Two-level recommendation hierarchy, where the first level shows a brief summary for each topic category the user might prefer, and the second level gives a specific list of news articles similar to the user's reading interest. In [7] the system recommends personalized sets of videos to users based on their activity on the site. In order to keep users entertained and engaged, it is imperative that these recommendations are updated regularly and reflect a user's recent activity on the site. Zhu et al. [37] propose VideoTopic to decompose the recommendation process into video representation and recommendation generation. First, both visual and textual features of videos are extracted. Using a topic model, each video is then represented as a mixture of set of topics, and each topic is a mixture distribution of textural and visual words/content extracted from a video collection. User interests are then estimated based on users' previously watched videos, and can also be represented as a distribution over topics. A hybrid recommendation approach has been proposed in [34] for video recommendation over social networks by considering the user relationship strength and the interest degree of video. For a given user, the recommendation score of a video candidate is decided by the interest degree of the video by the user's friends, and relationship strengths between the user and his friends. [30] presents a novel online video recommendation system based on multimodal fusion and relevance feedback. It shows the feedback can adjust the recommendation results dynamically and generate satisfying recommendation.

Feature illustration The features used to describe social data include content information such as aural text and social context information. Text related to a social document can be grouped into two categories: (1) explicit text, referring to the surrounding text provided by publisher(2) implicit text, referring to the (hidden) categories and their probabilities obtained by automatic text categorization based on a set of predefined taxonomies (i.e., category hierarchy). Turnbull et al. [26] utilize the audio content (related to timbre and harmony) and two social sources (social tags and web documents) information. It analyse the audio signal by using two acoustic feature representations, one related to timbre and onerelated to harmony. The two features are represented as bags of feature vectors for each song. It also uses two more socially-situated sources, one based on social tags and one based on web documents. For each of these four representations, we describe algorithms that evaluate the relevance of a song to all tags from a given vocabulary. Using such algorithms, we can retrieve songs from a test corpus, ordered by their relevance to a given text query. In [16], the document of each individual modality is represented by a set of features, feature vector. Furthermore, as different parts of video may have different degree of interest to a suer. Assigned with a feature weight, reflecting the degree of interest of this user to this segment. Adjust it by feedback. We use the Vector Space Model (VSM) and probabilistic model to describe the explicit and implicit texts, respectively. The visual relevance is measured by color, motion, and shot tempo(the average number of shots per second).

Another issue is about how to decide the different contribution of different features? It has two clues: intra-weights and inter-weights. Intra-weights mainly focus on the relationship among same features but inter-weights considers the relevance between isomerous features. [18] builds the photos in sharing community as a graph, nodes representing same feature and edge representing the relationship among same and isomerous features.

Profile construction The goal of the recommendation system is to provide personalized recommendations that help users find high quality relevant to their interest. So tt is necessary to construct user profile to reveal user's interest and behaviour in recommendation system. In general, user profiles mainly come from two types of sources: (1) direct profile, that is, users selection of a list of predefined interests such as the user profile in demographic-based recommendation which is created with the classification of users into stereotypes; (2) indirect profile, that is, users rating of a number of items. Regardless of what kinds of items are recommended by these systems, the objective is to collaboratively recommend the items matching to the profile or interest. However, content-based recommendation dedicated to video has not yet been deeply studied.

3.2 Group Recommendation

Users in sharing community tend to join a group in social network to share, study, and organise events such as music selection in public places, tourist attractions, holiday destinations, movies, and TV programs. Recommending social items to a group helps alleviate users effort of information retrieval and enhance data click rate. Group recommendation approaches are typically based on generating an aggregated preference by using the users individual preferences. The main approaches to recommend information to group can be grouped into two parts as follows.

- Merging the recommendation made for individuals. Group recommendation is usually based on aggregating individual users' preferred items into a single list of recommendations to a group using consensus function. When doing group recommendation, there are two consensus function: aggregated voting and least misery.
- Constructing a group preference model. Most of the previous works in group recommendation consider the preferences of every member of the group with the same degree of importance and try to satisfy the preferences of every individual. However, groups of people can have very different characteristics, like size, the relationships among their members or the distribution of people with similar or antagonistic personal preferences. So mining theses factors helps to construct more precise preference model.

Next, we will review the group recommendation methods has been done from the above two categorises. Seko et al. [24] calculates recommendation scores using a feature space that consists of the behavioral tendency of a group and the power balance among group members based on individual preference and the behavioral history of group. At first, the system determines the rating of each watched content using the content genre rating list of each member. Assuming that n is the number of genres that cover content c and $r_{r,g}$ is the rating score ascribed to member i for genre g, the rating of member i for content $c(u_{i,c})$ is defined by Equation:

$$u_{i,c} = \frac{1}{n}\sum_{k=1}^{n} r_{i,k} \tag{4}$$

Moreover, the system creates the rating vector to plot the watched content on the space which has each member's rating dimension. The rating vector is a multidimensional vector composed of the rating of the content by each member. Assuming that m is the number of group members, the rating vector of content c is given by:

$$V_c = \{u_{1,c}, u_{2,c}, ..., u_{m,c}\} \tag{5}$$

According to predicted user rating, the group recommendation algorithm in Wang et al. [28] calculates predicted group rating, which is usually achieved by a certain aggregate function. Its aggregate function consider the relationship between various users in the group, and optimize the aggregate function according to users different influence on the group, which can better reflect the social characteristics of group. PageRank algorithm is introduced in the recommendation method to calculate the member's importance in the group respectively, and to amend the aggregate function of individual preferences. The aggregate function consider the relationship between various users in the group, and optimize the aggregate function according to users different influence on the group, while can better reflect the social characteristics of group.

Ma et al. [14] takes social and trust relationship into consideration to measure the individual's preference. The matrix decompose algorithm decompose the user's preference matrix and social network matrix to calculate a feature vector of the user and objects, and use of the feature vector point multiplication to measure the user's preference . The aggregate function consider the relationship between various users in the group, and optimize the aggregate function according to users different influence on the group, which can better reflect the social characteristics of group.

$$pred(G,i) = \sum (f_{u,G} * pred(u,i)) \tag{6}$$

where $f_{u,G}$ stands for influence of user u on the group G, and $pred(u,i)$ can get from the recommendation system.

Quijano-Sanchez et al. [20] proposes a new recommendations to groups by using existing techniques of collaborative filtering, while taking into account several social factors that improve the accuracy of the system: the composition of the group personality and the social connections among the individuals. It simulates the argumentation process followed by groups of people when agreeing on a common activity in a more realistic way. Personality and social trust factors used to improve the recommendations.

How to recommend to a group of users who may or may not share similar tastes. Amer-Yahia et al. [2] analyses the desiderata of group recommendation and proposes a formal semantics that accounts for both item relevance to a group and disagreements among group members.

In [22], any group member can provide a vector of preferences. It has been previously shown that user preferences may vary depending on mood, context, and company (i.e., other people in the group). They are interested in a specific scenario where users are provided the flexibility to update their preferences during recommendation time by choosing items they would like or not to see, and the system accounts for those newly provided preferences to compute recommendations to the group. This new feature is useful in a number of practical applications such as travel planning, online games, and book clubs, or strategic voting, where users are likely to be in a different mindset at recommendation time and do not want the system to solely rely on their past preferences. In [23], homophony and social influence suggest that preferences (e.g., over products, services, political parties) are likely to be correlated among people whom directly interact in a social network. They develop a model, preference-oriented social networks that captures correlations of individual preferences, where preferences take the form of rankings over a set of options. It is widely recognized that individuals' behaviours and preferences are correlated with those of their friends or connections.

Berkovsky and Freyne [3] investigates the use of aggregated group data in collaborative filtering recipe recommendations. Recommendations are tailored to the entire group, to ensure maximum satisfaction of each member and the group as a whole. Aggregated group-based data can be achieved by weighting the data of individual users accordingly. Four models are proposed for weighting user data.

1. The uniform model weights users uniformly, i.e., $\omega(u_x, f_a) = 1$
2. The heuristic model is role-based, where a role refers to a user's function within a family: applicant, partner, or child.

3. Two other weighting models are based on the observed user interactions with the content. The weights assigned to usedrs reflect their activity $act(u_x)$, i.e., number of ratings $rat(u_x, item_i)$, as a predictor of their degree of engagement.
   – The role-based model weights users according to the activity $act(u_x)$ of users in the same role across the entire community:

$$\omega(u_x, f_a) = \frac{\sum_{y \in U} act(u_y) | role(U_y) = role(u_x)}{\sum_{y \in U} act(u_y)} \tag{7}$$

   – The family-log model weights users according to their activity in relation to other family members:

$$\omega(u_x, f_a) = \frac{act(u_x)}{\sum_{y \in f_a} act(u_y)} \tag{8}$$

The individual data of group members need to be aggregated in a weighted manner, such that the weights reflect the observed interaction of group members, focusing on interactions observed with as localized as possible boundaries.

Oconnor et al. [18] presents a new collaborative filtering recommender system designed to recommend items for groups of users, rather than for individuals. There are two issues to address when forming recommendations for groups. First, they define a social value function that describes how the tastes and opinions of individuals affect the groups recommendation. Then they propose algorithmic implementation of that social value function to create an efficient recommendation based on the tastes of many users. They found that users not only valued group recommendations, but were willing to yield some privacy to get the benefits of group recommendations.

Recio-Garcia et al. [21] presents a method for recommendation to groups that distinguishes among the different types of individuals according to their personality.They study how the group personality composition influences the recommendation accuracy for the group. The novelty of our approach lies in the use of the member personalities to choose the most interesting movie that would better satisfy the whole group. In conflict situations, they describe the behaviour of an individual along two basic dimensions: assertiveness and cooperativeness.

According to predicted user rating, the group recommendation algorithm in [28] calculates predicted group rating, which is usually achieved by a certain aggregate function. The innovation of this paper is that our aggregate function consider the relationship between various users in the group, and optimize the aggregate function according to users different influence on the group, which can better reflect the social characteristics of group.

$$pred(G, i) = \sum (f_{u,G} * pred(u, i)) \tag{9}$$

where $f_{u,G}$ stands for influence of user u on the group G, and $pred(u, i)$ can get from the recommendation system.

3.3 Evaluation Metrics

Recommendation system uses two main methods to evaluate the recommendation performance, one based on user study and the other based on history data.

Based on user study To conduct a user study of a recommendation system, the researches invite multiple subjects to use the recommender system and evaluate its performance. For each recommendation task, the subjects need to evaluate the top-k recommendations suggested by the recommendation system. There are six main metrics are usually used as follows.

- Average accuracy (AC). AC is defined as the proportions of videos with the rating bigger than 4 to all recommended videos. It indicates the proportion of correct recommendations.
- Average cumulative gain (AGC). AGC indicates the average rating of all recommendations. It is the mean of non-interpolated average precisions(AP). The videos with scores no less than 4 are defined as relevant documents when computing AC. AGC indicates the ranking order of correct recommendations in the list.
- Normalized discounted cumulative gain (nDCG). Researchers aggregate all the feedback provided by the subjects to create an ideal ranking list. As recommendations are based on result rankings, the normalized discounted cumulative gain(nDCG) is used to measure the effectiveness of the recommendation list. A higher nDCG value means that more relevance items appear first in the results list.
- mean average precision (MAP)
- Precision@k. The percentage of top-k answers retrieved that are correct. Generally, the correct recommendations are regarded as the ones which judge values from subjects are above a predefined threshold.
- Recommendation Satisfaction Index(RSI). It is used to measure user's evaluation for recommendation. It is described the ration of satisfactory recommendation(SP) to entire recommendation(ST)

$$RSI = \frac{\sum SP}{\sum ST}. \tag{10}$$

Based on history data In this evaluation method, a user's view history is divided into training set and test set. Training set is used to build the recommendation items based on proposed model and user's preferences. Then the system is evaluated by whether it can suggest items that the user has actually viewed in testing set. The main metrics used in this effectiveness evaluation are as follows:

- Hit Number #hit@k. A successful recommendation is called a hit. For a user u, the hit number of first k recommendations is the number of videos in the intersection of these k recommendations and the clicked items of user u during testing period. Define #hit@k for a single test case as either the value 1, if the clicked items appears in the top-k recommendations, or the value 0, if otherwise.
- Precision@k. Precision is the ratio of the number of relevant items (hit number) to the total number of retrieved items (all the items in recommended list).

$$Precison@k = \frac{\#hit@k}{k}. \tag{11}$$

- Recall. Recall is the ratio of the number of relevant items (hit number) which are retrieved to the total number of relevant items (usually all the items in testing set).

– F-measure. F-measure and F1 are proposed by combining Precision and Recall as a synthetic metric. F-measure is the harmonic mean of precision and recall.

$$F1 = \frac{2 * precision * recall}{precison + recall}. \tag{12}$$

– MAE (Mean Absolute Error). MAE calculates the error between the predicted rating and actual rating of the users, to measure the accuracy of the forecasts.

$$MAE = \frac{\sum_{i}^{|T|} |r_i - r'_i|}{|T|}. \tag{13}$$

where r is the real score, r' is the score calculated by recommendation system, T is the testing set.

– RMSE (Root Mean Squared Error). RMSE is similar with MAE. It is defined as:

$$RMSE = \sqrt{\frac{\sum_{i}^{|T|} (r_i - r'_i)^2}{|T|}}. \tag{14}$$

The smaller the MAE or RMSE is, the better the recommended system is.

– Appropriate Precision

– Novelty Precision. Herlocker suggests that recommendation systems should achieve not only high accuracy, but also usefulness. As usefulness factors, he introduced the terms Novelty and Serendipity. Novelty or Serendipity means interesting but unknown to the user. In this paper, we define “unknown” as genre combinations that have yet to be watched. Both Appropriate Precision and Novelty Precision are extremely important since known content cannot increase user satisfaction even if the content is appropriate for the user. In this paper, we evaluated the proposal’s ability to find novel content by calculating “Novelty Precision”.

Table 1 concludes the evaluation methods in recent literatures.

Table 1: Literature Review

papers	methods	metrics
Kuo et al. [10]	Based on History Data	AR
Chen et al. [4]	Based on History Data	#hit@k, Accuracy@k
Yang et al. [30]	Based on user study	AR, AC, MAP
Yin et al. [31]	Based on History Data	#hit@k, Accuracy@k
Guan et al. [9]	Based on History Data	Precision, MAP, nDCG
Öztürk and Cicekli [19]	Based on History Data	Recall, hit@N
Wong et al. [29]	Based on History Data	RSI
Wang et al. [28]	Based on History Data	AR
Yin et al. [32]	Based on History Data	Recall, hit@N
Zhou et al. [36]	Based on user study	AR, AC, MAP
Cui et al. [6]	Based on user study	Precision

4 Methodology

In this section, we present a brief outline of the methodology we plan to use towards addressing the research questions of Section 2. Our research questions focus on efficient and effective recommendation to different subjects, i.e., item recommendation to an individual user, item recommendation to a group users, and group recommendation to an individual user. So we will illustrate our proposed solution from three aspects.

4.1 Individual Recommendation

In this section, we describe our proposed approach for the sub-problems in Section 2.1. We will propose correlation-based feature selection, graph-based social model, and multi-level index structure for addressing the challenges in Section 2.1 (a)-(c) respectively.

Correlation-based feature selection We will propose two alternative feature selection algorithms, global-based feature selection and group-based feature selection, to balance the feature selection quality and processing time. We plan to exploit the information theory based measure for capturing non-linear correlations what exist in our application.

We select features based on a training dataset, which is manually labeled by human relevance judgment. A set of 10 classes $C = \{C_i\}$ are labeled over the training dataset. We will use symmetrical uncertainty(SU) to measure the correlation between a feature and a class to judge whether a feature is relevant. A feature whose SU value equals to 0 is irrelevant and can be filtered. Then we propose a new concept feature contribution based on joint entropy, denoted as FC, to reveal the real information contribution of a feature when facing a group of features. The feature selection process includes two parts:

- **Relevance checking**: For a data set D with m features $F_o = \{f_1, f_2, , f_m\}$ and a class set C, we calculate the $SU(f_i, C)$ over each feature f_i and delete the features whose SU are equal to 0. All the remaining features comprise the class-relevant feature subset $F_r = \{f_1, f_2, , f_n\}$.
- **Redundancy removal**: Our global-based algorithm adopts sequential backward selection called Feature Contribution based Redundancy Removal (FCRR), and continuously processes every feature until all are finished. In every loop, for any $f_i \in F_r$, we compute its feature contribution regarding with remaining feature set $S = \{F_r - f_i\}$. Then the feature with smallest feature contribution is selected and compared with a predefined threshold. If this feature is smaller than the threshold, it will be removed from the feature set.
  To improve the efficiency of feature selection, we plan to propose a group-based algorithm, which will first find a number of groups, and then apply FCRR algorithm to each group to remove the redundant features. To identify the groups, we first construct a weighted complete graph that takes every feature as a node, and the edge weight of each pair of nodes is the SU value between them. Then, we build a minimum-spanning tree (MST) based on this weighted graph, and cut the MST into a forest. Each subtree of the forest is a group in which vertices are co-related and potentially redundant features exist. We finally perform FCRR over each group

Graph-based social data model We will construct a graph model over media data and user profile. Each media data is represented as a Feature Interaction Graph (FIG), and each user profile is described as a FIG set. We plan to design a graph similarity measure for matching media data (FIG) to user profiles (FIG sets). If a media data matches any one in a user profile, this media data is relevant to this user. We will employ a pair-nodes clique model over FIG to calculate the similarity between two media data. In FIG graph, a pair-nodes clique is the clique comprising with two nodes and the edge connecting them. Suppose we have two media data graphs G and G', the clique similarity between two pair-nodes cliques $\phi(n_i, n_j, e_{ij})$ and $\phi(n'_i, n'_j, e'_{ij})$ from G and G is defined as

$$cs(\phi, \phi') = \gamma(d_i + d_j) + (1 - \gamma)\rho(e_{ij}, e'_{ij}), \tag{15}$$

where d is the distance between two nodes, ρ is the distance between e_{ij} and e'_{ij}, and γ denotes the weights of nodes in clique similarity. We treat FIG graph as a set of pair-nodes cliques ϕ, and the similarity between G and G' can be measured by:

$$S(G, G') = \sum_{i=1}^{|\phi|} \lambda_i cs_i(\phi_i, \phi'_i) \tag{16}$$

where $\lambda_i \in \Lambda$ is the weight of this pair-node clique.

Index structure We design a multi-level index structure that consists of an R-tree structure, an array and a number of multiple level inverted files. FIG graphs are stored in the inverted files, and the R-tree is to find the content index keys of a given query. The array is used to store multiple index keys to different attributes, and a number of media data ids that can be used to identify potential relevant media data graph information in inverted files

4.2 Group Recommendation

In this section, we describe research methodology for the sub-problems in Section 2.2. We will propose group multi-profile construction, recommendation aggregation, and profile summarization for addressing the challenges in Section 2.2 (a)-(b) respectively.

Group multi-profiles construction Existing work for group profile construction mainly focus on establishing a new virtual profile based on each member in group. However, individuals in a group may have many different interest points. The virtual profile in existing work in fact is generated using an average method, which assumes a group only has a single interest, which makes the profile only represents a small part of interest. In our proposal, we fully capture the interest differences among group members, and the importance of different interest point construct a profile for each member based on his/her browsing history, uploaded commented items. The constructed multi-profile of G are represented as a set $P = \{pi\}$.

Recommendation aggregation Based on multi-profiles constructed, our system extracts k recommendation lists based on each profile. The challenge is how to aggregate the different recommendation lists. The main solution in existing work has two clues: average aggregation and least misery. However, different members has different significance to a group. For example when a group of people have dinner, the guests should be given high priority, or for a group activity, the most active members should also be given more consideration. In order to get more accuracy recommendation, we introduce *viscosity* of a profile to measure the significance of a profile. *Viscosity* of a profile is the sum of profile weight w and its contribution c to the group, such as the frequency of attending group activity, the number of followers and so on. Based on the relevance function $r_I(p_i, v)$ in individual recommendation, the group relevance function is defined

$$r_G(G, i) = \sum v_i \cdot r_I(p_i, v) \tag{17}$$

Profile summarization For very large groups, it is slow to process each profile of a group one by one. We plan to conduct hierarchical summarization on multi-profiles offline. The efficiency of recommendation is affected by hierarchical grain.

Evaluation metrics Recommendation systems should achieve not only high accuracy, but also usefulness. We use Appropriate Precision to measure accuracy. As usefulness factors, we introduce a metric called *Novelty* which means interesting but unknown to the user to evaluate.

4.3 Continuous Streaming Media Data Recommendation

In this section, we present our proposed approach for the sub-problems in Section 2.3. We will propose context information supplement, media data and user profile matching, and online hashing and incremental computing for addressing the challenges in Section 2.3 (a)-(b) respectively.

Context information supplement For a incoming streaming media data to sharing community, v, only its uploader's information can be obtained because it has little change to communicate with other users, which makes the available information very limited. Fortunately, a user usually has specific interests. We can find similar media data in his uploading history by media data based on content. The context information c_u of these media data can be treated as the supplement contexts of the incoming social media data. Besides, the closest friends of a user usually share similar interests, which provides another way to extract context information from the friends of uploaders.

Media Data and user profile matching After supplementing the context information, we use FIG model in research question 1 to construct media data and user profile and then conduct match.

Online hashing and incremental computing In order to recommend the online stream in a time window $[t, t+T)$ instantly, we plan to map these incoming media data to a space by transformation. For the next time window $[t+T, t+2T]$, we also execute online hashing transformation. If there is hash collision between a media data in the second window and another in the previous window, we compute the similarity between these two media data. We recommend the second media to the same group of users in case that these two data are similar. By incremental computing, we take advantage of previous time window result to accelerate stream processing.

Evaluation metrics

- Accept rate: we recommend media data stream to different users in sharing community and adopt the ration of accepted ones to whole number of recommended ones as our effectiveness evaluation.
- Efficiency: we compare the time cost with incremental computing and without increment computing to show the efficiency.

5 Time Schedule

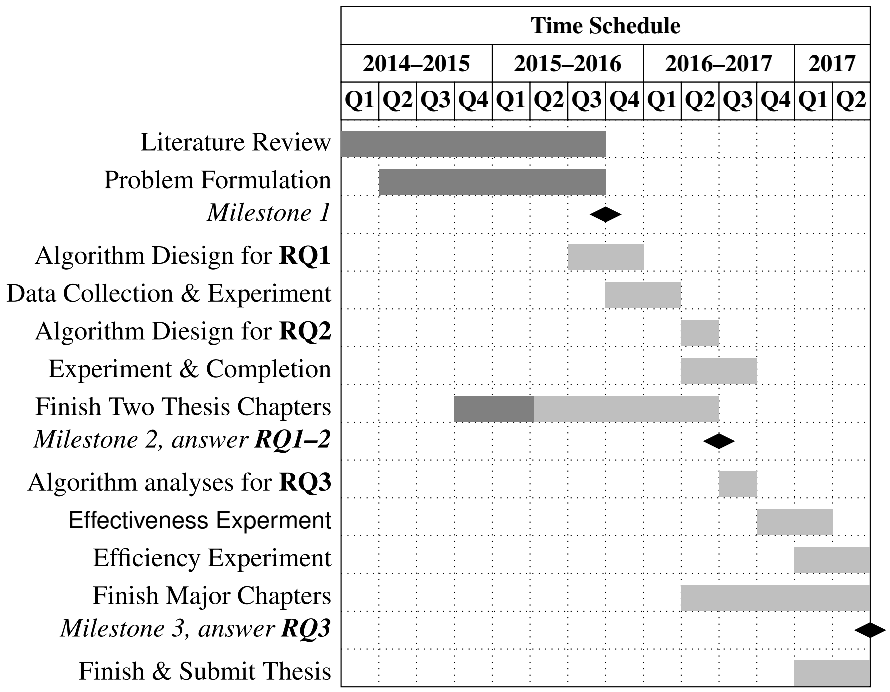

Fig. 1: Gantt Chart for a three year PhD project. The dark grey part indicates the percentage of tasks finished. Each grid cell represents a time span of three months. The starting date was 21st July, 2014.

References

[1] http://www.marketingcharts.com/online/online-viewers-prefer-socially-recommended-videos-21011/.

[2] Sihem Amer-Yahia, Senjuti Basu Roy, Ashish Chawlat, Gautam Das, and Cong Yu. Group recommendation: Semantics and efficiency. *Proceedings of the VLDB Endowment*, 2(1):754–765, 2009.

[3] Shlomo Berkovsky and Jill Freyne. Group-based recipe recommendations: analysis of data aggregation strategies. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 111–118. ACM, 2010.

[4] Bisheng Chen, Jingdong Wang, Qinghua Huang, and Tao Mei. Personalized video recommendation through tripartite graph propagation. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 1133–1136. ACM, 2012.

[5] Chi-Yin Chow, Jie Bao, and Mohamed F Mokbel. Towards location-based social networking services. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, pages 31–38. ACM, 2010.

[6] Bin Cui, Anthony KH Tung, Ce Zhang, and Zhe Zhao. Multiple feature fusion for social media applications. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pages 435–446. ACM, 2010.

[7] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296. ACM, 2010.

[8] Lucia Del Prete and Licia Capra. differs: A mobile recommender service. In *Mobile Data Management (MDM), 2010 Eleventh International Conference on*, pages 21–26. IEEE, 2010.

[9] Ziyu Guan, Can Wang, Jiajun Bu, Chun Chen, Kun Yang, Deng Cai, and Xiaofei He. Document recommendation in social tagging services. In *Proceedings of the 19th international conference on World wide web*, pages 391–400. ACM, 2010.

[10] Fang-Fei Kuo, Meng-Fen Chiang, Man-Kwan Shan, and Suh-Yin Lee. Emotion-based music recommendation by association discovery from film music. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 507–510. ACM, 2005.

[11] Ken Lang. Newsweeder: Learning to filter netnews. In *Proceedings of the 12th international conference on machine learning*, pages 331–339, 1995.

[12] Qing Li and Byeong Man Kim. An approach for combining content-based and collaborative filters. In *Proceedings of the sixth international workshop on Information retrieval with Asian languages-Volume 11*, pages 17–24. Association for Computational Linguistics, 2003.

[13] Hangzai Luo, Jianping Fan, Daniel A Keim, and Shinichi Satoh. Personalized news video recommendation. In *Advances in Multimedia Modeling*, pages 459–471. Springer, 2009.

[14] Hao Ma, Haixuan Yang, Michael R Lyu, and Irwin King. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM*

conference on Information and knowledge management, pages 931–940. ACM, 2008.

[15] J Mangalindan. Amazons recommendation secret. *CNN Money http://tech. fortune. cnn. com/2012/07/30/amazon-5*, 2012.

[16] Tao Mei, Bo Yang, Xian-Sheng Hua, and Shipeng Li. Contextual video recommendation by multimodal relevance and user feedback. *ACM Transactions on Information Systems (TOIS)*, 29(2):10, 2011.

[17] Bradley N Miller, Istvan Albert, Shyong K Lam, Joseph A Konstan, and John Riedl. Movielens unplugged: experiences with an occasionally connected recommender system. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 263–266. ACM, 2003.

[18] Mark Oconnor, Dan Cosley, Joseph A Konstan, and John Riedl. Polylens: a recommender system for groups of users. In *ECSCW 2001*, pages 199–218. Springer, 2001.

[19] Gizem Öztürk and Nihan Kesim Cicekli. A hybrid video recommendation system using a graph-based algorithm. In *Modern Approaches in Applied Intelligence*, pages 406–415. Springer, 2011.

[20] Lara Quijano-Sanchez, Juan A Recio-Garcia, Belen Diaz-Agudo, and Guillermo Jimenez-Diaz. Social factors in group recommender systems. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(1):8, 2013.

[21] Juan A Recio-Garcia, Guillermo Jimenez-Diaz, Antonio A Sanchez-Ruiz, and Belen Diaz-Agudo. Personality aware recommendations to groups. In *Proceedings of the third ACM conference on Recommender systems*, pages 325–328. ACM, 2009.

[22] Senjuti Basu Roy, Saravanan Thirumuruganathan, Sihem Amer-Yahia, Goutam Das, and Cong Yu. Exploiting group recommendation functions for flexible preferences. In *Data Engineering (ICDE), 2014 IEEE 30th International Conference on*, pages 412–423. IEEE, 2014.

[23] Amirali Salehi-Abari and Craig Boutilier. Preference-oriented social networks: Group recommendation and inference. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 35–42. ACM, 2015.

[24] Shunichi Seko, Takashi Yagi, Manabu Motegi, and Shinyo Muto. Group recommendation using feature space representing behavioral tendency and power balance among members. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 101–108. ACM, 2011.

[25] Márcio Soares and Paula Viana. Tuning metadata for better movie content-based recommendation systems. *Multimedia Tools and Applications*, 74(17):7015–7036, 2015.

[26] Douglas R Turnbull, Luke Barrington, Gert Lanckriet, and Mehrdad Yazdani. Combining audio content and social context for semantic music discovery. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 387–394. ACM, 2009.

[27] Mark Van Setten, Mettina Veenstra, Anton Nijholt, and Betsy van Dijk. Prediction strategies in a tv recommender system-method and experiments. In *ICWI*, pages 203–210, 2003.

[28] Jing Wang, Zhijing Liu, and Hui Zhao. Group recommendation based on the pagerank. *Journal of Networks*, 7(12):2019–2024, 2012.

[29] Fung Wong, Shijun Lee, and Quanrui Wong. Points of interest recommendation based on context-aware. *International Journal of Hybrid Information Technology*, 8(3), 2015.

[30] Bo Yang, Tao Mei, Xian-Sheng Hua, Linjun Yang, Shi-Qiang Yang, and Mingjing Li. Online video recommendation based on multimodal fusion and relevance feedback. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 73–80. ACM, 2007.

[31] H. Yin, X. Zhou, Y. Shao, H. Wang, and S. Sadiq. Joint modeling of user check-in behaviors for point-of-interest recommendation. In *Proc. Conf. Information and Knowledge Management*, pages 1631–1640. ACM, 2015.

[32] Hongzhi Yin, Bin Cui, Jing Li, Junjie Yao, and Chen Chen. Challenging the long tail recommendation. *Proceedings of the VLDB Endowment*, 5(9):896–907, 2012.

[33] Zhijun Yin, Liangliang Cao, Jiawei Han, Jiebo Luo, and Thomas S Huang. Diversified trajectory pattern ranking in geo-tagged social media. In *SDM*, pages 980–991. SIAM, 2011.

[34] Xiaojian Zhao, Jin Yuan, Richang Hong, Meng Wang, Zhoujun Li, and Tat-Seng Chua. *On video recommendation over social network*. Springer, 2012.

[35] Li Zheng, Lei Li, Wenxing Hong, and Tao Li. Penetrate: Personalized news recommendation using ensemble hierarchical clustering. *Expert Systems with Applications*, 40(6):2127–2136, 2013.

[36] Xiangmin Zhou, Lei Chen, Yanchun Zhang, Longbing Cao, Guangyan Huang, and Chen Wang. Online video recommendation in sharing community. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1645–1656. ACM, 2015.

[37] Qiusha Zhu, Mei-Ling Shyu, and Haohong Wang. Videotopic: Content-based video recommendation using a topic model. In *Multimedia (ISM), 2013 IEEE International Symposium on*, pages 219–222. IEEE, 2013.