JAM: Joint Action Matrix Factorization for Summarizing a Temporal Heterogeneous Social Network

Yu-Ru Lin Hari Sundaram Aisling Kelliher

Arts Media and Engineering Program, Arizona State University, Tempe, AZ 85281
{yu-ru.lin, hari.sundaram, aisling.kelliher}@asu.edu

Abstract

This paper presents *JAM* (Joint Action Matrix Factorization), a novel framework to summarize social activity from rich media social networks. Summarizing social network activities requires an understanding of the relationships among concepts, users, and the context in which the concepts are used. Our work has three contributions: First, we propose a novel summarization method which extracts the co-evolution on multiple facets of social activity – who (users), what (concepts), how (actions) and when (time), and constructs a context rich summary called "activity theme". Second, we provide an efficient algorithm for mining activity themes over time. The algorithm extracts representative elements in each facet based on their co-occurrences with other facets through specific actions. Third, we propose new metrics for evaluating the summarization results based on the temporal and topological relationship among activity themes. Extensive experiments on real-world Flickr datasets demonstrate that our technique significantly outperforms several baseline algorithms. The results explore nontrivial evolution in Flickr photo-sharing communities.

Introduction

Today, we are witnessing the large scale use of rich media social networks such as Flickr, YouTube and MySpace. There, users can upload media, tag them using text keywords, and comment on other user's activities. This paper focuses on summarizing social activities in rich media networks. A summary can prove valuable to the end user, as it allows her to contextualize her activities in terms of the broader community. Summaries are valuable for understanding why certain content has become highly popular. Such knowledge will be of use to both the companies that support the social network (e.g. Flickr) for managing resources as well for companies interested in extracting cultural trends. Frequency based approaches (tag clouds, list of important users, ranked list of popular images etc.) for characterizing content are important, but limited in the insight they can provide. Rich-media social networks are interesting not only due to the diversity of content available, but also because social actions on and around the content, including tagging / uploading of

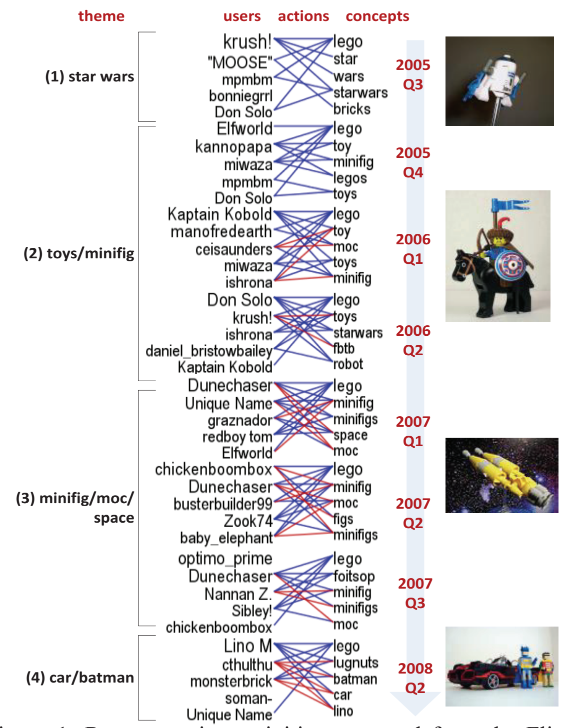

Figure 1: Representative activities extracted from the Flickr group "Lego Freaks" by using our method. An activity theme is represented by a set of users, concept terms with respect to actions (blue edges represent *post* actions and red edges represent *comment* actions).

content, and exchange of comments between users. Hence summarizing social network activities requires an understanding of the relationship between concepts, users, and the context in which the concepts are used.

Related work. Social activity summarization deals with constantly changing mishmash of interrelated users, concepts and media objects. Recent analysis of social groups and their temporal dynamics (e.g. [4]) has focused on dynamic but homogeneous networks, i.e. the edges represent homogeneous actions (e.g. posting). Such temporal analysis cannot be directly applied to analysis of co-evolution on multiple facets in a context rich online social network. Although heterogeneous interrelated entities have attracted considerable interest (e.g. [1,5]), these multi-graph mining algorithms do not consider the temporal evolution of the interrelated entities.

Our approach. We propose *JAM* (Joint Action Matrix Factorization), a novel method for summarizing social activity over time. The key idea is to show community activity as *co-evolution on multiple important facets*: who (users), what (concepts), how (actions) and when (time). An example is given in Figure 1 – an empirical result of our approach, which summarizes the social activity during 2005-2008 in a Flickr photo group called "Lego Freaks" (http://www.flickr.com/groups/51035542119@N01/). In each quarter, an *activity theme* consisting of multiple facets is extracted for summarizing the group's activity. We highlight some observation: (a) The group activity exhibit complex dynamics. From the concept facet we see mini-figures (termed "minifig") are of the group's top interests during 2006-2007. However, the users who are interested in the same concept have changed. (b) Users and concepts can be related through different actions. Those who post photos on a topic might rarely give comments on the same topic. The distribution of different type of actions suggests how an element (user or concept) gains importance in the community. In comparison with single-facet presentation (e.g. tag cloud), our automated activity summary provides a rich context to understand *why* certain people, concept or media objects have enhanced importance.

Our work has three key contributions: (1) a novel summarization framework that extracts co-evolution on multiple facets of social activity to construct a context rich summary called "activity theme", (2) an efficient algorithm for mining activity themes over time by using a non-negative matrix factorization (NMF) method, and (3) new metrics for evaluating the summarization results based on the temporal and topological coverage and coherence among representative elements in activity themes.

We have conduct extensive experiments on real-world Flickr datasets. The results show our technique significantly outperforms several baseline methods such as interestingness measure [2] and HITS [3], with a 27.2% improvement on the average. The experiments suggest that the activity themes extracted by JAM are more coherent and consistent than baseline methods, in terms of the proposed evaluation metrics. We present some interesting observation from the Flickr datasets to illustrate meaningful and insightful summary of community activity.

Problem Definition

Data model. We construct a data model consisting of the interrelated data entities, including media objects, people, comments and tags. This data model (ref. Figure 2 for a summary of notations) is based on Flickr's social groups and can generalize to other social networks. The data includes different object sets – U (users), P (timestamped media objects such as photos), C (timestamped comments on media) and Q (media description such as tags). Tags are assigned by users and have been commonly used to annotate and retrieve the relevant concepts of a photo. Thus we use tags to represent the concepts of a photo. We shall use "concept", "tag" and "term" interchangeably.

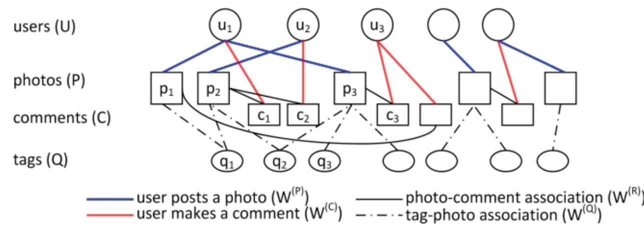

Figure 2: Data model in a photo-sharing space, including four sets of heterogeneous data objects U (users), P (photos), C (comments), Q (tags) and four types of relationships among these objects, which can be represented by matrices: $\mathbf{W}^{(P)}$ – user-photo matrix where each entry $\mathbf{W}_{ij}^{(P)}$ indicates user u_i posts photo p_j, $\mathbf{W}^{(C)}$ – user-comment matrix, $\mathbf{W}^{(Q)}$ – tag-photo matrix and $\mathbf{W}^{(R)}$ – photo-comment matrix.

The heterogeneous relationships among these objects can be represented by corresponding matrices (ref. Figure 2). These matrices are basic relationships in our data model, as other relationships can be derived from a combination of these matrices. We formulate the summarization problem in terms of extracting temporally representative social activities. The key idea is to extract activity as a composite of multiple important facets that provide a rich context to understand the social meaning of the media.

An activity a is defined as a co-presence of multiple important facets: *user*, *action* and *term*. An action (identified by *media object* and *time*) indicates how these facets are associated, e.g. a user *posts* (action) a photo with respect to some tagged concepts, or *comments* (action) on a photo posted by another user. An *activity theme* A_t for a time t is a set of activities i.e. A_t:=$\{a|t\}$. Finally, a temporal social media summary A is a sequence of activity themes, i.e. $A := \{A_1, A_2, \ldots, A_t, \ldots\}$.

Problem. Given interrelated social data entities (U, P, C, Q), extract an activity theme A_t for each time t, to construct the temporal activity summary A.

Activity Theme Extraction

In this section we propose a unified matrix factorization framework for extracting activity themes over time.

Joint Action Matrix Factorization. We solve the problem of extracting temporally representative activity themes in a non-negative matrix factorization (NMF) framework. We formalize theme extraction as a multi-graph clustering problem – each activity theme is a cluster of strongly co-occurring users, actions and terms. Assume there are K activity themes. We first examine two actions among users and terms. The *post* actions comprise user-photo ($\mathbf{W}^{(P)}$) and term-photo ($\mathbf{W}^{(Q)}$) relationships. The *comment* actions comprise user-comment ($\mathbf{W}^{(C)}$) and term-comment relationship. The term-comment relationship, denoted by $\mathbf{W}^{(QR)}$, can be derived by combining term-photo and photo-comment relationships, i.e. $\mathbf{W}^{(QR)} = \mathbf{W}^{(Q)}\mathbf{W}^{(R)}$.

We begin with the $|Q|\times|P|$ term-photo matrix $\mathbf{W}^{(Q)}$. Similar to a term-document matrix, the K-dimensional latent space can be factorized into a $|Q|\times K$ matrix $\mathbf{Y}$ and a $K\times|P|$ matrix $\mathbf{Z}^{(P)}$, where each column of $\mathbf{Y}$ is the axis of each dimension. Photos can be projected to each dimension by the coefficients in the corresponding rows of $\mathbf{Z}^{(P)}$. We put non-negative constraints on $\mathbf{Y}$ and $\mathbf{Z}^{(P)}$ so that each theme is represented by an additive combination of terms. Using $\mathbf{YZ}^{(P)}$ to approximate the matrix $\mathbf{W}^{(Q)}$, we seek to minimize the following objective function:

$$J(\mathbf{Y},\mathbf{Z}^{(P)}) = D(\mathbf{W}^{(Q)} \,\|\, \mathbf{YZ}^{(P)}) \qquad <1>$$
$$s.t\ \mathbf{Y} \in \Re_+^{|Q|\times K}, \mathbf{Z}^{(P)} \in \Re_+^{K\times|P|}, \sum_i \mathbf{Y}_{ij} = 1\ \forall j$$

where $D(\mathbf{A}\|\mathbf{B}) = \sum_{i,j} (\mathbf{A}_{ij} \log \mathbf{A}_{ij}/\mathbf{B}_{ij} - \mathbf{A}_{ij} + \mathbf{B}_{ij})$ is the Kullback-Leibler (KL) divergence between matrices $\mathbf{A}$ and $\mathbf{B}$, which is used to measure how the factorization deviates from the observed data.

The latent space $\mathbf{Y}$ is now solely produced by users' posting actions. We then incorporate different actions – users' commenting actions with respect to the terms. We use the term-comment matrix $\mathbf{W}^{(QR)}$ to relate terms to the latent space through commenting actions. To combine two different actions, we use the same latent space $\mathbf{Y}$ to approximate the matrix $\mathbf{W}^{(QR)}$ as follows:

$$J(\mathbf{Y},\mathbf{Z}^{(C)}) = D(\mathbf{W}^{(QR)} \,\|\, \mathbf{YZ}^{(C)}) \qquad <2>$$
$$s.t\ \mathbf{Y} \in \Re_+^{|Q|\times K}, \mathbf{Z}^{(C)} \in \Re_+^{K\times|P|}, \sum_i \mathbf{Y}_{ij} = 1\ \forall j$$

Next, we relate users to the latent space. Given the user-photo matrix $\mathbf{W}^{(P)}$ and the user-comment matrix $\mathbf{W}^{(C)}$, with the same coefficient matrices $\mathbf{Z}^{(P)}$ and $\mathbf{Z}^{(C)}$ we can find corresponding K-dimensional latent space of which axes are represented by users. Thus, we approximate $\mathbf{W}^{(P)}$ and $\mathbf{W}^{(C)}$ by the following objective functions:

$$J(\mathbf{X},\mathbf{Z}^{(P)}) = D(\mathbf{W}^{(P)} \,\|\, \mathbf{XZ}^{(P)}) \qquad <3>$$
$$s.t\ \mathbf{X} \in \Re_+^{|U|\times K}, \mathbf{Z}^{(P)} \in \Re_+^{K\times|P|}, \sum_i \mathbf{X}_{ij} = 1\ \forall j$$

$$J(\mathbf{X},\mathbf{Z}^{(C)}) = D(\mathbf{W}^{(C)} \,\|\, \mathbf{XZ}^{(C)}) \qquad <4>$$
$$s.t\ \mathbf{X} \in \Re_+^{|U|\times K}, \mathbf{Z}^{(C)} \in \Re_+^{K\times|C|}, \sum_i \mathbf{X}_{ij} = 1\ \forall j$$

We can combine the above objective functions by $J_1 = J(\mathbf{X},\mathbf{Z}^{(P)})+J(\mathbf{X},\mathbf{Z}^{(C)})+J(\mathbf{Y},\mathbf{Z}^{(P)})+J(\mathbf{Y},\mathbf{Z}^{(C)})$. Minimizing the combined objective function gives two K-dimensional latent spaces that correspond to K activity themes. In the space of $\mathbf{X}$, each dimension axis is represented by a column of $\mathbf{X}$, i.e. $\mathbf{X}_j$, and each entry $\mathbf{X}_{ij}$ indicates the strength of user u_i associating with the j-th activity theme. Similarly in $\mathbf{Y}$, $\mathbf{Y}_{ij}$ indicates the strength of term q_j associating with the j-th theme.

To extract themes that have temporal correlation, we introduce time indicator matrices for actions. We segment the data duration into T time slots. For $|P|$ posting actions, we construct a $T\times|P|$ matrix $\mathbf{H}^{(P)}$, with each entry $\mathbf{H}_{ti}^{(P)}=1$ indicating that the photo p_i is posted during time t; 0 otherwise. Similarly for commenting actions, we construct a $T\times|C|$ matrix $\mathbf{H}^{(C)}$. To make the K-dimensional latent spaces to align with these time slots, we let $K=T$. We regularize the objective function by these time indicators. Combining with time regularization and different actions, the goal is to minimize:

$$J(\mathbf{X},\mathbf{Y},\{\mathbf{Z}^{(\kappa)}\}) = \qquad <5>$$
$$\sum_{\kappa=1}^{M} D(\mathbf{W}^{(u,\kappa)} \,\|\, \mathbf{XZ}^{(\kappa)}) + D(\mathbf{W}^{(q,\kappa)} \,\|\, \mathbf{YZ}^{(\kappa)}) + D(\mathbf{H}^{(\kappa)} \,\|\, \mathbf{Z}^{(\kappa)})$$
$$s.t\ \mathbf{X} \in \Re_+^{|U|\times T}, \mathbf{Y} \in \Re_+^{|Q|\times T}, \mathbf{Z}^{(\kappa)} \in \Re_+^{T\times I_\kappa},$$
$$\sum_i \mathbf{X}_{ij} = 1\ \forall j, \sum_i \mathbf{Y}_{ij} = 1\ \forall j$$

where $\mathbf{X}$ and $\mathbf{Y}$ represent the latent spaces of users and terms respectively, $\{\mathbf{Z}^{(\kappa)}\}$ is a set of coefficient matrices for each action type κ. I_κ is the dimensionality (total amount) of type κ actions. $\mathbf{W}^{(u,\kappa)}$ represents the user-action-κ relationships where each entry indicates a user u perform an action of type κ. Similarly $\mathbf{W}^{(q,\kappa)}$ represents the term-action-κ relationships. $\mathbf{H}^{(\kappa)}$ is a time indicator matrix for type κ actions. E.g., if κ represents the posting action, the respective matrices are $\mathbf{W}^{(P)}$, $\mathbf{W}^{(Q)}$ and $\mathbf{H}^{(P)}$. For simplicity we do not consider a weighted combination of individual objective functions, but this extension is straightforward.

Iterative Solution. We provide an iterative algorithm to solve the optimization problem defined by Eq.<5>. Based on the concavity of log function, a local minima solution to Eq.<5> can be found by the following update rules:

$$\mathbf{X}_{it} \leftarrow \sum_\kappa \sum_j \mathbf{W}_{ij}^{(u,\kappa)} \sigma_{ijt}^{(u,\kappa)},\ \ \mathbf{Y}_{kt} \leftarrow \sum_\kappa \sum_j \mathbf{W}_{kj}^{(q,\kappa)} \sigma_{kjt}^{(q,\kappa)}, \qquad <6>$$
$$\mathbf{Z}_{tj}^{(\kappa)} \leftarrow \sum_i \mathbf{W}_{ij}^{(u,\kappa)} \sigma_{ijt}^{(u,\kappa)} + \sum_k \mathbf{W}_{kj}^{(q,\kappa)} \sigma_{kjt}^{(q,\kappa)} + \mathbf{H}_{tj}^{(\kappa)},$$

then normalize such that $\sum_i \mathbf{X}_{ij} = 1\ \forall j, \sum_i \mathbf{Y}_{ij} = 1\ \forall j,$ where

$$\sigma_{ijt}^{(u,\kappa)} = \frac{\mathbf{X}_{it}\mathbf{Z}_{tj}^{(\kappa)}}{(\mathbf{XZ}^{(\kappa)})_{ij}},\ \ \sigma_{kjt}^{(q,\kappa)} = \frac{\mathbf{Y}_{kt}\mathbf{Z}_{tj}^{(\kappa)}}{(\mathbf{YZ}^{(\kappa)})_{kj}}.$$

Due to the sparseness of $\mathbf{W}^{(u,\kappa)}$ and $\mathbf{W}^{(q,\kappa)}$, the total time complexity of each update is $O(mT)$, where m is the number of non-zero entries in the input matrices, and T is the number of time slots. If we consider the number of time slots T and the degree of nodes (users and terms) in the data is bounded by some constant, the complexity is linear in the total number of entities in the networks.

Experiments

In this section, we report experimental studies on a collection of real-world Flickr group datasets.

Flickr Dataset. We have collected the data using Flickr API. We sample 191 groups based on the group size distribution. For each group, we collect photos, comments, tags and relevant users, i.e. users who have either sent a photo in the group or created a comment on a photo in the group. The average number of users per group is 2621. The average time span of the group data is 37 months.

Performance metrics. Evaluating the quality of temporal group activity summarization is another challenge due to the lack of ground truth. Here we propose a quantitative method for evaluating the extracted summaries. We define three metrics based on the topological relationship between the extracted activities and the total activities. Let $A_o = (U_o, E_o, Q_o)$ be the set of all observable activities (users, actions and terms) for a given time t, and let $A_s = (U_s, E_s, Q_s)$ be the set of activities extracted from A_o for representing an activity theme at t. Let G_o and G_s be the corresponding bipartite graphs for A_o and A_s respectively. We define three metrics to evaluate different aspects of an activity theme:

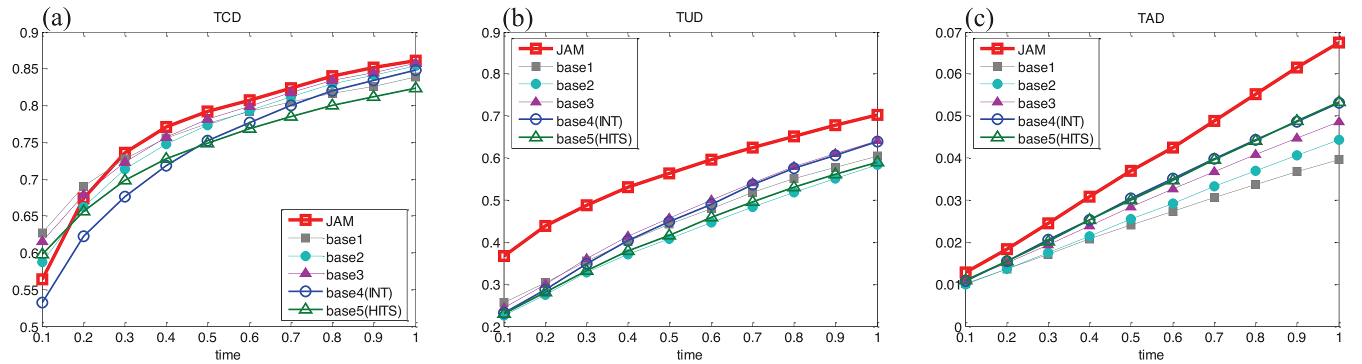

Figure 3: Quantitative evaluation for temporal group activity summary. We compare our method JAM with five baseline methods based on three metrics: (a) TCD – temporal concept degree coverage, (b) TUD – temporal user degree coverage (c) TAD – temporal action density. Our method outperforms baseline methods for all three metrics.

User degree coverage (UD) measures the amount of concepts covered by the extracted user U_s, relative to the amount of total concepts. It is defined as follows:

$$UD(A_s)=\frac{|\Gamma_{G_o}(U_s)|}{|\Gamma_{G_o}(U)|},\quad TUD(A,t)=\frac{\left|\bigcup_{i=1}^{t}\Gamma_{G_o}(U_i)\right|}{|\Gamma_{G_o}(U)|}, \qquad \langle 7\rangle$$

where $\Gamma_G(S)$ denotes the set of adjacent nodes $v\in V\backslash S$ in the graph $G=(V, E)$. The temporal user degree coverage (TUD) is a cumulative temporal measure, which prefers temporally diverse information to redundant information.

Concept degree coverage (CD / Temporal CD) measures the amount of users covered by the extracted concept, and is defined similarly, as:

$$CD(A_s)=\frac{|\Gamma_{G_o}(Q_s)|}{|\Gamma_{G_o}(Q)|},\quad TCD(A,t)=\frac{\left|\bigcup_{i=1}^{t}\Gamma_{G_o}(Q_i)\right|}{|\Gamma_{G_o}(Q)|}, \qquad \langle 8\rangle$$

where $\Gamma_G(S)$ is defined as above.

Action density (AD / Temporal AD) measures the amount of existing connections among extracted set of users and concepts, normalized by the total connections, and is defined as:

$$AD(A_s)=\frac{|E_s|}{|E_o|},\quad TAD(A,t)=\frac{\left|\bigcup_{i=1}^{t}E_i\right|}{|E_o|}, \qquad \langle 9\rangle$$

Results and Discussion. We compare our method, "JAM", with the following baseline methods: *posting frequency* (base1), *commenting frequency* (base2), *total frequency* (base3), *interestingness* [2] (or "INT" for short, base4), and *HITS* [3] (base5).

We partition the time by quarter, and run our algorithm with 5 baseline methods for the 191 groups. We extract the top 5 users and terms based on the ranking given by different methods for each time slot t and extract the observable actions among the extracted users and terms to generate activity themes $\{A_t\}$ as group activity summary A. For each group, we measure the quality of group activity summary based on three temporal performance metrics, TUD(A,t), TCD(A,t) and TAD(A,t) for $t\in[T_s,T_e]$ where T_s and T_e is the start and end of the group data. We normalize the time span of the group performance into [0.1,1] and average over all groups. The evaluation results are shown in Figure 3. Our method, JAM, outperforms baseline methods for all three metrics, with a 27.2% improvement on the average. Our method differentiates users/concepts by the types and times of actions, so the extracted users are diversely related to more different concepts.

We also illustrate how our method captures the dynamics of group activities by using the example group "Lego Freaks" (ref. Figure 1). More qualitative analysis of the mining results is presented in a full version of this work.

Conclusion and Future Work

We propose a method for summarizing and representing social activity over time. In our framework, we formulate the summarization problem as extraction of representative activity themes. This summarization framework helps identify who (users), what (concepts), how (actions) and when (time) to represent the collective activities. We conducted extensive experiments on real-world datasets. Based on quantitative evaluation and qualitative observation, we demonstrate that our method can construct meaningful community activity summary. There are some open issues: (1) constructing summary at different time resolution; (2) incorporating user feedback for evaluating the activity summary. We plan to address these issues in future work. Additional research directions include combining media content analysis to support rich media summarization.

References

[1] A. AGARWAL, et al. (2006). *Learning to rank networked entities*, SIGKDD, 14-23, 2006.

[2] M. DUBINKO, et al. (2006). *Visualizing Tags over Time*, WWW, 193-202, 2006.

[3] J. M. KLEINBERG (1999). *Authoritative sources in a hyperlinked environment.* J. ACM **46**(5): 604-632.

[4] R. KUMAR, et al. (2006). *Structure and evolution of online social networks*, SIGKDD, 611-617, 2006.

[5] S. ZHU, et al. (2007). *Combining content and link for classification using matrix factorization*, SIGIR, 487-494, 2007.