

Recommending Interest Groups to Social Media Users by Incorporating Heterogeneous Resources

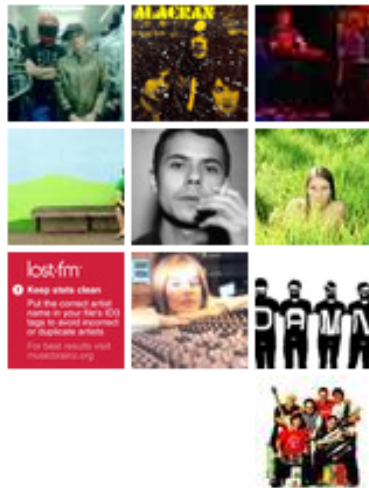
Wei Zeng and Li Chen
(lichen@comp.hkbu.edu.hk)

Department of Computer Science
Hong Kong Baptist University

Motivation

Traditional recommendations

Recommended Artists (see all)



- ▶ Grafton Primary
- ▶ Alacran
- ▶ BumblebeeZ
- ▶ Andrea Echeverri
- ▶ Colder
- ▶ Muscles
- ▶ Damn Arms
- ▶ Fabiana Cantilo
- ▶ Portishead & Moloko
- ▶ Jarabe de Palo















Our focus: recommending interest groups

Groups are based on a common interest, an artist or genre, or anything really!

Community » Groups

Recently Active Groups

- **People with no social lives that listen to more music than is healthy who are...**
46,021 members
Latest activity: [rewoorc800](#) joined this group.
- **Ayria**
1 member
Latest activity: [HectorOmarRoman](#) connected [Ayria](#) to the group [Ayria](#).
- **Metalheads who don't give a fuck if they are true or not**
9,225 members
Latest activity: [mindsphere](#) added the article [2013 Ratings \(in progress\)](#) to this group.
- **I'd die without music**
19,994 members
Latest activity: [YunSan](#) joined this group.
- **Zombie Slayers United**
60 members
Latest activity: [guitarheroine2](#) connected [45 Grave](#) to the group [Zombie Slayers United](#).
- **lastfmlogos.info**
1,376 members
Latest activity: [marlos783](#) joined this group.
- **GD Game Threads**
449 members
Latest activity: [ToluTolu](#) joined this group.
- **Addicted to Last.fm**
16,887 members
Latest activity: [iNLITY](#) joined this group.
- **Flagcounter**
3,507 members
Latest activity: [Wasurezaki](#) joined this group.
- **For those who don't sleep enough due to staying up late at night for no apparent...**
19,844 members
Latest activity: [CelesteMarque](#) joined this group.
- **The Extreme Music Forum**
108 members
Latest activity: [Registeel](#) left this group.
- **The Ancient (Dis)Order of the Last FM Platinum Round Table**
55 members
Latest activity: [deculus](#) joined this group.

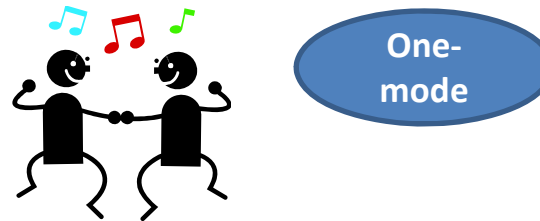
Challenge

- Data sparsity problem
 - Last.fm, 100,000 users, user-item pairs **29,908,020**, but **user-group pairs 1,132,281**
- Related work
 - **Community/affiliation recommendation** based on graph proximity model (Vasuke et al. RecSys'10), combinational collaborative filtering (Chen et al., KDD'08), or Latent Dirichlet Allocation (Chen et al., WWW'09)
- **But**, few have fully incorporated other available resources to further increase the recommendation accuracy

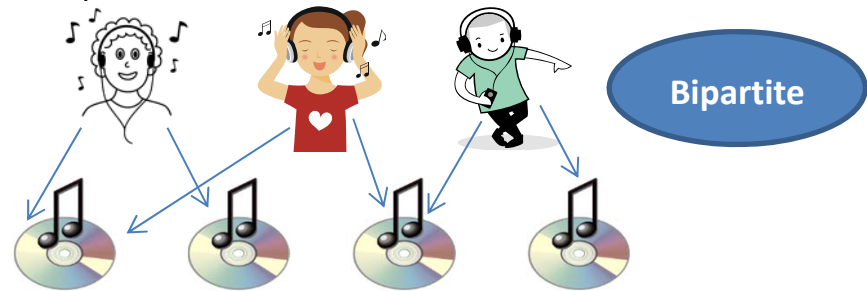
Our methodology

- To fuse two auxiliary resources

- User-user friendship



- User-item preferences



- Research questions

- *How to fuse, due to the different properties?*

- *Which resource takes more effective effect?*

- *What about their combined effect?*

Algorithm

- Fusion framework: **Matrix Factorization**
 - Advantages: scalability, efficiency, potential accuracy
- To fuse friendship
 - Regularization model
 - Advantage: for minimizing the gap between two entities
- To fuse user-item preferences
 - Factorization model
 - Advantage: for effectively factorizing user-item relations into two components

To fuse friendship

Basic matrix factorization
of user-group matrix

equals 1 if the user u joined group g

$$\min_{u^*, g^*} \sum_{u, g} c_{ug}^* (\underline{p_{ug}^*} - x_u^T z_g)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_g \|z_g\|^2 \right)$$

$$+ \lambda_f \left(\left\| x_u - \frac{1}{|F(u)|} \sum_{f \in F(u)} \widehat{sim}(u, f) x_f \right\|^2 \right)$$

Regularization of
user-user
friendship

coefficient for the friendship regularization

normalized similarity degree between the user u and her/his friend f , based on common items, common groups, or common friends

To fuse user-item preferences

Basic matrix factorization of user-group matrix

$$\alpha \min_{u^*, g^*} \sum_{u, g} c_{ug}^* (p_{ug}^* - x_u^T z_g)^2 + \lambda (\sum_u \|x_u\|^2 + \sum_g \|z_g\|^2) +$$

$$(1 - \alpha) \min_{u^*, i^*} \sum_{u, i} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda (\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2)$$

Factorization of user-item matrix

equals 1 if the user u clicked item i (implicit feedback)

used to adjust the relative weights of user-item matrix

Cont.

- Alternatively, for the comparison purpose

$$\min_{u^*, g^*} \sum_{u, g} c_{ug}^* (p_{ug}^* - x_u^T z_g)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_g \|z_g\|^2 \right) + \lambda_f \left(\left\| x_u - \frac{1}{N(u)} \sum_{n \in N(u)} \omega_{un}^* * x_n \right\|^2 \right)$$

user u 's neighbors who have common items with u

weight of similarity between two users u and n , based on their common items

Regularization of user-item relation

To fuse them together

Basic matrix factorization of user-group matrix

$$\alpha \min_{u^*, g^*} \sum_{u, g} c_{ug}^* (p_{ug}^* - x_u^T z_g)^2 + \lambda (\sum_u \|x_u\|^2 + \sum_g \|z_g\|^2) +$$

$$\lambda_f (\|x_u - \frac{1}{|F(u)|} \sum_{f \in F(u)} \widehat{sim}(u, f) x_f\|^2) +$$

$$(1 - \alpha) \min_{u^*, i^*} \sum_{u, i} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda (\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2)$$

Factorization of user-item matrix

Regularization of user-user friendship

Experiment

	Element	Size	Element	Size
Last.fm	#user	100,000	#user-item pair	29,908,020
	#item	22,443	#friendship pair	583,621
	#group	25,397	#user-group pair	1,132,281
Douban	#user	71,034	#user-item pair	12,292,429
	#item	25,258	#friendship pair	273,832
	#group	2,973	#user-group pair	373,239



Evaluation: leave-one-out, **hit-ratio metric**

Results – fusing user-item preferences

	<i>Last.fm</i>		<i>Douban</i>	
Method	Hits@5	Hits@10	Hits@5	Hits@10
Group.MF (baseline)	0.0530	0.0875	0.1995	0.2933
<i>Fusing user-item preferences (via Factorization)</i>				
Group.MF.I.F@train.20	0.0573	0.0899	0.2030	0.2950
Group.MF.I.F@train.40	0.0678	0.1026	0.2102	0.3013
Group.MF.I.F@train.60	0.0714	0.1068	0.2113	0.3079
Group.MF.I.F@train.80	0.0722	0.1070	0.2120	0.3095
<i>Fusing user-item preferences (via Regularization)</i>				
Group.MF.I.R@train.20	0.0559	0.0885	0.2025	0.2932
Group.MF.I.F@train.40	0.0559	0.0885	0.2026	0.2936
Group.MF.I.R@train.60	0.0560	0.0886	0.2026	0.2936
Group.MF.I.R@train.80	0.0561	0.0887	0.2027	0.2937
<i>Fusing friendship</i>				
Group.MF.F.R	0.0566	0.0910	0.2072	0.2973
Group.MF.F.F	0.0553	0.0876	0.2038	0.2928
Group.MF.F.FCos	0.0549	0.0861	0.2075	0.2974
Group.MF.F.GCos	0.0593	0.0923	0.2093	0.2999
Group.MF.F.ICos	0.0569	0.0897	0.2062	0.2921

Result 1: the accuracy of factorization model (Group.MF.I.F) is improved with the increase of the density

Result 2: the accuracy of regularization model (Group.MF.I.R) is lower and does not obviously change when the data density varied

Results – fusing user-user friendship

	<i>Last.fm</i>		<i>Douban</i>	
Method	Hits@5	Hits@10	Hits@5	Hits@10
Group.MF (baseline)	0.0530	0.0875	0.1995	0.2933
<i>Fusing user-item preferences (via Factorization)</i>				
Group.MF.I.F@train.20	0.0573	0.0899	0.2030	0.2950
Group.MF.I.F@train.40	0.0678	0.1026	0.2102	0.3013
Group.MF.I.F@train.60	0.0714	0.1068	0.2113	0.3079
Group.MF.I.F@train.80	0.0722	0.1070	0.2120	0.3095
<i>Fusing user-item preferences (via Regularization)</i>				
Group.MF.I.R@train.20	0.0559	0.0885	0.2025	0.2932
Group.MF.I.F@train.40	0.0559	0.0885	0.2026	0.2936
Group.MF.I.R@train.60	0.0560	0.0886	0.2026	0.2936
Group.MF.I.R@train.80	0.0561	0.0887	0.2027	0.2937
<i>Fusing friendship</i>				
Group.MF.F.R	0.0566	0.0910	0.2072	0.2973
Group.MF.F.F	0.0553	0.0876	0.2038	0.2928
Group.MF.F.FCos	0.0549	0.0861	0.2075	0.2974
Group.MF.F.GCos	0.0593	0.0923	0.2093	0.2999
Group.MF.F.ICos	0.0569	0.0897	0.2062	0.2921

Result 3: the regularization model (Group.MF.F.R) outperforms the factorization model (Group.MF.F.F)

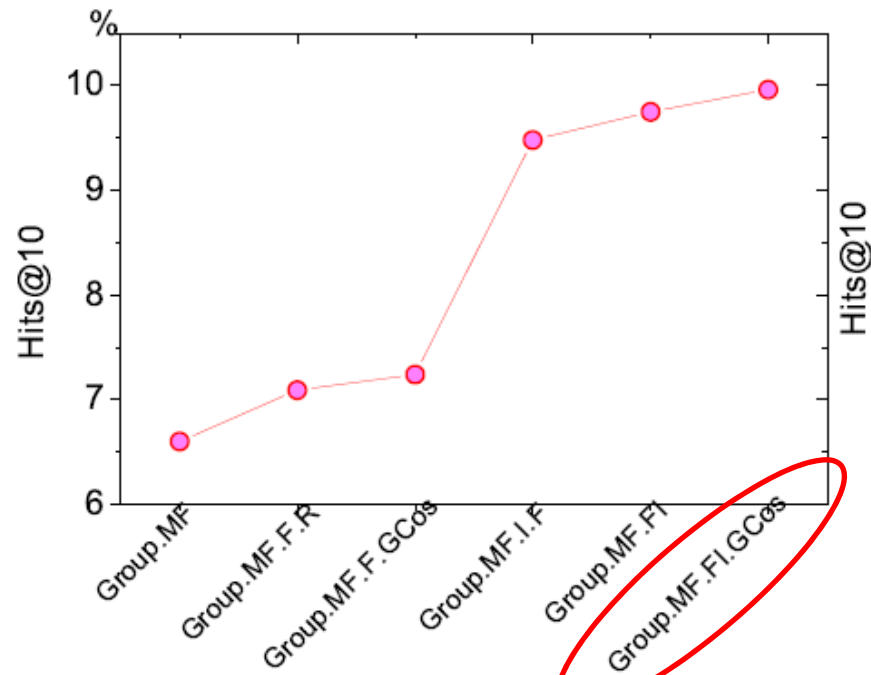
Result 4: the integration of group-based similarity measure (Group.MF.F.GCos) outperforms the others

Results – comparison of the two resources

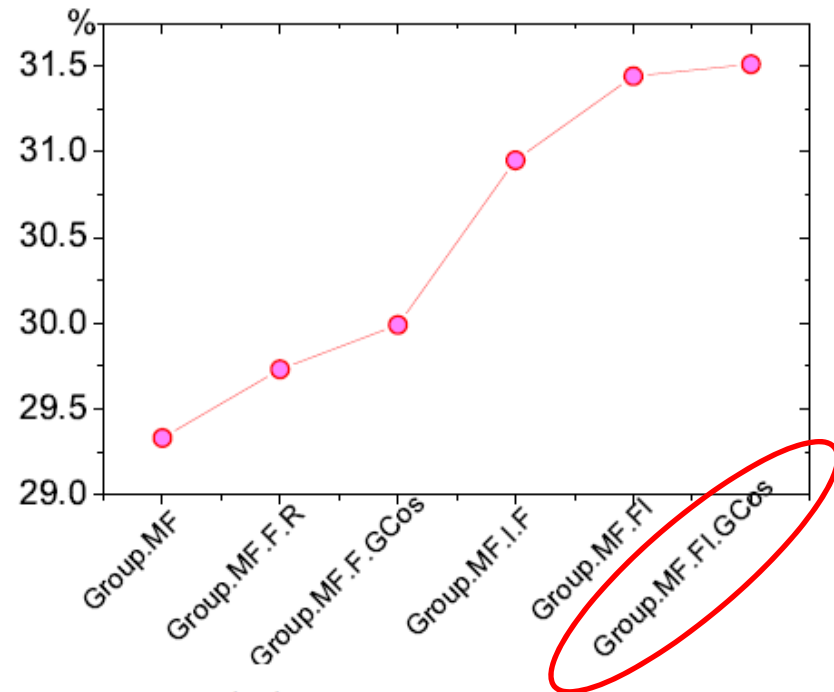
	<i>Last.fm</i>		<i>Douban</i>	
Method	Hits@5	Hits@10	Hits@5	Hits@10
Group.MF (baseline)	0.0530	0.0875	0.1995	0.2933
<i>Fusing user-item preferences (via Factorization)</i>				
Group.MF.I.F@train.20	0.0573	0.0899	0.2030	0.2950
Group.MF.I.F@train.40	0.0678	0.1026	0.2102	0.3013
Group.MF.I.F@train.60	0.0714	0.1068	0.2113	0.3079
Group.MF.I.F@train.80	0.0722	0.1070	0.2120	0.3095
<i>Fusing user-item preferences (via Regularization)</i>				
Group.MF.I.R@train.20	0.0559	0.0885	0.2025	0.2932
Group.MF.I.F@train.40	0.0559	0.0885	0.2026	0.2936
Group.MF.I.R@train.60	0.0560	0.0886	0.2026	0.2936
Group.MF.I.R@train.80	0.0561	0.0887	0.2027	0.2937
<i>Fusing friendship</i>				
Group.MF.F.R	0.0566	0.0910	0.2072	0.2973
Group.MF.F.F	0.0553	0.0876	0.2038	0.2928
Group.MF.F.FCos	0.0549	0.0861	0.2075	0.2974
Group.MF.F.GCos	0.0593	0.0923	0.2093	0.2999
Group.MF.F.ICos	0.0569	0.0897	0.2062	0.2921

Result 5: the user-item preferences act more positive than the friendship in terms of enhancing group recommendation

Results – combined effect



(a) Last.fm



(b) Douban

Result 6: combination of Group.MF.F.GCoS and Group.MF.I.F@train.80 for fusing the two resources friendship and user-item preferences together achieves accuracy improvement

Conclusion

- Fused both **friendship** and **user-item preference** data to improve the accuracy of recommending interest groups
- Proved the outperforming suitability of **regularization model** for handling the one mode friendship data, and the **factorization model** for processing the user-item bipartite data
- Future work: more auxiliary resources, more algorithm comparisons