

Generalized Group Profiling for Content Customization

Mostafa Dehghani¹

Hosein Azarbonyad²

Jaap Kamps¹

Maarten Marx²

¹Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands

²Informatics Institute, University of Amsterdam, The Netherlands

{dehghani, h.azarbonyad, kamps, maartenmarx}@uva.nl

ABSTRACT

There is an ongoing debate on personalization, adapting results to the unique user exploiting a user's personal history, versus customization, adapting results to a group profile sharing one or more characteristics with the user at hand. Personal profiles are often sparse, due to cold start problems and the fact that users typically search for new items or information, necessitating to back-off to customization, but group profiles often suffer from accidental features brought in by the unique individual contributing to the group. In this paper we propose a generalized group profiling approach that teases apart the exact contribution of the individual user level and the 'abstract' group level by extracting a latent model that captures all, and only, the essential features of the whole group.

Our main findings are the followings. First, we propose an efficient way of group profiling which implicitly eliminates the general and specific features from users' models in a group and takes out the abstract model representing the whole group. Second, we employ the resulting models in the task of contextual suggestion. We analyse different grouping criteria and we find that group-based suggestions improves the customization. Third, we see that the granularity of groups affects the quality of group profiling. We observe that grouping approach should compromise between the level of customization and groups' size.

Keywords: Group Profiling, Contextual Suggestion, Content Customization.

1. INTRODUCTION

Context is pervasive on the modern web, due to cloud-based and mobile applications, making every information access interaction part of an eternal user session. Effective ways to leverage this context are key to further enhancing the user experience, both in terms of better quality of results as in terms of easier ways to articulate complex information needs. This requires both effective ways of personalization

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHIIR'16, March 13–17, 2016, Carrboro, NC, USA

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-3751-9/16/03...\$15.00

DOI: <http://dx.doi.org/10.1145/2854946.2855003>

to an individual user as well as customization to a profile based on groups of users. For group level analysis, there is a need for extracting a group profile that captures the essence of the group, separate from the sum of the profiles of its individual members. This profile should be "specific" enough to distinguish the preferences of the group from other groups, and in the same time, "general" enough to capture all shared tastes, expectations, and similarities of its members.

Group profiling can help understand both explicit groups, like Facebook groups, and implicit groups, like groups extracted by community detection algorithms. There is a wide range of applications for group profiling, like understanding social structures [9], network visualization, recommender systems [1, 4, 8], and direct marketing [3]. One of the important applications of group profiling is in the content customization problem. Content customization is the process of tailoring content to individual users' characteristics or preferences. However, using individual preferences for content customization is not always possible. For example sometimes there is a new user in the system with no historical interactions and no rich information about the preferences, or sometimes the user is not able to determine his/her preferences explicitly. In these situations, group based content customization would be beneficial to suggest content to the user based on the preferences of the groups that the user belongs to.

The main aim of this paper is *to develop a language model for a group of users based on the group preferences, which contains all, and only, essential shared commonalities of the group members, and to employ such a group profile to customize content suggestion for individual users*. We break this down into three concrete research questions:

- RQ1** *How to estimate models for group profiles capturing exactly the shared commonalities of their members?*
- RQ2** *How effective are group profiles to customize content suggestion for individual users?*
- RQ3** *How does the user's group granularity affect the quality of the group's profile?*

There is various research done on the task of group profiling which, given the individual attributes and preferences, aims to find out group-level shared preferences [6, 7]. Tang et al. [9] presented three different methods for group profiling: *Aggregation*, which tries to find features that are shared by the whole group; *Differentiation*, which tries to extract features that can help to differentiate one group from others; and *Egocentric differentiation*, which tries to extract features that can help to differentiate members of one group from the neighbour members. In recent work, Hu et al.

[4] proposed a deep-architecture model to learn a high level representation of group preferences.

For group recommendation there is research on building a model of a group by forming a linear combination of the individual models [5]. Some of them construct the group’s preference model on the basis of individual preference models, using a notion of distance between preference models [10]. Some approaches try to divide the group into several categories of homogeneous users and specify the preference model for each subgroup. Then they create the group model as a weighted average of the subgroup models, with the weights reflecting the importance of the subgroups [2].

In this paper, we assume that the model of a user’s preferences is a mixture of general, specific and his/her group preferences, and we try to extract the latent group model. We utilize the extracted profile in the task of contextual suggestion to improve content customization based on a user’s group memberships.

The rest of the paper is structured as follows. First, in Section 2, we explain our approach for group profiling in detail. Section 3 presents the results of our experiments on the task of group-based contextual suggestion as well as some analysis on the effect of group granularity on the quality of group profiling. Finally, Section 4 concludes the paper and discusses possible extensions and future work.

2. GROUP PROFILING

In this section, we investigate our first research question: “How to estimate models for group profiles capturing exactly the shared commonalities of their members?”

Group profiling refers to the task of extracting a descriptive model for a group of users which addresses the particular aspects that bind group members together. The goal of the proposed method for group profiling is to extract the latent common language model of the group which represents the shared group preferences. We assume that there are three different models from which the group members sample to express their preferences: *group* model, *general* model, and *specific* model. The group model is supposed to represent the shared preferences of the group. The general model represents the general terms that anybody may use (very common observed terms), and specific model represents the terms that are attached to an individual user’s preferences but not others (partially observed terms).

Each model is a distribution over terms. We consider collection language model θ_c as the general model:

$$p(t|\theta_c) = \frac{c(t, C)}{\sum_{t' \in V} c(t', C)} \quad (1)$$

This way, terms that are well explained in the collection model get high probability and are considered as general terms. To estimate the probability of terms given the specific model, θ_s , we use the Equation 2 and normalize all the probabilities to form a distribution:

$$P(t|\theta_s) = \sum_{u_i \in G} \left(P(t|\theta_{u_i}) \prod_{\substack{u_j \in G \\ j \neq i}} (1 - P(t|\theta_{u_j})) \right) \cdot idf_G(t) \quad (2)$$

where u_i is the user i in group G , and θ_{u_i} is the language model representing the user’s preferences. Here, $idf_g(t)$ represents inverse document frequency of term t in group G . Equation 2 calculates probability of term t to be an specific

term. To this end, it simply considers the probability of a term to be important in one of the user models but not others, marginalized over all user models as well as considering group document frequency. In this way, terms that are well explained in only one user model but not others get high probability and are considered as specific terms.

Based on the generative model, each term in a user model is generated by sampling from a mixture of these three models—group, general (or collection), specific—independently. Thus, the probability of generating term t in user model u would be:

$$p(t|u) = \lambda_{u,g}p(t|\theta_g) + \lambda_{u,c}p(t|\theta_c) + \lambda_{u,s}p(t|\theta_s), \quad (3)$$

where $\lambda_{u,x}$ stands for $p(\theta_x|u)$ which is the probability of choosing model θ_x given the user u .

The goal is to fit the log-likelihood model of generating all terms in the user models to discover the exact term distribution of the group model, θ_g . Let $G = \{u_1, \dots, u_n\}$ be a group of users. The log-likelihood of the group would be:

$$\log p(G|\Lambda) = \sum_{u \in G} \sum_{t \in V} c(t, u) \log \left(\sum_{x \in \{g, c, s\}} \lambda_{u,x} p(t|\theta_x) \right), \quad (4)$$

where $c(t, u)$ is the frequency of term t in user model u and Λ determines the set of all parameters:

$$\Lambda = \{\lambda_g, \lambda_c, \lambda_s\}_{u \in G} \cup \{\theta_g\} \quad (5)$$

As we have mentioned, we estimate θ_c and θ_s based on the collection language model as well as the patterns of terms occurrences in the documents and we make them fixed in the model. Finally, to fit our model, we estimate the parameters using the maximum likelihood (ML) estimator. So we solve the following problem:

$$\Lambda^* = \underset{\Lambda}{\operatorname{argmax}} p(G|\Lambda) \quad (6)$$

Assuming $X_{u,t} \in \{g, c, s\}$ as a hidden variable indicating which model has been used to generate term t in user model u , we can compute the parameters efficiently using Expectation-Maximization (EM) algorithm. The stages of EM algorithm would be:

E-Step

$$p(X_{u,t} = x) = \frac{\lambda_{u,x} p(t|\theta_x)}{\sum_{x' \in \{g, c, s\}} \lambda_{u,x'} p(t|\theta_{x'})} \quad (7)$$

M-Step

$$p(t|\theta_g) = \frac{\sum_{u \in G} c(t, u) p(X_{u,t} = g)}{\sum_{t' \in V} \sum_{u \in G} c(t', u) p(X_{u,t'} = g)} \quad (8)$$

$$\lambda_{u,x} = \frac{\sum_{t \in V} c(t, u) p(X_{u,t} = x)}{\sum_{x' \in \{g, c, s\}} \sum_{t \in V} c(t, u) p(X_{u,t} = x')} \quad (9)$$

After convergence of the EM algorithm, all the parameters are estimated including the group model, θ_g , which is a distribution over terms representing shared group preferences as well as $\lambda_{u,g}$, $\lambda_{u,c}$, and $\lambda_{u,s}$ for each user u which determine the contribution of each model in each user’s preferences.

In this section, we proposed a generative user model as a mixture of group, general (or collection), and specific models, and showed that the latent distribution of group model over terms can be extracted as the group’s profile.

3. EXPERIMENTS

In this section, we present our experiments to evaluate the effectiveness of the estimated language model of groups in the task of contextual suggestion. Furthermore, we analyse the effect of group granularity on group profiling. We first explain the data collection used in our experiments and then present the evaluation results.

3.1 Data Collection

In this research, we have made use of the TREC 2015 contextual suggestion¹ Batch task dataset. Contextual suggestion is the task of searching for complex information needs that are highly dependent on both context and user interests. The dataset contains the information from 207 users including their age, gender, and set of rated places or activities as the user preferences (rates are in the range of -1 to 4). The task is to generate a list of ranked suggestions from a set of candidate attractions, by giving the user information as well as some information about the context, including location of trip, trip season, trip type, tripe duration, and the type of group the person is travelling with. For each user, we consider rated suggestions that are annotated with rates of more than 2 as relevant. Furthermore, we generate the user language models as a mixture of their relevant preferences considering the rates. Based on the information in the dataset, we divide users into several groups. Groupings are based on the users information and context information. Table 1 presents grouping criteria, the groups, and number of users in each group.

3.2 Group Profiling for Contextual Suggestion

In this section, we investigate our second research question: “How effective are group profiles to customize content suggestion for individual users?”

We generate group-based rankings of suggestions to evaluate the quality of group profiles in content customization. To this end, one of the grouping approaches given in Table 1 is chosen, e.g. based on users’ age. Then we estimate language model of each group employing the approach explained in Section 2. Afterward, regarding the information of the given request, i.e. the user information and context information, the group which the user belongs to is selected and based on the similarity of the language model of the selected group and the language model of candidate, the ranked list of the suggestions is generated.

Beside the group-based ranking, we generate a ranked list of suggestions based on the preferences of the user as a baseline. To do so, a language model is estimated as the mixture of the model of user preferences regarding their ratings and based on the similarity of the preferences language model and the candidate language model, a ranked list is generated.

Furthermore, according to the explanation in Section 2, the contribution of each of *specific*, *group*, and *general* models in each user model is learned as the model parameters, i.e. $\lambda_{u,s}$, $\lambda_{u,g}$, and $\lambda_{u,c}$. Having these parameters empowers us to efficiently combine the group-based model with the preferences-based model for content customization. To this end, we smooth the preferences-based model of user with both the group model and the general model using JM-smoothing employing the learn parameters. To evaluate

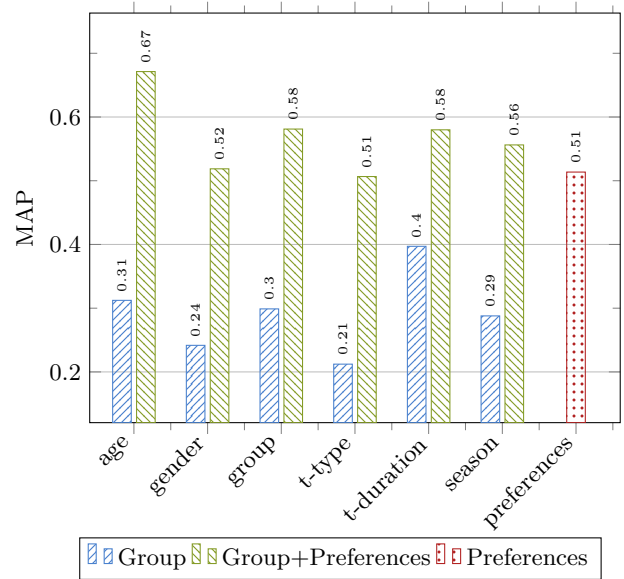


Figure 1: Performance of employing user preferences-based and group-based customization on contextual suggestion task. Improvements of combining group-based and preferences-based approach over the preferences-based approach and corresponding group-based approaches are statistically significant based on one-tailed t-test, with p-value < 0.05.

the quality of the combination, we have done experiments considering different grouping criteria.

Figure 1 presents the performance of employing different grouping approaches for group-based suggestion as well as preferences-based suggestion. The combinations of preferences-based suggestion and group-based suggestion are also reported.

As can be seen, among the group-based strategies, suggestions based on the duration of the trip is the most effective strategy. Also age of the user and the type of the group the user travels with, are rather important while type of the trip is not so important. This could be due to the fact that most of the time, the user’s interests and beloved attractions do not change based on the type of trip which could be “business” or “holiday”. On the other hand, combining the preferences-based suggestions with group-based suggestions in all grouping strategies leads to improvement. This means in case of incompleteness of user’s profile, customizing the content based on the groups that user belongs to, implicitly fills the missing information and improves the performance of suggestions. However, this depends on the quality of the groups profiles that should reflect essential common (not general, not specific) characteristics of the groups.

3.3 Effect of Group Granularity

In this section, we investigate our third research question: “How does the user’s group granularity affect the quality of the group’s profile?”

In the grouping stage, sometimes users can be grouped based on different levels of granularity. For example, having the age of users, discretization can be done using binning with different sizes of bin. In this section, we analyse the effect of granularity of groups, and consecutively the size of the groups with a fixed volume of train data, on the quality of group profiling.

¹<https://sites.google.com/site/trecontext/trec-2015>

Table 1: Statistics of users groups resulted by grouping based on different criteria

Grouping Criterion	Age					Gender		Group Type				Trip Type				Trip Duration				Season			
	< 20	20 - 30	30 - 40	40 - 50	> 50	male	female	alone	family	friends	other	holiday	business	other	night out	day trip	weekend	longer	spring	summer	autumn	winter	
Groups																							
Group Size (#users)	9	87	60	22	23	107	91	23	101	65	4	176	7	12	8	17	99	74	20	99	52	22	

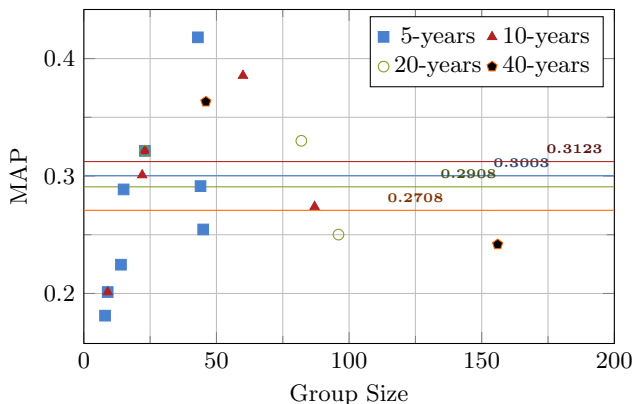


Figure 2: Effect of groups granularity on the performance group profiling

We have selected “age” of users as the grouping criterion and tried different bin sizes for discretization: 5 years, 10 years, 20 years and 40 years. Figure 2 shows the quality of groups profiles on different sizes of groups in the task of contextual suggestion. Each point in the figure represents a group of users and its position determines its size and the performance of group-based contextual suggestion for the users within the group. Moreover the horizontal lines represent overall performance of different levels of granularity. As can be seen, since the number of sample users is limited fine-grained grouping leads to having smaller groups. So small number of samples affects the group profiling quality and slightly decreases the performance. While coarse-grained grouping leads to having large groups that leads to not being able to adequately customize the group profile.

In our dataset, 10 years granularity for “age” has the best performance since the formed groups are big enough so that the group profiling approach is able to estimate high quality models, and they are small enough so that the group profiles are easily distinguishable which leads to a more effective customization.

4. CONCLUSIONS

In this paper, we dealt with the problem of group profiling. The main aim of this paper was to develop a language model for a group of users based on the group preferences, which contains all, and only, essential shared commonalities of the group members, and to employ such a group profile to customize content suggestion for individual users

Our first research question was: *How to estimate models for group profiles capturing exactly the shared commonalities of their members?* We proposed to consider each user preferences as a mixture of general, specific and its group preferences and estimated the latent group preferences as the shared concerns among all group members.

Our second research question was: *How effective are group*

profiles to customize content suggestion for individual users? We utilized the proposed group profiling approach for the task of contextual suggestion, and our experimental results showed that considering group-based suggestions along with user preferences-based suggestions can improve the content customization.

Our third research question was: *How does the user’s group granularity affect the quality of the group’s profile?* We designed an experiment to investigate how group granularity may affect the group profiling quality. We found that the grouping approach should result in groups that are big enough to enable group profiling to infer high quality models and small enough to enable the extracted model to make customization for group members.

As the future work, we are going to find a way for learning how to employ group-based suggestions on the basis of different grouping criteria simultaneously. For example, how to combine suggestions based on the age with suggestions based on the gender. A further development would be to evaluate the proposed group profiling approach on other tasks and other kinds of data including non-textual data.

Acknowledgments This research is funded in part by the European Community’s FP7 (project meSch, grant # 600851) and the Netherlands Organization for Scientific Research (WebART project, NWO CATCH # 640.005.001; ExPoSe project, NWO CI # 314.99.108; DiLiPaD project, NWO Digging into Data # 600.006.014).

References

- [1] S. Amer-Yahia, S. B. Roy, A. Chawlat, G. Das, and C. Yu. Group recommendation: Semantics and efficiency. *VLDB*, 2:754–765, 2009.
- [2] L. Ardissono, A. Goy, G. Petrone, M. Segnan, and P. Torasso. Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. *Applied Artificial Intelligence*, 17(8-9):687–714, 2003.
- [3] B. Custers. Effects of unreliable group profiling by means of data mining. In *Discovery Science*, pages 291–296, 2003.
- [4] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and W. Cao. Deep modeling of group preferences for group-based recommendation. In *AAAI*, 2014.
- [5] A. Jameson and B. Smyth. The adaptive web. chapter Recommendation to Groups, pages 596–627. 2007.
- [6] J. Masthoff. Group recommender systems: Combining individual models. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*, pages 677–702. Springer US, 2011.
- [7] C. Senot, D. Kostadinov, M. Bouzid, J. Picault, and A. Aghasaryan. Evaluation of group profiling strategies. In *IJCAI*, pages 2728–2733, 2011.
- [8] S. Shang, Y. Hui, P. Hui, P. Cuff, and S. Kulkarni. Beyond personalization and anonymity: Towards a group-based recommender system. In *SAC*, pages 266–273, 2014.
- [9] L. Tang, X. Wang, and H. Liu. Group profiling for understanding social structures. *TOIS*, 3(1):15:1–15:25, 2011.
- [10] Z. Yu, X. Zhou, Y. Hao, and J. Gu. Tv program recommendation for multiple viewers based on user profile merging. *User Modeling and User-Adapted Interaction*, 16(1):63–82, 2006.