

# Concept maps for Personalized Interest Management

Swaroop Kalasapur, Henry Song, Doreen Cheng,

Samsung R&D Center,  
95 W. Plumeria, San Jose, CA 95134  
{s.kalasapur, hsong, d.cheng}@sisa.samsung.com

**Abstract.** To assist users to access wanted services and information within a short attention span, personalization has gained tremendous importance. While systems such as recommenders for specific domains have been successfully deployed, there is no existing user-centric mechanism that can be utilized for all application domains. By capturing user's interests and relationships among such interests, it is possible to provide personalization for a wide range of applications. In this paper, we present our initial investigations in constructing concept maps for user interest management. Based on common sense, we have attempted to build a generic concept map that can be utilized for recommendation purposes to address the cold-start problem. We have also presented our experiment in generating personal concept maps, that are derived purely based on data corresponding to a particular user. Our aim is to create a platform for personalized interest management. We provide our observations, challenges and insights based on our experience.

**Keywords:** Personalization, Interest management, Concept Maps, Knowledge based personalization.

## 1 Introduction

Personalization is about customizing a variety of services according to user preferences. While current personalization techniques such as recommendations for shopping, movies, etc. [1, 2] aim at providing domain-specific personalization support, they do not account for implications of user actions out side of the domain. The devices that a user can use to interact with various services are proliferating, and it is now possible to capture user interests in a variety of domains for supporting across domain personalization. With mobile devices becoming the primary mode of information access, it becomes more important to manage user interests on a user device such as a mobile phone that can cater to a number of applications across multiple domains.

Collaborative technologies for personalization emphasize on the similarity among users to derive per-user metrics [3]. While such mechanisms are very useful for services such as online stores, they fall short of being applicable to support a wide range of user needs. Collaborative filtering relies on users' rating data. Typically, users rate only a very small portion of the entire item set, which leads to a very sparse population of the rating data. Under such condition, collaborative filtering uses other

users' interests for prediction using well-known cold-start techniques. A simple example is when a new item is added, since there is no information about users' preference on this new item, the system will have to wait until enough data is gathered, before making any decisions on the item. A variety of solutions have been employed for cold-start. For example, hybrid solution combines content-based [4] with collaborative filtering. Such an approach, however, is only applicable to text-based content. Other approaches combine ontology with collaborative filtering [1, 5], which requires domain knowledge expertise. Knowledge based recommender systems [7] attempt to avoid the cold-start problem, but they either require extensive user interaction to discriminate among options, or need detailed knowledge about the user. The concept map based approach presented in this paper can be effectively used as a vehicle for capturing user interest which can then be used by knowledge based recommenders.

The user assistance through personalization can be maximized if user interests can be gathered from evidences on different devices with which the user interacts. For example, with powerful mobile devices having become constant companions and primary mode of access to digital resources, a majority of user interests can be collected. By building solutions based on the observed evidence from the user activities on their mobile devices, we can provide personalization support to a large number of applications.

Many personalization efforts also take advantage of semantic computing. Research in semantic computing has focused on creating ontologies that represent the knowledge within a given domain. Although very promising, ontology based approach is still an art being practiced by a handful of scientists and domain experts. Since a domain can be represented in many ways, coming to a consensus on domain ontology is not a trivial challenge. When it comes to addressing multiple domains, corresponding to multiple user interests, integrating multiple domain ontologies is a very challenging task. Since each domain is modeled separately by corresponding domain experts, there is no easy way to manage cross-domain knowledge in an elegant manner.

Concept maps have been employed by many to facilitate a wide ranging solutions in education and learning, modeling, visualization, etc. Through concept maps, it is possible to capture the relationships among various concepts and represent them in a fashion that can be digested both by humans and computers. We have utilized concept maps to capture the relationships among the various user interests. By capturing such relationships, we aim to answer questions such as "Is the user interested in cars and how likely is he interested in restaurants?"

In this paper, we present two broad types of concept maps. The first is a generic concept map that is aimed at capturing relationships among various interest items from general population. The main purpose of the generic concept map is to capture common-sense relationships among various concepts in a field, and use it as a basis for personalization. The second type of concept maps is personal in nature. We attempt to build a concept map purely based on any evidence available from user usages of the devices. By keeping track of the usages, we can collect a variety of information about the user. We can then use this collected information to derive relationships among various concepts. The derived relationships among various concepts can guide a personalization scheme such as recommendation, by providing a means for reasoning based on user interests.

## 2 Concept Maps

Concept maps are graphical tools for organizing and representing knowledge [6]. Typically, the knowledge within a domain is identified through a number of concepts and the relationships among such concepts are also identified. Concept maps are built using the identified concepts as nodes and the relationships among concepts as edges between them. There are many efforts in literature that illustrate the applicability of concept maps in fields such as visualization, education and learning, etc. Concept maps provide very flexible and efficient facilities for knowledge representation. With such flexibility, it is possible to easily express the relationships among various concepts that make up a domain and cross domain.

In our investigation into concept map, we have started with one simple relationship that captures the relative interest of the user. We model the concept map as a directed graph with nodes representing concepts and the edges from node A to node B representing the possibility of the user being interested in B, given the user interest in A. The weights along the edges represent the strength of the relationship. An example concept map is shown in Figure 1. When we know the user interest in one of the concepts, we can utilize the concept map to reason about user interests in various other concepts. From Figure 1, if we know that the user has some interest in 'Health', there is a 50% chance that the user is also interested in 'Hiking', so if the user has given a rating of 5 for 'Health', there is a 50% chance that the user will give 'Hiking' a rating of 5.

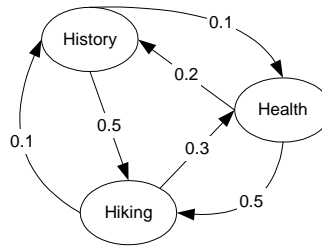


Figure 1. An example concept map.

### 2.1 Generic Concept Map

One of the major problems in personalization is the problem of cold start. A personalization system typically will have to wait for the user to utilize the system for a period of time before the system can start the process of personalization. We employ generic concept maps to address this problem. By identifying the relationships among various concepts, based on information available through generic sources, we can construct the generic concept map. The rationale behind this is to utilize the common sense to act as a starting point for personalization.

To construct the generic concept map, we start with a list of concepts. These concepts can belong to a single knowledge domain or cross domains. With each concept, we perform a query to a (set of) generic search engine(s) to retrieve a large number of results. We analyzed 1000 retrieved search results and computed the

conditional probabilities of occurrence of all other concept terms given the concept in the query. The probabilities are interpreted as “also interested” relations. This gives the relationships between concept in the query and all others. The process is repeated for all of the selected concepts. An example of the retrieved relationships is presented in Table 1. We have manually examined relationships in the generated generic concept map. Common sense tells us that the relationships are reasonably good. However, we have not verified it with prediction accuracy.

**Table 1.** An example of generated generic concept map.

| <i>Concept</i>               |   | 1    | 2    | 3    | 4    | 5    | 6    |
|------------------------------|---|------|------|------|------|------|------|
| Automotive                   | 1 | 1.00 | 0.12 | 0.19 | 0.00 | 0.05 | 0.06 |
| Legal and Financial Services | 2 | 0.04 | 1.00 | 0.21 | 0.00 | 0.10 | 0.05 |
| Computer and Internet        | 3 | 0.01 | 0.02 | 1.00 | 0.00 | 0.07 | 0.06 |
| Personal Care                | 4 | 0.02 | 0.01 | 0.14 | 1.00 | 0.08 | 0.10 |
| Education and Instruction    | 5 | 0.02 | 0.06 | 0.28 | 0.00 | 1.00 | 0.25 |
| Entertainment and Arts       | 6 | 0.01 | 0.01 | 0.17 | 0.00 | 0.13 | 1.00 |

With the generic concept maps, we can address the cold start problem in personalization. One of the major personalization tasks is recommendation. If we know the user interest in one of the concepts, we can utilize this information to recommend other concepts that the user is most likely to be interested in. We can utilize the following relationship for the purpose.

$$P_j = \left( \sum_{i=1}^k W_{i,j} \times R_i \right) / \sum_{i=1}^k W_{i,j}$$

Where  $P_j$  denotes the prediction on  $j^{th}$  keyword;  $W_{i,j}$  denotes the relation from  $i^{th}$  concept to  $j^{th}$  concept;  $R_i$  is the valid user interest level on the  $i^{th}$  concept.

## 2.2 Personal Concept Maps

While generic concept maps are very useful in capturing knowledge within a domain and across domains, we speculated that a personalized concept map might be more useful for supporting personalization, especially since user interests may change over time; new concepts may need to be added, old ones deleted, and the relationships updated. While personalizing a generic concept is a viable approach, we wonder whether we could directly generate personalized concept maps from user’s usage data. To investigate the feasibility of this approach, we conducted an experiment trying to automatically build a concept map using the usage data we collected from an eight-user three-month mobile phone usage data.

During the experiment we encountered a few challenges. The first challenge is to derive interested concepts from unstructured sources such as email text, SMS messages, URLs visited, documents received or edited. We employed the Yahoo’s Term API to extract the interested concepts. Given a text segment, Yahoo’s Term API attempts to identify terms that qualify as representative indicators to the supplied segment and we can directly use the identified terms as concepts to build the personal concept map. The strength of relationships among the various concepts can be directly

derived as the co-occurrence coefficient among the various identified concepts. Table 2 shows a snapshot of the concepts identified and the co-occurrence frequencies among the concepts based on 45 email messages for a single user.

**Table 2.** Snapshot of a personalized concept map.

| <i>Term</i>                   | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 15 |
|-------------------------------|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|
| <i>4<sup>th</sup> of July</i> | 1 |   |   |   |   |   |   |   |   |   |    |    |    |    |    |
| <i>American countries</i>     | 0 | 1 |   |   |   |   |   |   |   |   |    |    |    |    |    |
| <i>Costa Rica</i>             | 0 | 1 | 1 |   |   |   |   |   |   |   |    |    |    |    |    |
| <i>Europe summer</i>          | 0 | 1 | 1 | 1 |   |   |   |   |   |   |    |    |    |    |    |
| <i>Free kicks</i>             | 0 | 1 | 1 | 1 | 1 |   |   |   |   |   |    |    |    |    |    |
| <i>Love quote</i>             | 0 | 0 | 0 | 0 | 0 | 3 |   |   |   |   |    |    |    |    |    |
| <i>Mexico study</i>           | 0 | 0 | 0 | 0 | 0 | 0 | 2 |   |   |   |    |    |    |    |    |
| <i>California</i>             | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |   |   |    |    |    |    |    |
| <i>Study question</i>         | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |   |    |    |    |    |    |
| <i>Nutrition</i>              | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |    |    |    |    |    |
| <i>Performing arts</i>        | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1  |    |    |    |    |
| <i>poem</i>                   | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1  | 1  |    |    |    |
| <i>Spanish study</i>          | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1  | 0  | 1  |    |    |
| <i>Parents</i>                | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 0  | 0  | 2  |    |
| <i>San Luis Obispo</i>        | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0  | 0  | 1  | 0  | 2  |

The second challenge is very little we can derive from the extracted terms as shown in Table 2. There can be two reasons for this low co-occurrence among identified concepts. First of all, Yahoo’s term extraction gives better results if contextual information, such as the field of text being examined, is available. But such topical information is almost always not available in user data. Therefore, more powerful mechanisms based on natural language processing are needed for identifying concepts from these usage data. Secondly, the inspected user data covers a variety of interest domains and there can be a large set of concepts derived. But very few relationships among the concepts can be derived, possibly due to the small dataset.

The third challenge is that to use concept maps for modeling a user’s preferences, we need to capture both positive and negative relationships, where a negative relationship between concept A and concept B means that a user’s likes in A indicates the user’s dislike in B. This may require natural language processing techniques.

The fourth challenge is that a very large corpus of usage text may be required in order to build a useful concept map from usage data since the relationship between  $n$  concepts is  $O(n^2)$  where  $n$  is the number of concepts. This means that a user must first use the device for a very long time before a map can be constructed.

The fifth challenge is to distinguish the scope of a particular term. A single term such as *Apple* can have various meanings based on the usage context. And, therefore can be candidate concepts with varying relationships. Therefore, we need a mechanism that can accommodate such semantic differences.

Another challenge is shortened representations of words and terms commonly used in SMS e.g. abbreviations like ‘LOL’, ‘TTYL’, ‘BRB’. A filtering mechanism or a

mechanism to translate them into proper forms, such as the urban dictionary [8] is needed in order to use SMS text for the purpose.

### 3 Conclusion and Future Work

In summary, we believe that concept maps are useful tools in personalization. With powerful personal devices, it is now possible to provide personalized support without compromising user privacy. We explored ways to automatically construct cross-domain generic concept maps and learn that it is feasible to do so using search engines and the information available in the Web. We also experimented using the text from usage logs to build a personal concept map and found that in order to build a useful concept map, large corpuses and natural language processing tools are needed. We speculate that the most practical approach could be to use the Web to build generic concept maps for cold start and use user's usage data to personalize the generic concept map(s). We would like to verify this in the future research.

### References

1. Schickel, V. and Faltings, B., "Using Hierarchical Clustering for Learning the Ontologies used in Recommendation Systems", In ACM SIGKDD, pp 599 – 608, 2007
2. Papadogiorgaki, M., Papastathis, V., Nidelkou, E., Kompatsiaris, I., Waddington, S., Bratu, B., and Ribiere, M., "Distributed User Modeling for Personalized News Delivery in Mobile Devices", In Semantic Media Adaptation and Personalization, pp 80 – 85, 2007
3. Herlocker, J., Konstan, J., Borchers, A., and Riedl, J., "An Algorithmic Framework for Performing Collaborative Filtering". In ACM SIGIR'99, 1999
4. Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., and Sartin, M., "Combining Content-based and Collaborative Filters in an Online Newspaper", In SIGIR'99 Workshop on Recommender Systems, 1999
5. Schickel, V., and Ozden, B., "Inferring user's preferences using ontologies", In AAAI, pp 1413 – 1418, 2006
6. Novak, J. D. & A. J. Cañas, The Theory Underlying Concept Maps and How to Construct Them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition, 2008
7. Burke, R. Knowledge-based Recommender Systems. In A. Kent (ed.), Encyclopedia of Library and Information Systems. Vol. 69, Supplement 32. New York: Marcel Dekker, 2000.
8. Urban Dictionary: [www.urbandictionary.com](http://www.urbandictionary.com)