Automatic Learning of User Interests for Personalized Commmunication Services

Matthias Strobbe, Olivier Van Laere

Supervisor(s): Filip De Turck, Bart Dhoedt

Abstract— **In view of the overwhelming popularity of user generated content new intelligent services are needed to filter this content based on personal interests. In this paper we present a set of algorithms for retrieving content, based on dynamic user profiles and learning capabilities. To illustrate the approach taken, a rich communication service is presented.**

I. Introduction

During the last few years, we have witnessed the emergence of a whole range of popular web sites hosting user generated content (Flickr [1], YouTube [2], etc.). The amount of content is so overwhelming that users are experiencing more and more problems identifying content matching their interests. Intelligent services taking personal interests into account can provide an efficient way of filtering and ranking the content in such a way that the user easily finds the content he wants. To achieve this goal, the metadata attached to the content (usually tags) has to be matched with the user interests and user feedback has to be taken into account carefully to keep the user interests up to date.

II. User Interests Matching

A. Keyword Tree

In our approach user interests are modeled using a keyword tree with added weight values. Top level keywords represent categories. Lower level keywords represent subcategories and specific interests. A weight value represents the importance of a keyword for a certain user. These weight values are adapted when input or feedback is received from the user.

B. Algorithms

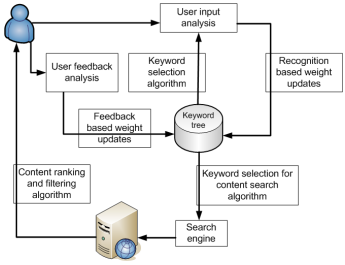

Figure 1. Overview of the interactions between the algorithmic components

Figure 1 shows a general overview of the interactions between the algorithms.

The *User Input Algorithm* increases the weight value of a keyword that was recognized from the input of the user (e.g. entered search terms or by speech analysis). We assume that these recognized keywords are an indication of the interests of the user. The *Keyword Selection for Content Search Algorithm* makes a selection of search terms based on the recognized keyword and related keywords with high weight values. After a search the results are ranked by the *Content Ranking Algorithm*. Ranking is based on the matching between the tags attached to the content and the keywords used for the search. When a user clicks on a returned result, the *User Feedback Algorithm* will be noti-

M. Strobbe is with the Department of Information Technology, Ghent University - IBBT, Gent, Belgium. His research is financed by a Ph. D. fellowship of the Fund for Scientific Research - Flanders (F.W.O.-Vlaanderen). E-mail: Matthias.Strobbe@intec.ugent.be .

fied and will increase the weight values of the tags attached to the content. The increases are bigger than for the *User Input Algorithm* as user feedback tells a lot about someone's interests. The *Keyword Selection Algorithm* identifies a relevant subset (the 'current keyword list') of the keyword tree. This list keeps track of the keywords that can be recognized for a user.

III. Use Case: Content Selection Based on Communication

Using the outlined algorithms we implemented a use case to provide users of a multimedia chat client with content that is an added value to their conversation, i.e. pictures about the topic they are discussing that at the same time match with their interests. The users chat with each other via SIP and the contents of their conversation are analyzed by a text analyzer. When a keyword is recognized, pictures are looked up in Flickr [1] based on the recognized keyword and related keywords. The resulting pictures are sent to the users and shown as a slideshow. When a user clicks on a picture to see it in more detail the *User Feedback Algorithm* is notified and the keyword tree of the user is updated.

IV. Evaluation

To evaluate our algorithms a number of simulations were performed with arbitrary keyword trees.

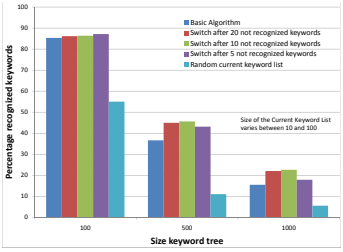

Figure 2. Comparison of the basic and advanced Keyword Selection Algorithm

In a first series of simulations the *Keyword Selection Algorithm* was tested for its adaptivity to switch to the current topic of the conversation. Two versions of the algorithm were tested: a basic version where keywords from the old topic are gradually replaced by keywords from the new topic and a more advanced version that recovers to a general current keyword list when no keywords are recognized during a certain period. Figure 2 shows the results in comparison with a random current keyword list.

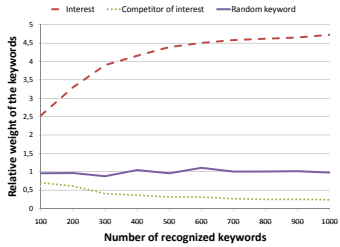

Figure 3. Comparison of the evolution of keyword weights

In a second series of simulations, the impact of the feedback on the weight values of the keywords was verified. The relative weight values of three keywords were tracked: a keyword that is of interest to the user (the keyword often gets feedback), a keyword that is a sibling of the interest of the user and a random other keyword not competing with any interests. Figure 3 shows the results. The random keyword stays around the average relative weight value of 1, but the user interest reaches a much higher value and as a consequence the competitor a very low value.

V. Conclusion

In this paper we presented a new way to match user generated content with user interests. We used a keyword tree to model the interests of the user in combination with five algorithms to update the keyword tree, incorporate user feedback and select relevant keywords for communication services.

References

[1] http://www.flickr.com
[2] http://www.youtube.com