

Content-Based Filtering and Personalization Using Structured Metadata

A. Mufit Ferman, James H. Errico, Peter van Beek, M. Ibrahim Sezan

Sharp Laboratories of America

5750 N.W. Pacific Rim Blvd., Camas, WA 98607, (+1) 360-817-7557

{mferman, jerrico, pvanbeek, sezan}@sharplabs.com

ABSTRACT

Structured descriptions of multimedia content and automatically generated user profiles are used to filter content.

Categories & Subject Descriptors: H.3.3 [Information storage and retrieval]: Information Search and Retrieval, Information filtering, Relevance feedback; H.3.7 [Information storage and retrieval]: Digital Libraries - Standards, Systems issues; H.5.1 [Information interfaces and presentation]: Multimedia Information Systems - Video; H.2.1 [Database management]: Logical Design - Data models; H.2.4 [Database management]: Systems - Multimedia Databases

General Terms: Algorithms

Keywords: Multimedia content, user profile, personalization, recommendation, XML, MPEG-7

1. OVERVIEW

User profiles that contain compact descriptions of users' interests and personal preferences provide a method for selecting from an increasing amount of multimedia content and for reducing information overload [1,2]. A user profile can be used to (a) filter input content, so that programs or items that the user has shown interest in are presented to the user, and/or (b) request from a content distribution service the programs of interest. Emerging standards such as MPEG-7 and TV-Anytime [3,4] provide an XML-based, interoperable framework for structured and rich description of multimedia content, along with tools that enable improved user interaction with such content, e.g., tools for representing users' preferences and history of content usage. In this paper, we propose novel algorithms for (i) automatically determining a user's profile from his/her usage history (profiling agent), and (ii) automatically filtering content according to the user's profile (filtering agent). The proposed algorithms are designed to support an MPEG-7 or TV-Anytime compliant description framework, although they can also be utilized in a non-standard environment.

We present a framework for integrated personalization and management of multimedia content (see Figure 1) which allows non-intrusive observation of user actions over time, enabling collection of consumption-related data without explicit user input. This usage history is subsequently utilized in conjunction with structured metadata that is available for each of the consumed content entities to identify what types of content the user prefers most. Specifically, various features are extracted from the content metadata and usage history database, and are then utilized by the proposed profiling agent that incorporates a fuzzy reasoning system to identify and update the user's preferences. The resulting preference descriptions may include a variety of descriptive attributes, such as genres, actors, channel information, etc., each of

which can be assigned a weight reflecting its desirability by the user. Furthermore, the preference descriptions support multilevel, hierarchical grouping of different attributes, and thus provide a highly flexible way of supporting a variety of applications. These descriptions are subsequently compared against the content metadata by the proposed filtering agent, to determine which content entities best match the user's interests. Each individual component of a user preference description corresponds to a test or matching operation against components of descriptions of multimedia content. After individual preference components are tested, the results of these tests are combined into a single composite test result that reflects the user's overall preferences. The filtering agent combines the individual test results according to the hierarchy of the user preferences description, where each parent test is a combination of its children tests. The combination of children tests within a single parent may depend on the types of the elements being tested, and involves operators such as intersection, union, maximum, minimum, averages, etc.

The proposed system has been tested using the MovieLens dataset [5], and preliminary recommendation experiments using only genre information yield results that are as good as, or better than, those previously published in [1,2]. Results of experiments that utilize richer content and user preference description metadata will be reported during the conference. Also, a prototype content browser based on the proposed framework will be demonstrated.

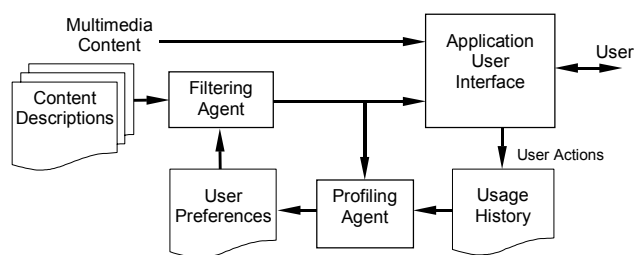


Figure 1. Content filtering and user profiling framework.

2. REFERENCES

- [1] Good, N. et al., Combining collaborative filtering with personal agents for better recommendations, Proc. AAAI'99, pp. 439-446, AAAI Press, Menlo Park, CA.
- [2] Sarwar, B. M. et al., Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system, Proc. ACM CSCW'98, p. 345, Seattle, WA.
- [3] MPEG, Introduction to MPEG-7, Sydney, 2001. <http://mpeg.telecomitalia.com>.
- [4] TV-Anytime Forum, <http://www.tv-anytime.org>.
- [5] GroupLens Research Project, <http://www.cs.umn.edu/Research/GroupLens>.