Using User Models in Music Information Retrieval Systems

Abstract

Most websites providing music services only support category-based browsing and/or text-based searching. There has been some research to improve the interface either for pull applications, e.g. query-by-humming systems, or for push applications, e.g. collaborative-filtering-based or feature-based music recommendation systems. However, for content-based search or feature-based filtering systems, one important problem is to describe music by its parameters or features, so that search engines or information filtering agents can use them to measure the similarity of the target (user's query or preference) and the candidates. MPEG7 (formally called "Multimedia Content Description Interface") is an international standard, which describes the multimedia content data to allow universal indexing, retrieval, filtering, control, and other activities supported by rich metadata. However, the metadata about the multimedia content itself are still insufficient, because many features of multimedia content are quite perceptual and user-dependent. For example, emotional features are very important for multimedia retrieval, but they are hard to be described by a universal model since different users may have different emotional responses to the same multimedia content. We therefore turn to user modeling techniques and representations to describe the properties of each user, so that the retrieval will be more accurate. Besides, user modeling can be used to reduce the search space, make push service easier and improve the user interface.

There are several important issues in user modeling for music information retrieval purpose or even more general multimedia retrieval. 1) How to model the user? User-programmed, machine-learning and knowledge-engineered methods can be used. 2) What information is needed to describe a user for music IR purpose? It may include both the user's indirect information (e.g. age, sex, citizenship, education, music experience, etc.) and direct information (e.g. user's interests, definition of qualitative features, appreciation habit, etc.). 3) How to represent, use and share the user model? Similar to MPEG7 concepts, we can use a standard language in text format to represent the user model, so that search engines or information filtering agents can use it to refine the result easily and efficiently, without repeating the long-time observation and learning of the user's behavior.

User modeling can be done on client-side or server-side. Issues including easy/hard to obtain the user information, hard/easy to use collaborative filtering model, far from/close to the music data, more/less privacy or safety, more scalable/higher load on the server, etc., need to be considered when choosing either of the paradigms.

We adopted the client-side user modeling paradigm in our MusicCat system. It is an agent that allows the user to define contexts and corresponding features of music that he wants to hear in those contexts correspondingly. Besides, the user can also define qualitative features of music based on quantitative features. For examples, the user just needs to tell the agent what kind of music he prefers to hear at what kind of context, like "I need fast and exciting music when I'm happy", "I need soft music to wake me up every morning at 8:00", "I need slow classical music, when I'm thinking", "I need rhythmic music when I'm walking", etc. Or, the user can define qualitative features, like "Romantic music for me means slow music with titles or lyric including word love", "My favorite music includes …" etc. Then, when the moment comes - the user tells the agent or a pre-defined time approaches, the agent can automatically, randomly and repeatedly choose music from the user's collection according to the pre-defined constraints. So far, MusicCat uses a profile-based user interface. And only midi files are used in the system.

In our system, we categorize the quantitative features of music into textual features including title, composer, genre, country, etc.; notation features including key signature, time signature, tonality, BPM, etc.; perceptual features including slow or fast, short or long, quiet or loud, non-rhythmic or rhythmic, and soft or exciting. User can define music corresponding to a particular context or qualitative feature based on these quantitative features. We manually input the textual information. Notation features were automatically extracted from midi files. Perceptual features were computed from the parameters including average duration per note, average tempo, tempo deviation, average pitch change per note, etc. The weights of each parameter contributing to each perceptual feature were set manually. Some psychoacoustic experiments should be necessary for more accuracy. To make the system more appealing, wireless communication technology can be used to make the system portable. We may also use sensors that can automatically detect the user context and then play corresponding music.

User information is very valuable and needs to be shared in the future to make universal information retrieval possible. Thus we propose an XML-like language, UMIRL (User Modeling for Information Retrieval Language), in which different systems may describe the user in this standard format to make the user model sharable and reusable. Although it is designed for music information retrieval purpose in this paper, it can be extended for general multimedia information retrieval as well.

To make music information retrieval systems more efficient, both user modeling techniques and descriptions are important. Those are prerequisites for open and efficient personalized services. We propose the user modeling language because there hasn't been much attention on the ways to share the valuable user data. We believe the main issues discussed in this paper will be the most significant but also the hardiest points in this research area.

Author Information

Wei Chai
Media Laboratory
Massachusetts Institute of Technology
chaiwei@media.mit.edu

Barry Vercoe
Media Laboratory
Massachusetts Institute of Technology
bv@media.mit.edu

Suggested Readings

Giorgio Braijnik, Giovanni Guida, Carlo Tasso. 1990. User Modeling in Expert Man-Machine Interfaces: a Case Study in Intelligent Information Retrieval . *IEEE Transactions Transactions on Systems, Man, and Cybernetics*, Vol 20. No.1, pp166-185.

Upendra Shardanand and Pattie Maes. 1995. Social Information Filtering: Algorithms for Automating "Word of Mouth". *Conference proceedings on Human factors in computing systems*, pp210-217.

Eric Scheirer. 2000. *Music-Listening Systems*, PhD dissertation.