

Moody Mobile TV: Exploring TV Clips with Personalized Playlists

Arne Berger¹, Robert Knauf¹, Maximilian Eibl¹, and Aaron Marcus²

¹ Chair Media Computer Science, Chemnitz University of Technology,
Strasse der Nationen 62, 09107 Chemnitz, Germany

{arne.berger, robert.knauf, eibl}@informatik.tu-chemnitz.de

² Aaron Marcus and Associates, Inc. (AM+A)

1196 Euclid Avenue, Suite 1F
Berkeley, California 94708-1640, USA
aaron.marcus@amanda.com

Abstract. An interface for filtering large video repositories for generating personalized playlists via navigation and selection of moods and emotion on a mobile device.

Keywords: mood navigation, customization, interface, query reformulation.

1 Users on the Move

The means of selecting and consuming audio-visual content in mobile settings varies significantly from ordinary domestic usage [1]. Users are in a hurry and have only a limited time available for enjoying entertainment on their mobile smart phones. Whether they are commuting or going out with friends, users usually avoid constructing complex search queries to find suitable content to watch. Besides time constraints, lack of concentration, contextual distractions or physical challenges (e.g., other people watching what they are doing or bright ambient light or too much noise in the subway or having to hold on to a strap in the subway and also to navigate a phone while keeping upright), obtrusive user interfaces, and constrictions from interpersonal circumstances, may prevent a search engine from gaining insight into the entertainment needs of the users.

1.1 Locally Driven Interest and Internet Based Video Consuming Behavior

Studies indicate that interest in local TV content also persists beyond the geographic broadcast limitation of individual stations. “Expats” and commuters eagerly look at “their” local station’s website to follow features from their hometown. Not least, this imparts a persistent bond to one’s hometown. While the local TV stations have a loyal target audience in large segments of the population, the target audience in the age group “under 30” is increasingly turning to the Internet as the primary medium for their video consumption. We seek to examine how the persistent interest in media coverage from one’s place of residence can be combined with the viewing habits of the Internet generation.

We strongly believe that a human component is crucial for recommendation based on emotions. Therefore we created and annotated in a crowd sourcing approach a database of 1000 videos in co-operation with local TV stations in Saxony (Germany) as a basis for personalized video playlists for mobile devices.

2 Simplicity and Plausibility of Video Recommendation

Finding interesting media items quickly and matching the users' entertainment requirements at just the right moment is difficult for both the content provider and the requesting consumers. The lack of both descriptive metadata and broad user feedback adversely affects retrieving suitable content inside a video portal with as few clicks as possible. Most often, the seeking consumers are left having to refine their search query several times or to use non-customized item lists such as "most viewed". Most commonly, retrieving media items referring to a local geographic area, a special interest, or a social group can only be achieved efficiently by following the "channels" of certain users or user groups. Social media sites such as www.youtube.com try to recommend relevant video clips as soon as users finish watching a clip. The system does not communicate the basis upon which a recommendation has been made to the users. Consequently, the algorithm might carry the user farther away from his entertainment source. Furthermore, the scheme on which the recommendation is carried out is a black box for users, leaving them with the experience that successive video clips are way too similar. Variants of the same video clip occur over and over again, or viewers are confused, because consecutive video clips seem to have no connection at all.

On the other hand, traditional TV stations handle the task of connecting single broadcasts very well. A moderator weaves a golden thread and guides viewers via this potentially emotional connection through a series of video clips. This TV format is well established and favorably to the algorithmic approach described above. It is, however, not readily transferable to the generation and consecutive play-out of personalized playlists.

Instead of providing a personal moderator, we introduce a semantic framework that enables the viewer to weave his own golden thread to select and connect interesting items. By applying readily available metadata from our database, this may be approached by selecting a list of video clips via time, genre and broadcasting station. This would allow answering four possible content searches: **When** did it occur, **What** is it about, **Where** did it happen and **Who** recorded it. However, all four lack any of the moderator's emotional features. For example, video clips about a traffic accident and offspring at the local zoo may follow right after each other in a playlist, just because both happen to occupy the same genre and happened around the same time. Since a human's moderation is lacking, the sequence might make sense, but may be disturbing for the audience.

3 Selecting Content vs. Constructing Playlists

Approaches to customize a content stream depending on users' interests usually learn by analyzing content description, genre data and users' viewing history. Consequently, they enable entertainment systems to adapt themselves to their owners' long-term habits. What they do not consider is the fact that peoples' individual moods vary at frequent and irregular intervals. To meet these requirements, media items need to be classified using emotional descriptions along with giving viewers an instrument to express their current mood.

Well-established systems that allow users to select content based on mood or emotion concentrate greatly on a single domain. To a certain extent, they are capable of plausibly recommending thematically related media objects as a playlist. Services such as www.last.fm, www.putpat.tv, or www.tape.tv merely contain music videos. Regarding this specific genre, recommendation is simply achievable, because (due to the domain limitation) it is based on a similarly limited plausible vocabulary. On the site www.tape.tv one can select material according to the genre or a mood. However, the moods cannot be chosen from a complete list of all imaginable possibilities, but from a list generated by the site's editors. It is also peculiar that one cannot combine several of these filter parameters. On the other hand, the site www.putpat.tv uses several slide controls (the parameters of which can be randomly occupied) to "adjust the station". These degrees of freedom enable competing queries such as "more Madonna, less pop"; at the same time. The concept of self-configurable sliders encourages users to try different things but, again, lacks an extensive list of moods to choose from.

While www.putpat.tv and www.tape.tv introduce novel, yet rudimentary, emotion based content filters, navigation functionality remains limited, compared to those available on www.youtube.com or www.last.fm. The first-mentioned two allow to skip the video clip currently playing in order to "ban it forever" or "love it", while the latter two evaluate user interaction and choices to aid the underlying algorithm in its learning. Both concepts are solely based on recommendation strategies based on an initial query in connection with users' listening histories.

All introduced services lack a well-defined emotional vocabulary and focus on mere algorithm based decisions. Both restrictions render the concept of a personalized playlist highly rigid and not customizable. In order to accumulate content based on emotion or moods, two main aids are missing in the concepts described above: First, the aforementioned list of emotions and second, a set of navigation options to allow a real personalization.

4 Using TV-Anytime to Contribute Emotional Metadata

In the recent past services for the purpose of TV personalization has been proposed. Obviously, all of them are subjected to descriptive and structured metadata. To unify this common feature, the TV-Anytime standard has been introduced to define metadata schemes, especially concerning genre and broadcast related data. [2] Subsequently, the standard influenced research efforts on TV services driven by users' needs, such as [3.]. Generally, emotional classification has been left out as the

basis of recommendations is still limited to genre data, content descriptions and users' long-term viewing history. Moreover, the TV-Anytime standard offers an emotion-based content description scheme consisting of 53 adjectives, which has been proved in reflecting emotional impacts to the viewers [4., 5.]. Regarding the advantages of a cross-system standard describing emotions and providing an anticipated well-structured order, our work takes a closer look at the suitability of the given listing.

5 Implementation and Evaluation

5.1 Design Methods

The creation of a personalized playlist is "more of an Art than a Science" [6]. The mere application of an algorithm is not enough for most users. [6] Therefore we adapted and applied a couple of design methods in order to gain insight into potential users mindsets regarding the construction of personalized playlists and the use of content filters, especially mood filters in a TV context. First, we asked participants to draw cognitive maps of possible ways for filtering video repositories down to satisfying playlists. This was done to get an insight how users think and associate when formulating content filters. Second, based on the subsequent results we crafted paper prototypes and discussed the sketched out possibilites for navigating and selecting with potential users in open-ended questionaries.

Questions we wanted to address are:

Which means are convenient for users to individually select video clips as a playlist?

Which means are necessary for users to perceive these playlists as such?

Is mood a key mean for that?

Are the moods suitably proposed by the

TV-Anytime subset for the local TV domain?

5.2 Moody TV - Navigation

With moody TV users shall be able to select and combine content filters fluently and independently. To find out what content attributes users are looking for, we asked participants to map out a virtual space of content properties and show how they thought to navigate within it. This method usually helps to discover pathways and interests in which people make sense of a particular content space. The results eventually help to make sense of how to construct queries for filter specification. Users were asked to individually draw a map or diagram of what comes to their mind when being on the move and having a mobile video handset available, whether sitting on public transportation alone or being in a pub with friends. The six users, four male, two female, between the age of 23 and 34 are all regular viewers of local TV. They had 15 minutes time to draw a map or scheme and were asked to freely associate parameters to form a personalized playlist. A discussion with all participants followed.

The results lead to the assumption that users indeed are interested in direct emotional filters. Most of the user generated maps feature emotion clusters or the simple question “how” in a list of questions. See Fig. 1 and 2 for excerpts of users cognitive maps.

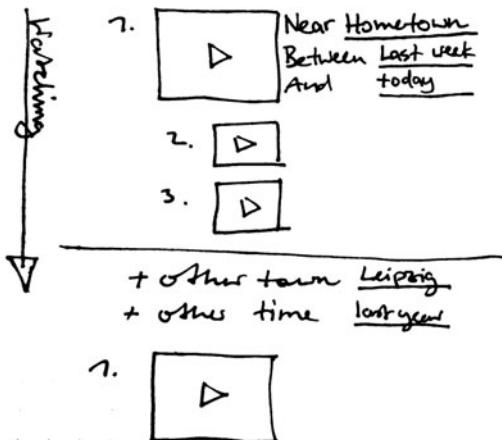


Fig. 1. Cognitive map drawn by a user

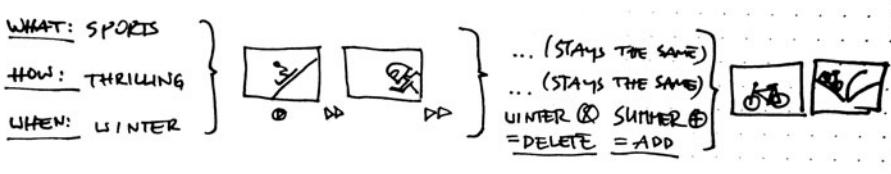


Fig. 2. Cognitive map drawn by a user

Given the mindset of being on the move, users formed questions from a simple vocabulary and subsequently wanted to change only certain parameters after watching a few video items. Although five users out of six proposed emotion in some way as a mean to filter, the results are far from being readily transformable into an interface. It is plausible to assume, that emotion is a relevant feature. However, cooperating with linguists may be necessary in the future to optimize semantic significance of the description vocabulary, which is for now beyond the scope of this work.

5.3 Moody TV - Prototyping

Based on the findings about emotion filters described above, we proposed a low fidelity prototype containing a filter named “How” together with more filters based on the other four cardinal questions Who, Where, When, What mentioned above.

To prove the concept we introduced it to twelve users between the age of 17 and 44. We rapidly sketched a series of mobile device screens with the main objective of constructing a personalized playlist with five clicks at the maximum. See Fig. 3. Findings are discussed in depth in [7].



Fig. 3. From [7]

Users' feedback on this approach was insightful in two ways. On one hand, users at large expressed their general approval on the advantages that might arise by constructing exhaustive content filters with just a few steps of interaction. On the other hand, the pre-structured characteristic was heavily criticized. The rigidly defined prototype inspired participants to incredibly rich feedback. It may be generalized as following:

All twelve users criticized the seven proposed moods as too general. Most users proposed mood sub-categories. *The landscape of moods in users' heads needs to be supported via a detailed browsing option for moods.*

Subsequent questions revealed that users have significantly varying cognitive styles. Although not statistically significant, at least two different styles have been discovered. *The decision to categorize moods is subject to the inclusion of different cognitive styles.*

Applying plain cardinal questions as categories has obscured the users. Six of our test subjects asked for a brief explanation of the questions and their combination. *We conclude that an interface that asks questions, should ask in a more conversational way.*

Three users criticized the interface proposal as too rigid for supporting granularity. Most users discovered that some categories need sub-categories, while others did not agree to that. *The interface needs to support this.*

Three users missed an option to sort the subsequent playlist. This could be addressed via the order in which users answer relevant questions.

The proposal of a low fidelity prototype in combination with open-ended questions has proved to be a fast and convenient way to gain user feedback on a large variety of issues without a lot of explanation. The main insight is, that all users found and used the filter option “how”. Most user feedback was given on only this feature.

6 Future Work: Moody TV - Emotion

Contrary to the proposed prototype a complete emotional vocabulary is featured in [2]. While a set of 53 emotions is sufficient to describe every kind of audiovisual data, it has not been widely implemented for selecting emotions, due to the lack of predefined sets of metadata. When content is tagged based on those metadata, the mere number of possible emotions makes it hard to subsequently navigate and select them in the form of, e.g. a list, a menu or even a tag-cloud.

Up to this point, the system has converged into an interface that asks questions in an order that is determined by the user.

Based the previous result that users’ cognitive styles are different, the next step is to adapt a design method to sort these emotions into categories that make sense to users in order to navigate the 53 option with ease.

Acknowledgement. This work was accomplished within the project sachsMedia – Cooperative Producing, Storage and Retrieval, funded by the program Entrepreneurial Regions of the Federal Ministry of Education and Research, Germany (FKZ: 03IP608). The authors are responsible for the content of this publication.

References

- [1] Schildbach, B., Rukzio, E.: Investigating selection and reading performance on a mobile phone while walking. In: Proceedings of the 12th international conference on Human computer interaction with mobile devices and services, pp. 93–102 (2010)
- [2] TV Anytime, pp. 1–160 (December 2010),
<http://www.etsi.org/deliver/etsits/102800102899/1028220301/01.02.0160/ts1028220301v010201.pdf>
- [3] Zhang, H., Zheng, S.: Personalized tv program recommendation based on tv-anytime metadata. In: Proceedings of the Ninth International Symposium on Consumer Electronics (ISCE 2005), pp. 242–246 (2005)
- [4] Lee, H.K., Kim, J.G., Yang, S.J., Hong, J.: Personalized tv services based on tv-anytime for personal digital recorder. IEEE Transactions on Consumer Electronics 51(3), 885–892 (2005)

- [5] Petersen, M., Butkus, A.: Extracting Moods from Songs and BBC Programs Based on Emotional Context. *International Journal of Digital Multimedia Broadcasting* (2008)
- [6] Cunningham, S.J., Bainbridge, D., Falconer, A.: More of an art than a science: Supporting the creation of playlists and mixes. In: *Proceedings of the 7th International Conference on Music Information Retrieval*, pp. 240–245 (2006)
- [7] Berger, A., Knauf, R., Eibl, M.: Constraints and simplification for a better mobile video annotation and content customization process. In: *Workshop Proceedings of the EuroITV*, 1–4 (April 2010)