Deep Modeling: Emotion-Based Modeling of Video Content Viewers

Shlomo Berkovsky
NICTA
Locked Bag 9013
Alexandria NSW 1435
Australia
Shlomo.Berkovsky@nicta.com.au

Sebastien Ardon
NICTA
Locked Bag 9013
Alexandria NSW 1435
Australia
Sebastien.Ardon@nicta.com.au

Proceedings of TVUX-2013: Workshop on Exploring and Enhancing the User Experience for TV at ACM CHI 2013,
27 April 2013, Paris, France.

Abstract
The acquisition of accurate user models is a challenging task that aggravates the generation of accurate video recommendations. In this paper, we propose a novel concept of deep user modeling, i.e., modeling of user emotions over the course of watching video content items. We argue that deep models can accurately reflect user preferences, and facilitate generation of high-quality personalized recommendations.

Author Keywords
Video content, user modeling, emotions, personalization, recommendations.

ACM Classification Keywords
H.5.2. User interfaces: Theory and methods; H.5.2. User interfaces: Input devices and strategies.

General Terms
Design, human factors.

Introduction
The convergence of IPTV services with the Web and the Social Web has resulted in unprecedented volumes of accessible video content. The nature of this content is very diverse: the quality varies from user-generated clips to multi-million dollar movies; the technology varies from black-and-white films to modern HD

effects; and the viewing platform varies from card-size smartphones to wall-size LCD screens. The amount and diversity of the accessible content are overwhelming and bring to the fore the need for intelligent systems that can deliver to users accurate and personalized video *recommendations* [1].

In order to generate high-quality recommendations, a recommender system needs to possess information about its users, typically encapsulated in the *user models*. These can be captured either explicitly (questionnaires, user feedback) or implicitly (mining observed user-system interactions) [2]. A rich, reliable, and up-to-date user modeling data is generally a scarce and hard-to-obtain resource. In the context of video content consumption, the granularity of the obtained user modeling data is often too coarse-grain, aggravating the generation of recommendations. For example, consider a 3-star rating given to a 3-hour movie. It is evident that this is a mediocre rating and the user probably liked some features/parts of the movie and disliked others. However, given a single movie rating, it is impossible to separate between the liked and disliked features and derive accurate user modeling data. Hence, in this work we discuss the use of emotions as a rich source of user modeling data.

Specifically, we present here the novel notion of *deep user modeling*, i.e., ongoing capturing, modeling, and mining of user emotions elicited by the video content over the course of watching. We present a method for obtaining emotion-based user modeling data, outline the use of deep models in the recommendation process, and discuss several limitations of deep modeling for video content recommendations.

Capturing Deep Models

It is evident that a single feedback given by a user after watching a complex video item expresses an aggregate opinion only. That is, it communicates the overall impression, but fails to reflect user preference towards specific features (actors/effects) or parts (plot/scene) of the video item. How can this fine-grained deep user modeling data be acquired? We argue that user emotions captured over the course of watching video content items can serve as a proxy for their preference towards various features of the watched videos.

Hence, we outline an explicit method for capturing user emotions, which requires users to specify the experienced emotions through a dedicated interface component. We presume that the most suitable location for this component would be a second-screen application, so not to clutter the main screen and minimize user distraction. Figure 1 depicts a second-screen emotion-capturing application used for crowd-sourcing purposes. We hypothesize that users may be motivated to express more emotions and use the interface more frequently, if it was integrated with a second-screen social media application.

Figure 1. Second-screen emotion capturing interface mock-up

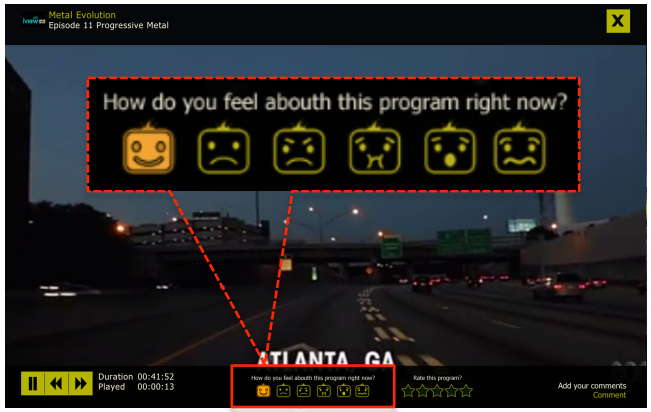

Figure 2. Main screen emotion capturing interface

Alternatively, the emotion interface can be integrated into the main screen. Figure 2 depicts the video player of our experimental Social TV platform [3]. There, six emotions are integrated as icons into the player's control bar. Users can activate the control bar at any given time and click on the icon representing the experienced emotion. The emotion clicks (along with their timestamps) are recorded, and the control bar disappears shortly afterwards, in order to allow users immerse back into the video.

The obtained user emotions can be considered as a labeling of the video content timeline (see example in Figure 3). The overall deep model of a user will be the union of the labeled video timelines. Unlike traditional user models, where the information units are triplets of *<user, item, rating>* or *<user, item_feature, score>*, deep models consist of complex triplets *<user, item, emotion_labeled_timeline>*. If expressed and mined accurately, a labeled video timeline contains more user information than an item rating or a feature score.

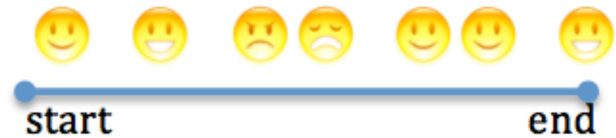

Figure 3. Emotion-labeled timeline of a video

Using Deep Models

The next question focuses on the ways to leverage the information encapsulated in the deep models for the generation of high-quality video recommendations. The intuition behind deep models is that the time-series of users emotions are indicative of the state of mind induced by the video content. This is typically a factor that cannot be easily expressed or captured, but may turn out instrumental in determining user enjoyment. We will exemplify two use-cases of deep models in two widely-used recommendation scenarios: collaborative and content-based filtering [1].

In collaborative filtering, deep models can be used to compute user-to-user similarity scores. These reflect the correlation of users' preferences and are computed through rating vectors, which are often sparse and have a limited overlap of rated items. The deep models can potentially facilitate a more accurate computation of the user-to-user similarity scores, even with only a few overlapping labeled video item timelines. The similarity scores can be used for personalized prediction of user ratings for yet unwatched video items.

Content-based filtering, however, requires some domain knowledge. Here we consider a simple knowledge, in form of content annotation of the video items (see a sample annotation in Figure 4). Coupling

0:35 Opening sequence; comet hits Earth

3:30 Comet fragments hit space shuttle; engineers scramble to understand

6:20 Shuttle explodes; asteroids begin to hit New York City

10:30 President learns of problem in space; asteroid is the size of Texas; 18 days until impact with Earth

12:00 Focus of films changes to Stamper Oil

17:00 Feasibility planning meeting at NASA; Oil is struck by Stamper Oil

22:30 Harry Stamper is taken to NASA to talk with engineers

27:00 Stamper begins to round-up his crew of roughnecks to assist NASA

Figure 4. Sample movie content annotation.

the content annotation of a video with the emotion-labeled timeline of a user can potentially improve the accuracy of ratings predictions for yet unwatched videos, and, next, of the generated recommendations.

Discussion

The scarcity of fine-grained user modeling data pertaining to specific features/parts of the watched videos motivated our work into the acquisition of deep emotion-based user models. In this position paper, we outlined our ideas around (1) capturing user emotions elicited over the course of watching video items, (2) mining these emotions to derive accurate deep user models, and (3) leveraging deep models in the recommendation process. We would like to discuss a few open issues surrounding the proposed ideas.

An interesting question refers to the emotions that can be accurately captured. There are several psychological studies into the emotions that videos elicit. One of the encompassing representations is the circumplex model of affect that positions emotions on a two-dimensional valence-arousal plane [4]. Another model includes six primary emotions that videos can elicit: anger, disgust, fear, joy, sadness, and surprise [5]. However, can the emotions be accurately perceived and expressed by users? For instance, recent user study discovered that the mood labeling can be reduced to two dimensions: seriousness and pace [6]. Can user emotions be further simplified to a binary like/dislike or thumbs up/down, or even to a unary negative feedback? Are there playful and creative ways to capture the emotions, e.g., through gamification and mobile interfaces?

Another practical concern refers to the feasibility of the proposed deep modeling approach. We posit that explicit emotion capturing is applicable for certain types of video content, e.g., talk shows or sport, but may turn inappropriate for others, e.g., movies. For the latter, users may want to immerse into the watched video content and refrain from labeling the video timeline. This would undermine the deep user models and aggravate the recommendation generation. Moreover, the proposed ideas may confront existing privacy regulations and ethical norms. This may preclude users from using such a system, regardless of the potential benefits of the recommendations. This should also be given a consideration.

Finally, we would like to highlight the pressing need for a thorough evaluation of the deep modeling ideas. We intend to leverage our Social TV platform to empirically validate the ideas outlined in this position paper.

References

[1] Ricci, F., Rokach, L., Shapira, B. Introduction to Recommender Systems Handbook. *Recommender Systems Handbook*, 2011.

[2] Berkovsky, S. Decentralized Mediation of User Models for a Better Personalization. In *Proc. AH 2006*, Springer, 2006.

[3] Ardon, S., Bensiali, S., Cinis, H., Wang, J., Berkovsky, S. Next-Generation Social TV Content Discovery. In *Proc. UMAP 2012*, Springer, 2012.

[4] Barrett, L.F., Russell, J.A. Independence and Bipolarity in the Structure of Current Affect. *Journal of Personality and Social Psychology*, 1998.

[5] Oliveira, E., Martins, P., Chambel, T. iFelt: Accessing Movies Through Our Emotions. In *Proc. EuroITV* 2011, ACM, 2011.

[6] Eggink, J., Bland, D. A Large Scale Experiment for Mood-Based Classification of TV Programmes, *White paper*, BBC, 2012.