Mining Affective Context in Short Films for Emotion-Aware Recommendation

Claudia Orellana-Rodriguez
Insight Centre for Data Analytics
University College Dublin
Dublin, Ireland
claudia.orellana@insight-centre.org

Ernesto Diaz-Aviles
IBM Research
Dublin, Ireland
e.diaz-aviles@ie.ibm.com

Wolfgang Nejdl
L3S Research Center
University of Hannover
Hannover, Germany
nejdl@L3S.de

ABSTRACT

Emotion is fundamental to human experience and impacts our daily activities and decision-making processes where, e.g., the affective state of a user influences whether or not she decides to consume a recommended item – movie, book, product or service. However, information retrieval and recommendation tasks have largely ignored emotion as a source of user context, in part because emotion is difficult to measure and easy to misunderstand. In this paper we explore the role of emotions in short films and propose an approach that automatically extracts affective context from user comments associated to short films available in YouTube, as an alternative to explicit human annotations. We go beyond the traditional polarity detection (i.e., positive/negative), and extract for each film four opposing pairs of primary emotions: joy–sadness, anger–fear, trust–disgust, and anticipation–surprise. Finally, in our empirical evaluation, we show how the affective context extracted automatically can be leveraged for emotion-aware film recommendation.

Categories and Subject Descriptors

H3.3 [**Information Search and Retrieval**]: Information filtering; K.4 [**Computer and Society**]

General Terms

Human Factors, Experimentation, Measurement

Keywords

Computational Social Science; Sentiment Analysis; Social Media Analytics; YouTube.

HT '15, September 1–4, 2015, Guzelyurt, Northern Cyprus.

ACM 978-1-4503-3395-5/15/09 ...$15.00.
DOI: http://dx.doi.org/10.1145/2700171.2791042.

1. INTRODUCTION

Amateur and professional filmmakers can reach large communities of viewers worldwide thanks to YouTube, which has drastically changed the way people access and critique films since its creation in 2005. This video-sharing platform provides its users with mechanisms – e.g., like/dislike buttons, comments – to publicly express their opinions about the uploaded movies, short films, or documentaries, thus shifting the role of users from being mere spectators to becoming active critics of the social media creations.

These user interactions, ratings and comments available in YouTube constitute a valuable source of data, which can be exploited to detect opinions, trends, or for ranking and recommendation tasks; however, the rapid pace at which users generate this content increases the difficulty of its analysis. Manually extracting useful information from different media is too slow, expensive, and requires considerable human effort, particularly if one needs to better understand users' emotions evoked by the videos.

In this paper, we explore two approaches for short film-emotion association. On the one hand, we obtain emotional annotations by crowdsourcing on Amazon Mechanical Turk and conduct an extensive study to better understand the affective content extracted by human intelligence from different short films and the role of emotions therein. On the other hand, we automatically extract affective context from the user-generated comments available in YouTube and empirically evaluate its usefulness by addressing an emotion-aware recommendation task. We do not limit our analysis to polarity extraction (positive-negative) but rather exploit Plutchik's four opposing emotion pairs (joy–sadness, anger–fear, trust–disgust, and anticipation–surprise) to better describe people's emotional context [32].

We focus on short films since they are motion pictures with all the components of a feature film but with less running time. As such, in only a few minutes, they express and elicit a whole gamut of emotions oriented to impact the audience and communicate a story, becoming an ideal test bed for our study of emotion association. We use a collection of short films that participated in two major festivals which leverage YouTube as their dissemination platform, namely *Tropfest* [5] and *Your Film Festival* [6].

In particular, our goal in this work is to address the following research questions:

RQ1. What is the role of emotions in short films?

RQ2. How can we leverage the wisdom of the crowd to detect emotions evoked by short films?

RQ3. How similar is the emotional context automatically associated to a short film compared to the one explicitly annotated by humans? and

RQ4. Are the automatically extracted emotions useful for personalized recommendations?

The association of emotions to films has several practical applications, for instance, to provide a personalized ranking of movies considering the emotions a person prefers to experience – e.g., by recommending *happy* short films to an audience who looks forward to experiencing *joy* – or for filmmakers to understand their audience and thus, provide more content with a higher degree of certain emotions to improve user experience. In summary, the main contributions of this paper are:

- We present an analysis on the responses elicited via crowdsourcing and on how the presence or absence of certain emotions may affect users decisions on what to watch.
- We propose a method for automatically detecting and extracting the different emotions and polarity evoked by short films.
- We show the usefulness of the extracted emotions in a context-aware film recommendation scenario.

2. RELATED WORK

We build upon the promising results of our previous work [31]. This paper differs from [31] by providing a more detailed and general overview: (i) We explore more extensively the crowdsourced association between short films and emotions, and share useful insights from this study in Section 3. (ii) We formally introduce our automatic approach for emotion extraction (AEX), offering clear details to ease its implementation and adoption for practitioners (Section 4). (iii) We also empirically compare the human emotion annotations against the automatically extracted affective context from user comments, using a context-aware recommender system as benchmark evaluation (Section 5).

The research we present in this work is mainly related to the areas of sentiment analysis, content ranking and recommendation exploiting user-generated posts, and context-aware recommendation.

Sentiment Analysis in the Social Web. Online Social Networks, with their growing popularity and availability, provide a constant stream of real users posts whose analysis sheds light on understanding human behavior, preferences, and diverse societal issues. Xu et al. [35] propose a fast training procedure to automatically recognize common emotions in bullying. The model is applied to Twitter posts and the findings reveal that the most common emotion in bullying traces is *fear*, followed by *sadness, anger*, and *relief*.

The detection of sentiments in short informal texts is described in [22] and in [27]. In [22] the authors present a system for sentiment detection based on a supervised statistical text classification approach and derive the sentiment features from tweet-specific sentiment lexicons. The work in [27] shows emotion-word hashtags as good labels for emotions in tweets and proposes the use of emotion-labeled tweets as a method to generate a word-emotion association lexicon.

In [17] the authors introduce a method for automatic emotion extraction from tweets and blog posts in Spanish – collected by tracking mentions of personal names of 18 Latin American presidents – in order to measure the emotional effect of each president over the public opinion.

In this paper we focus on exploiting social media comments to extract emotions related to short films accessible in YouTube. Our work goes one step beyond the aforementioned approaches as we present how the automatically extracted emotions can be useful for the task of personalized short film recommendation.

User-Generated Content for Ranking and Recommendation. The role of user-generated content for ranking and recommendation tasks is studied in diverse research scenarios. Zhang et al. [37] examine brands by incorporating users posts and information about users interactions. Their experiments on Facebook data show that negative comments generate greater awareness of a brand than positive comments. The usefulness of check-ins and text-based tips in improving location recommendation is explored in [36].

Lipczack et al. [24] study Flickr user actions and tags across network relations to gain understanding on ways to enhance similarity-based recommendation applications and in social network analysis. The authors in [20] analyze YouTube, Flickr, and Last.fm comments to predict the popularity of Web 2.0 items.

Our work explores YouTube comments to detect emotions evoked by short films. In contrast to recent works, we analyze the impact these emotions have in various contexts and leverage our findings for the task of emotion-aware recommendation.

Context-aware Recommendation. Context-aware recommender systems (CARS) [9, 10] improve recommendation quality by incorporating contextual information to the recommendation process – e.g., time [18], location [26], social ties [16], or mood [34] – which enables them to adapt to the specific user's situation. In this work we are particularly interested in emotional context and its impact in CARS.

Gonzalez et al. [19] introduce a system that embeds users' emotional information, which is captured incrementally via small surveys, to enrich recommendations. More recently, Odic et al. [29, 30] demonstrate that emotions can be influential contextual variables in making recommendations. The role of emotions in CARS is explored by Zheng and coauthors in [38] where they study the influence of emotional context for rating prediction.

Similar to these studies, which rely on surveys to extract the emotional context with respect to the items of interest (e.g., movies), we ask humans to associate emotional vectors to short films. However, we go further and also explore how the emotional context can be automatically extracted from social feedback (YouTube comments) and then use a CARS setting as benchmark to assess its usefulness.

3. ROLE OF EMOTIONS IN SHORT FILMS

Our goal is to create a collection of short film-emotion associations to better understand the role of affective context in short films (RQ1). In addition, the emotional context extracted by humans will provide a gold-standard[1] to evaluate our automatic approach for emotion extraction, which we introduce in Section 4. To this end, we use Amazon Mechanical Turk (AMT) [3], a crowdsourcing marketplace where businesses and individuals (known as *requesters*) have access to an on-demand, scalable workforce (*workers*). The

[1]Our dataset is available upon request to the first author.

requesters post jobs, known as Human Intelligence Tasks (*HITs*), and the workers complete them in exchange for a monetary reward.

We design a HIT per short film. The HITs are designed following the structure used in [28] but slightly modified in order to adjust them to the films. Each HIT involves watching one single short film and answering sixteen questions. Eight of these questions ask the user to rate the degree of association of the short film to each emotion presented in Plutchik's psychoevolutionary theory [32]. This theory considers that there are eight primary emotions forming four opposing pairs, joy–sadness, anger–fear, trust–disgust, and anticipation–surprise. The next two questions ask for the sentiment polarity that the user would associate to the film (i.e., positive or negative). Five questions aim to get additional context relevant data for the user-short film pair such as time, audience, companion, emoticon, and genre. Finally, one question asks the user to label the film with a *like* or a *dislike* according to her/his personal preference. Appendix A shows an example of a HIT.

Short film collection. A short film is a motion picture with running time of 40 minutes or less [2]. As such, it contains all the elements of a feature film with the difference that the entire plot is much more condensed: there are only a few minutes to tell the story, transmit emotions, and impact the audience.

As a video-sharing website, YouTube hosts content ranging from a funny episode in a family trip to full length movies and documentaries, and unless private, each one of the videos can receive views, ratings and comments from audiences worldwide. Due to its popularity, accessibility and reach, short films festivals such as *Tropfest* [5] and *Your Film Festival* [6] use YouTube to host the films submitted to the competitions.

Tropfest is the world's largest short film festival, it started in Australia twenty two years ago and has expanded to regions including South East Asia and North America [1]. The finalists and winner films of *Tropfest* are uploaded to the festival's YouTube channel which has more than 68K subscribers and 35M views. *Your Film Festival*, sponsored directly by YouTube, took place in 2012 and it offered a \$500K grant as the prize. Over 15K films were uploaded to YouTube to enter the competition and the 10 finalists were selected by YouTube users.

Focusing on the two aforementioned festivals, we used YouTube's Data API [7] to collect a set of the participant short films. For each of them, we collected its metadata – i.e., number of views, likes, dislikes – and corresponding comments. The final collection consists of 235 short films and a total of 21,043 comments. The minimum and maximum number of comments per short film is 3 and 996, respectively. The average is 91 and the median is 27 comments per video.

Participants. Each user decided which HIT to annotate or skip. In total we had 107 different participants of which 27 were discarded after a quality control screening. The annotations of the remaining 80 participants constitute the input of this study. The users annotated 7 HITs on average and each HIT was completed by approximately 3 users. In total we captured 631 user-item interactions for the 235 short films.

Emotional vectors from the crowdsourced annotations. To extract emotional vectors from the crowdsourced annotations we considered the questions related to the emotions and polarity and associated a numerical value to each response. Table 1 shows the possible responses and the associated numerical values. Using the numerical scores, we normalized the responses to obtain probability vectors, i.e., whose values add up to 1.

Response	Value
This short film is not associated with the emotion e	0
This short film is weakly associated with the emotion e	1
This short film is moderately associated with the emotion e	2
This short film is strongly associated with the emotion e	3

Table 1: **Mapping responses to numerical values. In each question, e denotes the corresponding emotion.**

For example, an emotional vector for a short film over dimensions [joy, sadness, anger, fear, trust, disgust, anticipation, surprise], can correspond to: [0.22, 0.15, 0.07, 0.08, 0.10, 0.06, 0.17, 0.15]. The components of the vector add up to 1 and each of them is a positive number between 0 and 1. Similarly, a polarity tuple (positive, negative) can correspond to: (0.6, 0.4).

3.1 Emotions and Contexts

Contextual information is important if one wants to better understand the users' needs and provide them with higher quality ranking or recommendations. On each HIT, besides the emotions and polarity questions, we asked the user to label the film according to: (i) the audience(s) she perceived as being the most appropriate for the film, (ii) the companion(s) she would choose to watch the film with, (iii) the time when she would watch the film, (iv) the emoticon(s) that would be more representative of her experience while watching the film, and (v) the genre(s) she thinks the film belongs to. We also asked the user to indicate if she liked or disliked the film. Note that we did not ask the users to associate emotions to these different contexts.

The purpose of these questions is to explore if and how emotions, implicitly, affect the decision on what to watch.

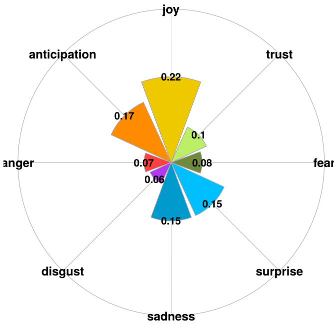
In what follows we analyze the emotions present in the videos associated by the users to each context.

Likes and Dislikes

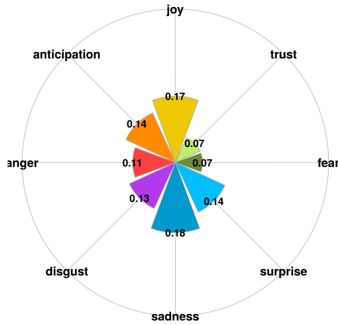
We obtained 631 crowdsourced responses (e.g., user-item interactions). In 72.9% (460) of these cases the users indicated they liked the short film while in 27.1% (171) they disliked it.

Since each user also indicated what emotions and polarity she associates to the video, we explore how these emotions are distributed for both the *liked* and *disliked* films. For example, for all the *liked* videos we calculate the overall joy score by adding all the scores given for the emotion joy and dividing it by the number of videos (average joy). We follow the same procedure for all the emotions.

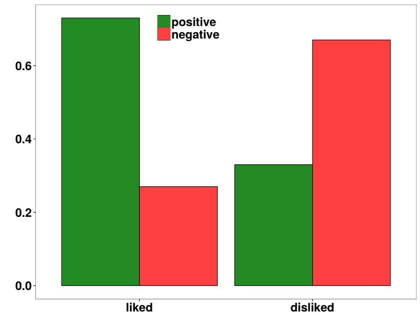
Figure 1 shows the average distribution of emotions and polarity in *liked* and *disliked* videos. As we can see in Figure 1(a), the most prominent emotion in the *liked* videos is joy, followed by anticipation, surprise and sadness. In the *disliked* videos the distribution of emotions is slightly different, sadness is the most prominent, followed by joy, anticipation and surprise (Figure 1(b)). It is also worth noting that the emotions disgust and anger increase when compared with those in the *liked* videos.



(a) Liked videos.



(b) Disliked videos.



(c) Polarity.

Figure 1: **Emotions and polarity in *liked* and *disliked* videos.**

Besides our questions related to Plutchik's eight basic emotions, we also asked users to indicate how positive or negative a film was. The possible responses were similar to those shown in Table 1.

In all cases where users labeled the video with a *like* we find an average score of 0.73 (the max. possible score is 1) for positive polarity and a score of 0.27 for negative polarity, while in the videos labeled with a *dislike*, there is an average score of 0.33 and 0.67 for positive and negative polarity, respectively. Figure 1(c) shows the distribution of polarity in the *liked* and *disliked* videos.

As shown in Figure 1, *Liked* videos are more associated with joy, anticipation, surprise, sadness, trust, and positive polarity, while *disliked* videos are more associated with sadness, joy, anticipation, surprise, a higher degree of disgust and anger, and negative polarity.

Audience

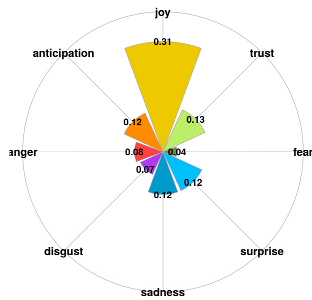
On each HIT, we asked the user to select the audience she/he thought as being the most adequate for the given short film. The question had six possible answers: *children, teenagers, young adults, adults, seniors*, and *all audiences*.

One reason behind this question is to better understand if there are certain emotions which a user considers more appropriate for a given demographic. Another reason is to explore how users preferences would change according to the context, for example, parents may forbid their child to watch a film if it is known to contain a high degree of anger.

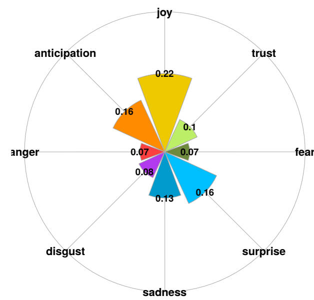
As in this question the users were allowed to select more than one answer, for the analysis we considered each film as many times as audiences were thought as appropriate. 5% of the films (56 films) were marked as appropriate for *children*, 16% (184) for *teenagers*, 28% (322) for *young adults*, 29% (337) for *adults*, 10% (112) for *seniors*, and 12% (135) for *all audiences*.

Figure 2 shows the average distribution of the emotions associated to the films that the users selected as adequate for each audience. It is interesting to note that for the audiences *children*, *teenagers* and *all audiences* the emotion with higher average is joy; however, as the audience's age increases (*young adults*, *adults*, *seniors*) the appropriate films contain less joy and a higher presence of sadness, disgust and fear, according to the respondents.

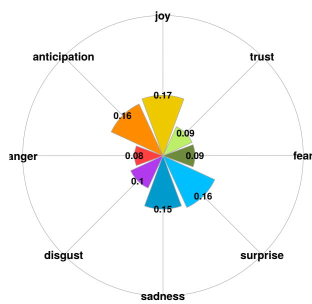
Children (Figure 2(a)) and *seniors* (Figure 2(e)) are the audiences for which the users selected the lowest number of videos when compared to the rest of the groups. The films for *children* have a high score for joy (0.31) and a low score for fear (0.04), disgust (0.07) and anger (0.08), while the films for *seniors* have the highest sadness score (0.21) and the lowest joy score (0.14) among all the audiences.



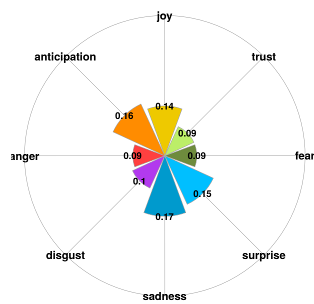
(a) Children.



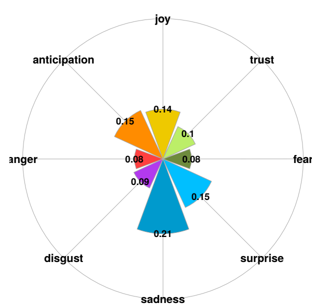
(b) Teenagers.



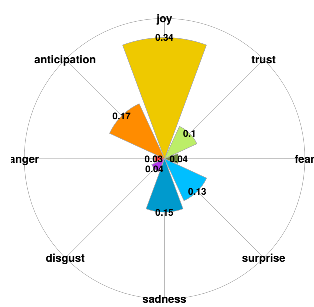
(c) Young adults.



(d) Adults.



(e) Seniors.



(f) All audiences.

Figure 2: **Emotions per audience.**

The short films marked as appropriate for *all audiences* (Figure 2(f)) are those associated with joy (0.34), anticipation (0.17), sadness (0.15), surprise (0.13), and trust (0.10). The low scores for disgust (0.04), fear (0.04) and anger (0.03) could indicate the influence of these particular emotions on the selection of an adequate audience.

The highest number of associated videos are for the audiences *young adults* (Figure 2(c)) and *adults* (Figure 2(d)). Contrary to other audiences, the videos in these groups show a more balanced distribution of emotions (scores ranging from 0.08 to 0.17), and while disgust, anger and fear are not dominant, there is an increase with respect to the other groups.

Time

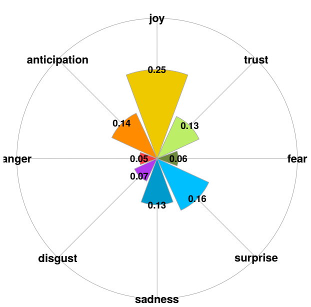
Time is one of the most studied contexts when trying to determine meaningful items for users. Users preferences vary according to the time of the day, day of the week, or weekday and weekend.

We asked users to indicate when they would prefer watching the given film. The question had four possible answers: *to relax after work*, *during a break at work*, *for entertainment during weekends or on vacation*, and *at anytime*. More than one answer was permitted.

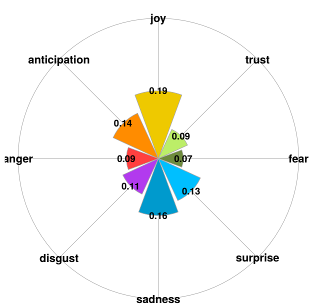
Rather than specifying a time of the day (i.e., morning, afternoon) we consider a combination of activity and time.

For *relaxing after work* 13% (89) of the films would be watched, 17% (115) during a *break at work*, 28% (195) *for entertainment during weekends or on vacation*, and 42% (292) at *anytime*. Figure 3 shows the average distribution of emotions for each time.

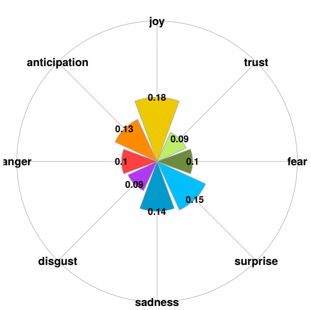
Joy, surprise, and anticipation are the strongest emotions in the videos selected for *relaxing after work*, with anger, fear and disgust being the least present (Figure 3(a)). *During a break at work* (Figure 3(b)) and *for entertainment on weekends or vacations* (Figure 3(c)) the films show an increase in anger, disgust, sadness, and fear. For the videos that would be watched at *anytime* (Figure 3(d)) the dominant emotions are joy, anticipation, sadness, and surprise.



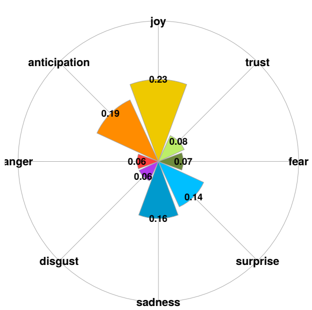
(a) After work.



(b) A break at work.



(c) Weekends or vacations.



(d) Anytime.

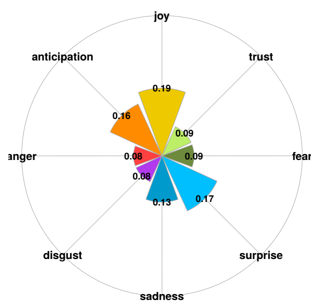
Figure 3: **Emotions per time.**

Companion

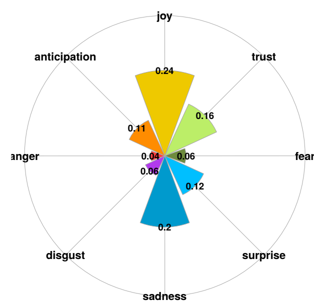
One of the factors that influences users' decisions on what to watch is their companion at that specific moment. For example, a teenager might not watch a war film if she is with family, but would be keen on watching it when she is with friends or alone, or a person would not dare to watch a horror movie while alone, but would watch it when surrounded by friends.

We asked the users to indicate the companion with whom they would watch the short film presented in the HIT. The possible answers were: *friends, family, partner, alone,* and *anybody.* More than one answer was permitted.

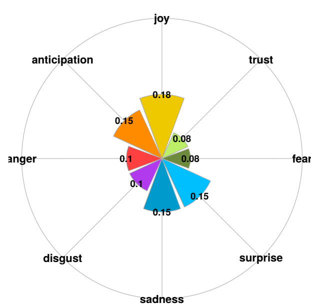
Figure 4 shows the average distribution of emotions per companion. 7% (59) and 12% (94) of the films would be watched with *family* and *partner* (e.g., boyfriend/girlfriend,



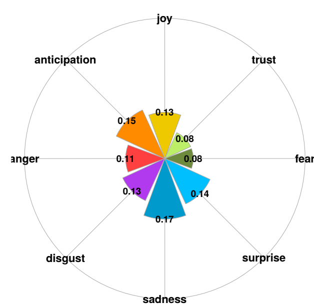
(a) Friends.



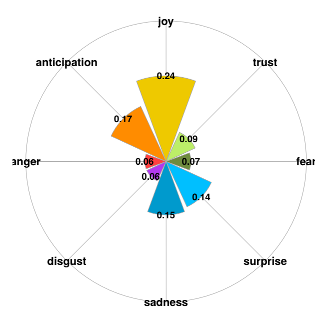
(b) Family.



(c) Partner.



(d) Alone.



(e) Anybody.

Figure 4: **Emotions per companion.**

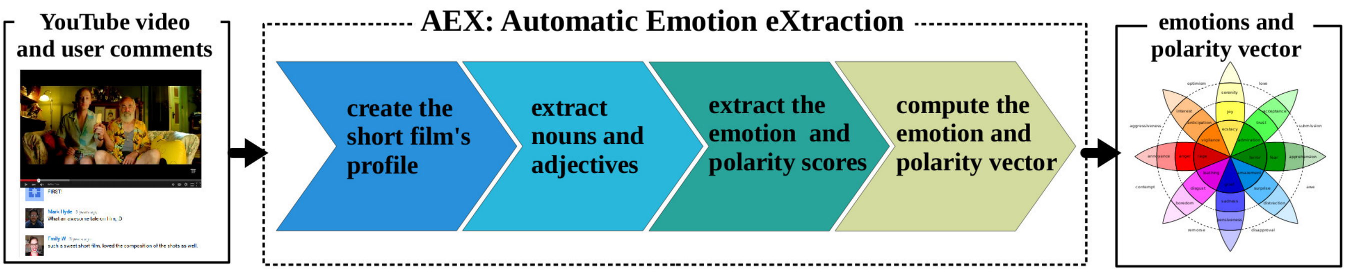


Figure 5: **Extraction of emotions and polarity vectors from YouTube Comments. (1) Aggregate the short film's comments to form the film's profile. (2) Tag the profile with part-of-speech and extract the identified nouns and adjectives. (3) Search a match in EmoLex for each noun and adjective. When there is a match, add and process the related scores to form the emotions and polarity vector.**

husband/wife), respectively, 20% (155) with *friends*, 23% (187) *alone*, and 38% (303) with anybody.

The films that the users would watch with *friends* (Figure 4(a)) have a high association with the emotions joy, surprise, anticipation, and sadness, and the highest presence of fear among all the companions. With *family* (Figure 4(b)), joy, sadness and trust are the dominant emotions, and fear has a lower score than in any other group of companions. For the option *alone* (Figure 4(d)), users selected those films with the lowest average of joy but with the highest average of disgust and anger.

With *partner* (Figure 4(c)), the average distribution of emotions is similar to that shown in (Figure 4(d)) with the main differences being joy, sadness, and disgust scores. The films to watch with *anybody* have joy as their dominant emotion, followed by anticipation and sadness. Anger, fear, and disgust have the lowest average scores among the emotions in this category.

In summary, under the settings of this study:

- Emotions impact the decisions on what to watch given different contexts.
- The most popular emotions (in terms of average score) found across the different contexts are joy, sadness, surprise, and anticipation. The least popular emotions and possibly the most discriminative – a strong presence of one or more of these emotions may help users distinguish in which contexts to watch the film – were anger, disgust, trust, and fear.
- Contexts involving children, seniors, or family resulted in a more unbalanced (very different average scores) distribution of emotions in the videos considered as appropriate, while for audiences such as adults, the selected videos had a more balanced (more similar average scores) presence of emotions.

4. AUTOMATIC EMOTION EXTRACTION

The explicit associations between short films and emotions via crowdsourcing provide very useful insights as detailed in Section 3. However, such effort can be costly and time consuming, particularly if done at large scale. In addition to that and given the accelerated pace of film production, an automatic approach for emotion detection represents a more desirable alternative.

In this section we address RQ2 and present our Automatic Emotion eXtraction approach (AEX) that exploits the large amount of YouTube comments associated to a short film – as an important source of opinions and discussions about the video – in order to track the emotions it evokes.

AEX is comprised of the steps outlined in Procedure 1 and illustrated in Figure 5. Our approach is motivated by promising results reported in recent studies for emotion detection from short and sparse text [17, 31, 22].

Procedure 1 AEX: Automatic Emotion eXtraction

Input: User comments associated to a short film p.
Output: Emotional vector e_p and polarity tuple τ_p for the short film.

1: Create a *profile* for short film p by aggregating its user comments.
2: Extract the terms from the profile – i.e., nouns and adjectives.
3: Automatically associate each extracted term to a set of *emotions* and *polarity*.
4: Compute the *emotion vector* e_p and *polarity tuple* τ_p for the short film p.
5: **return** $[e_p, \tau_p]$

First we build a *profile* for each of the short films. The profile consists of all the user comments collected for the corresponding short film.

After building the profiles we perform part-of-speech tagging on each of them, using LingPipe [8] and MorphAdorner [4], in order to extract the *nouns* and *adjectives*, which have been shown to be good predictors for this task [17]. Part-of-speech tagging is necessary because the same word can act as a different part-of-speech depending on the context of the sentence.

Then, using a term-based matching technique, we associate each term with emotion and polarity values. To this end, we use in our study the NRC Emotion Lexicon (EmoLex) [28], a large set of human-provided word-emotion association ratings annotated according to Plutchik's psychoevolutionary theory [32]. Besides including annotations for each primary emotion, namely joy–sadness, anger–fear, trust–disgust, and anticipation–surprise, EmoLex also includes positive and negative sentiments associated to the words.

For instance, the word *friend* is associated to the emotion *joy* and to *positive* polarity, whereas the word *violent* is associated to the emotions of *anger, disgust, fear, surprise*, and *trust*, and a *negative* polarity.

We use EmoLex in order to be consistent with the affective context associated by humans in Section 3 and we also use the same data set – 235 short films and a total of 21,043 comments. The minimum and maximum number of comments per short film is 3 and 996, respectively. The average is 91 and the median is 27 comments per video.

AEX's goal is to automatically obtain an emotional vector and polarity tuple with the same structure as the ones elicited via crowdsourcing (see Section 3). To this end, we define the emotional vector, e_p, for short film p as follows. Let T_p be the set of terms extracted from the short film's profile p, and T_m the set of all terms in EmoLex annotated with emotion m, where $m \in M$; $M := \{$*joy, sadness, anger, fear, trust, disgust, anticipation, surprise*$\}$, i.e., Plutchik's eight basic emotions. Then, the m^{th} dimension of emotional vector $e_p \in \mathbb{R}^{|M|}$ is given by:

$$e_p[m] := \sum_{t \in T_p} \mathcal{I}_m(t) * \text{tf-idf}(t)$$

where $\mathcal{I}_m(t)$ is an indicator function that outputs 1 if the term $t \in T_p$ is associated to emotion m, and 0 otherwise. tf-idf(t) denotes the tf-idf score of the term t, which weights the importance of the term contribution to the emotional vector.

Finally, we normalize vector e_p to produce a stochastic vector $\hat{e_p} = \frac{e_p}{Z_e}$, where $Z_e = ||e_p||_1$ is a normalization constant.

Similarly as in the case of emotions, we compute the polarity tuple $\tau_p = ($*positive, negative*$)$ of short film p as follows:

$$\tau_p = \frac{1}{Z_{polarity}} \Big(\sum_{t \in T_p} \mathcal{I}_+(t) * \text{tf-idf(t)}, \sum_{t \in T_p} \mathcal{I}_-(t) * \text{tf-idf(t)} \Big)$$

where the indicator functions and normalization constant are defined analogously as in the case of the emotions.

AEX offers the possibility to associate emotions to videos based on the social comments and discussion about it. Note that while this work focuses on short films, AEX is applicable to other scenarios using other types of textual resources, such as tweets, blog posts or users product reviews. In the next section we quantify to what extent the emotions extracted automatically compare to the ones associated by humans using a recommendation task as a benchmark.

5. EMPIRICAL EVALUATION

In this section we address RQ3 and measure to what extent the emotional context extracted using AEX compares to the one explicitly assigned by humans. We also assess the usefulness of the extracted context for a recommendation task, which tackles RQ4.

5.1 User Self-Similarity

AEX should capture the same, or a very similar, emotional context as the one extracted from the explicit human annotations. We take a user-centric perspective in order to measure the similarity among both strategies for emotion extraction.

We represent each observed user-item pair interaction – e.g., user watches a short film – using two emotional vectors corresponding to the manual and automatic strategy, respectively. Then, we compute the Cosine Similarity (CS) and Pearson Correlation Coefficient (PCC) between these two vectors and average the corresponding scores user-wise in order to measure the self-similarity of the user across the approaches.

The average CS and PCC over all users is $CS = 0.6893$ ($\sigma = 0.166$) and $PCC = 0.4992$ ($p < 0.001$) respectively, which indicates a positive correlation between the strategies.

5.2 Emotion-aware Recommendation

Our goal in this part of the evaluation is to answer RQ4, i.e., we want to assess how useful is the emotional context extracted by AEX for a recommendation task.

Emotions and CARS. A natural way to include emotional information into a recommender system is by following a context-aware approach. Context-aware recommender systems (CARS) [10] generate more relevant recommendations by (i) incorporating contextual information in the recommendation process and by (ii) adapting the recommendations to the specific contextual situation of the user. In this evaluation we are concerned with the former and leave the latter for future work, e.g., leveraging the different dimensions explored in Section 3 for adaptation.

We incorporate the emotional context into the recommender system by following a Collaborative Ranking (CR) approach [12, 16] – other state-of-the-art alternatives for CARS include Factorization Machines [33] and N-dimensional Tensor Factorization [21]. The main idea behind this approach is to cast the recommendation problem into a (personalized) learning-to-rank task [25].

The emotion-aware recommendation problem can be placed into the learning-to-rank framework [25] by noting that the users correspond to queries and short films to documents. For each user-item pair the observed binary rating (like/dislike) indicates the relevance of the corresponding item (short film) with respect to that user and can be used to optimize the parameters of a ranking function that will be used for recommendation. As in the case of the self-similarity computation, we represent each observed user-item pair interaction by the corresponding emotional vectors given by the manual and automatic strategy (AEX). Figure 6 shows an example of a standard CARS-CR data representation.

```
-----------------------------------------------------------------
 1 qid:1 1:0.05 2:0.17 3:0.04 4:0.06 5:0.27 6:0.06 7:0.11 8:0.22
-1 qid:1 1:0.12 2:0.12 3:0.16 4:0.06 5:0.16 6:0.06 7:0.16 8:0.12
-1 qid:1 1:0.00 2:0.20 3:0.03 4:0.06 5:0.23 6:0.00 7:0.07 8:0.38
 1 qid:2 1:0.11 2:0.12 3:0.12 4:0.10 5:0.15 6:0.09 7:0.07 8:0.20
-1 qid:2 1:0.04 2:0.23 3:0.03 4:0.02 5:0.22 6:0.01 7:0.14 8:0.26
-----------------------------------------------------------------
```

Figure 6: **Collaborative Ranking data representation for CARS. In each row the first column represents the label, the second represents the user id (qid), and the rest of the columns are the emotional features of the user-item pair interactions, which correspond to the 8-dimension emotional vector (the 2 dimensions representing the polarity tuple are omitted in this example).**

CARS-CR learning-to-rank method. We use LambdaMART as the learning-to-rank method [13] for our emotion-aware recommendation task. We choose this method due to its ability to directly optimize Information Retrieval (IR) metrics and for the good performance exhibited in recent ranking competitions [14, 15].

LambdaMART is a learning-to-rank (ensemble) method that uses Multiple Additive Regression Trees or MART as base learners. One of the main features of LambdaMART is that it can optimize arbitrary IR metrics by guiding the learning process using so-called λ-gradients, which reflect small changes in the IR metric while iterating over the training set. We omit a detailed description of the model, since it is out of the scope of this report, and refer the reader to [13] and [14] for a comprehensive analysis of LambdaMART.

Protocol. The dataset is split user-wise into training, validation, and test set. First, 20% of the users – with their respective items – are randomly assigned to the test set. Then, we randomly sample 10% of the remaining users to form a validation set, which will help us to select the hyperparameters of our model. Finally, the rest of the user-item interactions are used for training.

After the selection of the best hyperparameters through grid-search, we retrain our models on the union of the training and validation sets, i.e., using 80% of the users. To account for variability, the results reported are the average of 100 rounds of experiments considering 95% confidence intervals.

Metrics. In order to measure the recommendation quality and to reflect our goal of recommending a short list of items to a user, we look to *precision* and *mean reciprocal rank* at cut-off level 5 – i.e., $P@5$ and $MRR@5$ [11]. The metrics are computed for each user in the test set and then averaged.

Parameters settings. Using the validation set we found that for LambdaMART a number of 64 base trees, with 5 leaves each, and a learning rate (or shrinkage) equal to 0.1 led to good results. We use RankLib's implementation of LambdaMART to learn the ranking function [23].

Results. Figure 7 shows the recommendation performance for the CARS-CR model trained using the emotional vectors derived from human intelligence (HI) and the one trained using the affective context automatically extracted with AEX. The P@5 for the CARS-CR models using HI and AEX are 75.95% and 69.50%, respectively, which corresponds to a difference of 6.45 percentage points.

CARS-CR using HI achieves a $MRR@5 = 83.43\%$ which ties with the AEX-based approach ($MRR@5 = 81.83\%$) since the difference is not statistically significant.

5.3 Discussion

After our experimental evaluation, we can answer our research questions RQ3 and RQ4.

The emotional context automatically associated to a short film is similar and positively correlated to the one explicitly annotated by humans ($CS = 68.93\%$ and $PCC = 49.92\%$). This level of user self-similarity between the emotional contexts is important for recommender systems, for example, user- or item-based collaborative filtering (CF), which use a k-nearest-neighbor approach to deliver recommendations, could use the AEX emotional vectors – instead of the human annotated ones – to compute the user or item neighborhoods required to produce a short list of personalized recommendations. This answers RQ3.

For RQ4, the results illustrated in Figure 7 show that the recommender system based on AEX is very competitive, specially in terms of $MRR@5$. This indicates that the automatic approach is capable of inferring an affective context useful in learning an emotion-aware personalized ranking model. Given the inherently human nature of emotions, the fact that the HI-based approach achieves a slightly better performance is somewhat expected; however, the promising results of the AEX-based model open the doors to compute

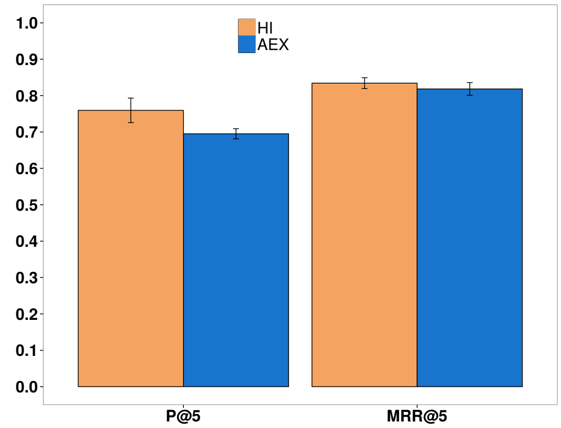


Figure 7: **Recommendation performance for the CARS-CR model using human intelligence (HI) and automatically extracted affective context (AEX).**

emotional context at scale and then leverage it to improve recommender systems.

6. CONCLUSION

In this paper we tackled the issue of extracting emotions associated to short films. We approached the task from two different fronts: (i) using crowdsourcing to obtain emotional labels and (ii) automatically extracting affective context based on the users criticism expressed in YouTube comments. Furthermore, we explored the usefulness of the extracted emotions for the task of context-aware recommendation.

We crowdsourced emotional labels by asking people on Amazon Mechanical Turk to complete HITs of sixteen questions, which aimed at identifying the emotions and polarity they associated with the short film and other relevant information on the different contexts (i.e., time, audience) at the moment of making the decision on whether to watch a film or not. The results obtained reveal that the emotions and polarity associated to short films have an impact on users liking/disliking the films and on the circumstances they would consider appropriate for watching a film or not.

We proposed AEX, an automatic emotion extraction approach, which makes use of YouTube comments to identify emotions associated to short films. Our findings reveal that users' comments indeed provide the information required to automatically extract emotions associated to short films.

Lastly, we measured the similarity of the emotional context automatically associated to a short film to that explicitly annotated by humans. In addition, we assessed the usefulness of the emotional context extracted by AEX for a recommendation task. Our results reveal that the emotional contexts extracted by the two different approaches are similar, and that the automatic approach is capable of inferring a useful affective context for emotion-aware personalized ranking.

The dataset used in this work was collected from two popular short film festivals available in YouTube. One of the reasons being reducing the inclusion of spam and ensuring a better quality of the selected films. We are aware, however, that the size of our final dataset is one limitation of this work. In the future, we plan to enlarge and diversify our dataset and to address tasks of short films ranking and recommendation considering, along with emotions and polarity, other contextual information such as time, companion, or target audience.

Acknowledgements. This work was supported in part by Science Foundation Ireland - Grant Number: 12/RC/2289.

7. REFERENCES

[1] About Tropfest. `http://tropfest.com/about/`. Accessed: 2015-04.

[2] Academy of Motion Picture Arts (AMPAS). `http://www.oscars.org/oscars/rules-eligibility`. Accessed: 2015-03.

[3] Amazon Mechanical Turk. `https://www.mturk.com/mturk/welcome`. Accessed: 2015-04.

[4] MorphAdorner. `http://morphadorner.northwestern.edu`. Accessed: 2015-03.

[5] Tropfest YouTube Channel. `https://www.youtube.com/user/tropfest`. Accessed: 2015-03.

[6] Your Film Festival. `https://www.youtube.com/user/yourfilmfestival`. Accessed: 2015-03.

[7] YouTube API. `http://developers.google.com/youtube/v3/`. Accessed: 2015-04.

[8] Alias-i. LingPipe 4.1.0. `http://alias-i.com/lingpipe`, 2008. Accessed: 2015-03.

[9] G. Adomavicius, B. Mobasher, F. Ricci, and A. Tuzhilin. Context-aware recommender systems. *AI Magazine*, 32(3), 2011.

[10] G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*, pages 217–253. Springer US, 2011.

[11] R. A. Baeza-Yates and B. A. Ribeiro-Neto. *Modern Information Retrieval - the concepts and technology behind search, Second edition*. Pearson Education Ltd., Harlow, England, 2011.

[12] S. Balakrishnan and S. Chopra. Collaborative ranking. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 143–152, New York, NY, USA, 2012. ACM.

[13] C. J. Burges. From ranknet to lambdarank to lambdamart: An overview. Technical Report MSR-TR-2010-82, Microsoft Research, June 2010.

[14] C. J. C. Burges, K. M. Svore, P. N. Bennett, A. Pastusiak, and Q. Wu. Learning to rank using an ensemble of lambda-gradient models. In *Yahoo! Learning to Rank Challenge*, pages 25–35, 2011.

[15] O. Chapelle and Y. Chang. Yahoo! learning to rank challenge overview. In *Yahoo! Learning to Rank Challenge*, pages 1–24, 2011.

[16] E. Diaz-Aviles, H. T. Lam, F. Pinelli, S. Braghin, Y. Gkoufas, M. Berlingerio, and F. Calabrese. Predicting user engagement in twitter with collaborative ranking. In *Proceedings of the 2014 Recommender Systems Challenge*, RecSysChallenge'14, 2014.

[17] E. Diaz-Aviles, C. Orellana-Rodriguez, and W. Nejdl. Taking the pulse of political emotions in latin america based on social web streams. *Web Congress, Latin American*, 2012.

[18] Z. Gantner, S. Rendle, and L. Schmidt-Thieme. Factorization models for context-/time-aware movie recommendations. In *Proceedings of the Workshop on Context-Aware Movie Recommendation*, CAMRa '10, 2010.

[19] G. Gonzalez, J. de la Rosa, M. Montaner, and S. Delfin. Embedding emotional context in recommender systems. In *Data Engineering Workshop, 2007 IEEE 23rd International Conference on*, 2007.

[20] X. He, M. Gao, M.-Y. Kan, Y. Liu, and K. Sugiyama. Predicting the popularity of web 2.0 items based on user comments. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '14, pages 233–242, New York, NY, USA, 2014. ACM.

[21] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver. Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys '10, pages 79–86, New York, NY, USA, 2010. ACM.

[22] S. Kiritchenko, X. Zhu, and S. M. Mohammad. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research (JAIR)*, 2014.

[23] Lemur Project. Ranklib. `http://www.lemurproject.org/`, 2015. Accessed: 2015-03.

[24] M. Lipczak, B. Sigurbjornsson, and A. Jaimes. Understanding and leveraging tag-based relations in on-line social networks. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*, HT '12, 2012.

[25] T.-Y. Liu. *Learning to Rank for Information Retrieval*. Springer, 2011.

[26] C. Mettouris and G. A. Papadopoulos. Ubiquitous recommender systems. *Computing*, 96(3):223–257, 2014.

[27] S. M. Mohammad and S. Kiritchenko. Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, 2014.

[28] S. M. Mohammad and P. D. Turney. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 2011.

[29] A. Odić, M. Tkalčič, A. Košir, and J. F. Tasič. Relevant context in a movie recommender system: Users' opinion vs. statistical detection. In *Proceedings of the 4th Workshop on Context-Aware Recommender Systems*, 2012.

[30] A. Odić, M. Tkalčič, J. F. Tasič, and A. Košir. Predicting and detecting the relevant contextual information in a movie-recommender system. *Interacting with Computers*, 25(1), 2013.

[31] C. Orellana-Rodriguez, E. Diaz-Aviles, and W. Nejdl. Mining emotions in short films: User comments or crowdsourcing? In *Proceedings of the 22nd*

International Conference on World Wide Web Companion (Posters)*, WWW '13, 2013.

[32] R. Plutchik. *A General Psychoevolutionary Theory of Emotion*. Academic press, New York, 1980.

[33] S. Rendle, Z. Gantner, C. Freudenthaler, and L. Schmidt-Thieme. Fast context-aware recommendations with factorization machines. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11. ACM, 2011.

[34] Y. Shi, M. Larson, and A. Hanjalic. Mining mood-specific movie similarity with matrix factorization for context-aware recommendation. In *Proceedings of the Workshop on Context-Aware Movie Recommendation*, CAMRa '10, 2010.

[35] J.-M. Xu, X. Zhu, and A. Bellmore. Fast learning for sentiment analysis on bullying. In *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, WISDOM '12, 2012.

[36] D. Yang, D. Zhang, Z. Yu, and Z. Wang. A sentiment-enhanced personalized location recommendation system. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, HT '13, 2013.

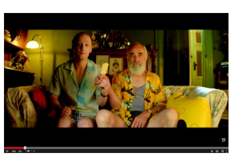
[37] K. Zhang, S. Bhattacharyya, and S. Ram. Empirical analysis of implicit brand networks on social media. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, 2014.

[38] Y. Zheng, R. Burke, and B. Mobasher. The role of emotions in context-aware recommendation. In *ACM RecSys' 13, Proceedings of the 3rd International Workshop on Human Decision Making in Recommender Systems*, 2013.

APPENDIX

A. EXAMPLE OF A HIT

Title: Emotions in Short Films
Keywords: Emotions, Sentiment Analysis, Video
Reward : $:0.15
Description: You will be shown a short film and asked to indicate the different emotions you associate to the film



Q1. Did you like this short film?
- yes
- no

Q2. How positive (good, praising) is this short film?
- this short film is not positive
- this short film is weakly positive
- this short film is moderately positive
- this short film is strongly positive

Q3. How negative (bad, criticizing) is this short film?
- similar choices as in Q2

Q4. How much would you associate this short film with the emotion joy? (for example, *happy* and *funny* scenes are strongly associated with joy)
- this short film is not associated with joy
- this short film is weakly associated with joy
- this short film is moderately associated with joy
- this short film is strongly associated with joy

Q5. How much would you associate this short film with the emotion sadness? (for example, *gloomy* and *hearthbreaking* scenes are strongly associated with sadness)
- up to Q11 the answers are similar to those in Q4

Q6. How much would you associate this short film with the emotion fear? (for example, *horror* and *scary* scenes are strongly associated with fear)

Q7. How much would you associate this short film with the emotion anger? (for example, *rage* and *shouting* scenes are strongly associated with anger)

Q8. How much would you associate this short film with the emotion trust? (for example, *loyalty* and *integrity* scenes are strongly associated with trust)

Q9. How much would you associate this short film with the emotion disgust? (for example, *gross* and *cruelty* scenes are strongly associated with disgust)

Q10. How much would you associate this short film with the emotion surprise? (for example, *astonishing* and *sudden* scenes are strongly associated with surprise)

Q11. How much would you associate this short film with the emotion anticipation? (for example, scenes that keep you *interested* and *expecting* something are strongly associated with anticipation)

Q12. Who would you watch this short film with?
- friends
- family
- boyfriend/girlfriend (husband/wife)
- alone
- with anybody

Q13. When would you watch this short film?
- for relaxing after a working day
- during a break at work
- for entertainment on weekends or on vacation
- at anytime

Q14. Who do you think would be the most appropriate audience(s) of this short film?
- children (0-12 years)
- teenagers (13-17 years)
- young adults (18-24 years)
- adults (25-64 years)
- seniors (> 65 years)
- all audiences

Q15. Please select the emoticons (one or more) representing the emotions you experienced while watching this short film

:) :(:D :o :& :@
|-) ;(:| (whew) :S :^)

Q16. What would you say is the genre (or genres) of this short film?

action	adventure	biography	animation
comedy	crime	documentary	drama
family	fantasy	film-noir	game-show
history	horror	music	musical
mystery	news	reality-tv	romance
sci-fi	sport	talk-show	thriller
war	western		