Abstract

Finding a good movie to watch is getting more difficult in this day and age because of the huge availability. Movie recommendations are thus getting more important. Mood-based and context-aware recommenders could help improve these by taking a viewer’s mood and context into account. In this paper, we constructed affective ‘profiles’ for movies that potentially characterize their affective influence on viewer’s moods. We then gathered data about people’s moods before and after watching movies to use these affective profiles in predicting viewer’s moods after the movie. We found that a linear model armed with these affective profiles and information about people’s mood and an estimated score for the movie could explain 22% of the variance in moods. While not perfect, we believe this shows that affective movie profiles based on metadata combined with traditional recommender systems could successfully use people’s moods to improve movie recommendations.

1 Introduction

Let’s say that you come home one day after a tough day at work and you want to watch a decent movie with your partner to release some of the tension in your body. You turn on your TV and start browsing through the enormous database of movies that your cable company’s subscription provides for online streaming. After having read summaries of a dozen movies, you’re getting irritated as none of them really appeal to you at this time. Some of them you’re quite confident you’d like, but you simply are not in the mood for them because they’re too “heavy” or too long. So finally, you give up and decide to watch a random TV show. What I just described is a situation that may occur more and more often in people’s lives as our access to instant realtime streaming services gets better and the libraries of available material grow.

1.1 Dealing with choice

How do you, a person wanting to watch a movie, decide what to watch when retrieval is instantaneous, and the number of possibilities endless? In 2005, Barry Schwartz gave a TED talk about the matter called ‘The paradox of choice’, in which he describes how people on one hand want more choice, but on the other hand feel suffocated by too much choice [Schwartz, 2006]. By guiding people to products they are likely to enjoy, recommender systems have been steering people away from the myriad of choices towards a small selection of items they are likely going to enjoy, thereby reducing people’s fatigue caused by having to make decisions [Häubl and Trifts, 2000]. But this filtering and steering could still be improved upon in many ways.

1.2 Increasing user satisfaction

The goal of any personalized recommender system and movie recommender systems in particular is to give the user or customer a satisfying experience [Liang et al., 2007], which in turn may lead to financial gain for the recommender. While giving people a satisfying experience is quite a vague goal, it seems obvious that a recommender would want to recommend movies that a user is most likely to enjoy (in whatever way). This kind of filtering and ordering based on a person’s personality and taste has been
done by many recommender systems for many years now, prime examples being Amazon’s book recommendations (Linden et al., 2003) and Netflix’ movie recommendations (Bennett and Lanning, 2007). Constructing the models necessary to analyze user personality and taste and predict how much an item will be liked, can be done in several ways and using many different sources of information.

1.3 Subject of this research

Two rather new fields of study are the fields of context-aware and mood-based recommender systems, i.e. recommender systems that not only take ratings of users and general user personality into account, but also consider the user’s (social) environment and the user’s mood. In this research, we investigate the feasibility of recommending movies based on a person’s mood. Specifically we construct affective profiles for movies, meaning that we will be classifying movies using content-based information that may be related to mood. We also investigate the influence of movies with certain affective profiles, on the person’s mood after watching a movie.

Aside from their particular purpose in our research, if successful, these affective profiles can be used to analyze the affective influence of products in general, and movies in particular. Furthermore they could be used to build a content-based recommender system using mood. And lastly they could be used to bootstrap collaborative recommenders that want to use mood, but do not have the necessary amounts of data yet to rely solely on collaborative information.

Thus our main research question is:

Question: Can we predict a person’s mood after watching a movie, based on the person’s current mood and information about the movie?

Practically, we will address this question in two parts, as can be seen in Figure 1:

1. How do we generate an affective profile of a movie based on its metadata?

2. Can people’s moods after watching a movie be predicted to a significant degree using these affective profiles?

1.4 Paper overview

In section 2 we elaborate on the current state of affairs in the recommender systems landscape to provide a context for the motivation of this research. In section 3 we discuss mood-based recommenders and the different underlying problems they have to deal with. Our experimental investigation consists of two parts: in section 4 we discuss the construction of the aforementioned affective profiles, while section 5 addresses our experiment for gathering data about people’s moods. In section 6 we will discuss the results of these endeavors, including the influences of these affective profiles and other available variables on a person’s mood and the feasibility of using this information to predict mood. Lastly, in section 7 we make some conclusions and suggest areas for future research.

2 Recommender systems

There are many ways in which recommendations can be generated, but mostly they fall in two categories: collaborative recommendations and content-based recommendations. We describe these below.

2.1 Common techniques

Collaborative recommendation (also known as collaborative filtering, CF) is probably the most widely used method of making recommendations and is also used by Amazon and Netflix (Linden et al., 2003, Bennett and Lanning, 2007). Collaborative recommendations are based on the similarity of different customers and items, calculated for example by tracking how often two items are bought by the same customers (similar items) and how often two customers buy the same items (similar users) (Sarwar et al., 2001). With this information users can be matched with items they are most likely going to enjoy (Fig. 2). One of the advantages of CF is that the system does not require any knowledge about the items or the users, i.e. it is content-free. One disadvantage is that its predictions are unreliable for items that are unpopular or newly added to the database, as well as for new users because the system does not have enough information yet to find similar items and users in these cases (see section 3.1.5).

Figure 2: Collaborative filtering uses information about e.g. ratings from users for certain items to find users similar to each other and items similar to each other. This can be used to estimate the rating a user will give for a particular item.

The second category, content-based recommendation, concentrates on exploiting information about users and items to filter irrelevant items from relevant items (Pazzani and Billsus, 2007). It develops a model
of preferences that a user has, which it then uses to recommend items exhibiting those preferences. The advantage of content-based recommendations is that they are less vulnerable to sparse data and can immediately build recommendations for new items that have not been bought or rated by any customers [Melville et al., 2002].

Another important way in which recommender systems differ, is in what exactly they are trying to predict or influence. Amazon's recommender system tries to maximize the probability that a customer will buy an item, while a movie recommender system such as Criticker or Netflix tries to maximize the scores users give movies. These latter systems do not have to maximize sales of particular items, since they are free and subscription-based respectively. Theoretically, a recommender system could try to influence any kind of variable, be it sales probability, rating or even mood (see section 3.1). Using collaborative filtering to predict mood has currently not been attempted yet.

### 2.2 Context-aware recommending

One aspect of currently implemented recommender systems is that they are mostly static, meaning that their database of recommendations for a single user does not change significantly from one moment in time or place to another. I.e. the system will recommend the same book or the same movie whether you’re tired or happy, and whether you’re at home or at a friend’s place, simply because it is not aware of these changing patterns. In other words, these systems only look at your past buying or watching behavior and generally do not take into account your current context or environment.

Your personality and your taste in books or movies is unlikely to change significantly in one day, so it makes sense to concentrate on the static aspects of recommendations. Koren (2009a) was one of the early adopters of time-aware models in collaborative filtering in order to take changing tastes of individuals and the community as a whole into account. However, what movie you would really like to watch at this very moment may depend on a large number of very dynamic factors that change much faster than in the span of months or years. Ranganathan and Campbell (2003) list a number of contextual factors including many that seem relevant to movie watching: physical contexts (e.g. location, time), environmental contexts (weather), personal contexts (health, mood, schedule) and social contexts (group activity, social activity). So if a system’s goal is to help you pick a movie that you’d like to watch at this very moment, it could possibly be much more effective if it not only took your past preferences and movie taste into account, but also your context. In this way we can filter out many movies that do not fit with your current context and thus reduce the amount time you need to make a choice.

### 3 Mood-based recommending

In this research we want to concentrate on one of those contextual clues that could possibly improve recommendations: mood. To a significant degree, mood determines what kind of entertainment you yearn for, and what motivations you have in consuming that entertainment, as was first shown by Zillmann (1988).

A prime example of a movie that would be heavily influenced by a recommender system taking context into account is Requiem for a Dream, a movie by Darren Aronofsky from 2000 that delivers a powerful, bleak and harsh view on drug addictions. Cinematically it is a critically acclaimed movie, so it is generally highly recommended, especially for cinema-afficionados. However, because of its bleak and harsh nature, it is not a movie that is easy to stomach: “Seldom has a film so powerfully affected me as Requiem for a Dream has — affected, in this case, as if my eyes and psyche have been bludgeoned.” writes Jason
Figure 3: Conceptual framework of a mood-based recommender. The green rectangle shows the realm of traditional recommender systems, with collaborative filtering at the heart. A mood-based recommender additionally uses the current mood of the user, and a mood profile of movies and users to estimate the mood of the user after watching the movie. Questions to be tackled are 1a) how to predict a person’s mood, 1b) how to build a mood profile of a movie, 2) how to model moods, 3) how to measure a person’s mood and 4) how to recommend movies using mood.

Gorber of Film Scouts magazine (Gorber, 2007).

In other words, this is the kind of movie that people are usually not in the mood for. It is unlikely to fare well on a girl’s night out, or after a depressing day at work. But nevertheless, it is one of many movies on the famous 1001 Movies You Must See Before You Die list (Schneider, 2012). A context-based recommender that takes mood into account would be able to find a suitable moment to recommend you a heavy movie like Requiem for a Dream. It would recommend lighter fare on other occasions, e.g. when you are tired, stressed out or generally not in a good mood.

To gain insight into the different mechanisms at play in a mood-based recommender, we provide a conceptual framework in Figure 3. Traditional recommender systems consist of a collaborative filtering engine with user ratings and movies as input and estimated ratings as output. A mood-based recommender engine would then use these estimated ratings, information about the movie and the current mood of the user to estimate resulting moods for each possible movie. Combined with the estimated ratings, a fitting recommendation can then be made.

There are a number of open questions that play important roles in actually building a mood-based recommender, including but not limited to the four points of interest discussed below, also indicated in red in the overview figure.

3.1 How to predict mood

The first question that comes to mind is the somewhat overarching question: how do you predict how a person’s mood has changed after watching a movie? Many factors will influence this mood change, many of which are either invisible, not obvious or not measurable. Whether these influences are so big that not being able to take them into account makes predicting moods impossible is one of the questions this research hopes to answer.

We hypothesize that if we can predict moods to a significant degree in an uncontrolled setting and with a limited amount of data, a larger effort using an existing user base, such as any of the user bases currently present on movie recommender sites, could yield substantial improvements to recommendations. The noise on a single individual’s mood will still be high, but any general trends in the population or smaller groups within the population should be more readily observable and exploitable.

The three most obvious (but definitely not the only!) factors that will influence this mood change are the characteristics of the person, the person’s mood before watching the movie and the characteristics of the movie.

3.1.1 Influence of personal characteristics

Personal characteristics are difficult to investigate with a limited dataset, however this is information usually already available in some form to recommender systems, such as Netflix. These collaborative recommender systems already know which people you seem to be similar to, and recommends movies on that basis. In similar fashion, if hypothetically a large database of mood-information were available, people could be matched with other people whose moods...
respond in similar ways to certain kinds of movies (Fig. 3). For example, such a system might automatically cluster people in two groups based on how they respond to horror movies: one group of people that become disgusted and terrified, and a second group that become excited or cheesy, a distinction hinted at in [Palmer] (2008).

Figure 4: Collaborative filtering could theoretically also be used to predict moods, since information about moods or mood changes from other users can be used to cluster people and movies, and thus predict mood changes in the same way as ratings can be predicted.

3.1.2 Influence of mood

The way you emotionally respond to a movie will likely to at least some degree depend on your mood at the time you started watching the movie. It is quite possible that a happy, excited person generally responds differently to a stylistic action movie, than a tired, sad person does. The tired person may not be aroused enough to respond to the movie, while the stylistic action scenes may resonate much more with the happy person.

3.1.3 Influence of movie characteristics

Though it is unclear how movies influence mood exactly, these influences do exist ([Canini et al.] 2010, Hubert and de Jong-Meyer [1991]), so it stands to reason that their influence on mood could be predicted to some accuracy using collaborative filtering in the same way user ratings are estimated in traditional recommender systems (Fig. 3). Similar films are likely to elicit similar affective responses in users. This is more or less a black box approach though, and may be more effective if accompanied by a psychology-based model of the influence of movies on mood.

3.1.4 Influence of social and other contexts

Lastly, environmental, social and other contexts also may have a significant influence on your overall experience of watching the movie and your mood in particular ([Ono et al.] 2007). Watching a movie with your mom may influence your mood in a different way than watching a movie with a bunch of friends. It may be that in the company of friends you respond better to high levels of action and humor, while in the company of your mom or partner you may respond better to other genres. Furthermore, watching a movie at home seems quite different from watching a movie in a movie theater or at a friend’s place. Using this information to better predict ratings has already been shown to be effective by [Adomavicius et al.] (2005), but using it to predict mood seems to be a new endeavor. We will investigate these influences only to a limited degree here, although this might be highly interesting for future research in context-based recommending.

3.1.5 The cold start problem

Even if you can take all these influences into account and put this information into a collaborative recommender to predict mood, collaborative filtering still suffers from what is called the cold start problem ([Schein et al.] 2002). When the matrix linking items to users is sparse, as is usually the case when a database has thousands or millions of users and items, for some items and users it will be difficult to find similar items and users. This is especially the case when a new user or a new item enters the database for which nothing is known yet. [Melville et al.] (2002) shows that this cold start problem can be alleviated by temporarily using content-based recommendations, using external information about either items (e.g. genres or keywords) or users (e.g. demographics).

In our research we investigate whether mood and genre-related metadata about movies, freely available in databases such as IMDb (imdb.com), The Movie Database (themoviedb.org) and Jinni (jinni.com), can be used to predict mood. If so, an affective profile constructed from this metadata could be used as a stepping stone to overcome the cold start problem in a working recommendation system. Particularly Jinni is of interest to our research, as their database contains mood tags for thousands of movies, such as bleak, emotional, witty, feel good, etc. We hypothesize that the general population’s mood change after watching a movie could be predicted to some significance using these mood tags.

3.2 How to model mood

Before mood descriptions can be used in a recommender system, they have to be mapped to numbers, since all recommender systems in one way or another translate words and properties to numbers before they can calculate with them. Moods (which are longer lasting and less intense than, but similar to, emotions) are particularly difficult to translate into numbers, but several quantitative models of emotional state exist, such as the Pleasure-Arousal-Dominance model by Russell and Mehrabian (1977) and Lövheim’s cube of emotion ([Lövheim] 2012).
Many lower-order systems exist as well such as the PANAS-X scale \cite{Watson1999}, but those contain many more variables or dimensions, which may make it difficult to draw conclusions that can be generalized to new data.

3.2.1 The Pleasure — Arousal — Dominance model

Because the PAD-model consists of only three dimensions and can be used in many general applications, we chose to use this as a quantitative model of user moods and affective profiles for movies. The PAD-model has been used before in the context of movies by e.g. \cite{Eliashberg1991}, although Payne and Shaw \cite{Payne1998} warn that the PAD-model may have to be modified slightly to accommodate moods that according to PAD theory should not occur together, but do seem to exist in unison in some leisure contexts according to their findings.

The three dimensions used in the PAD-model (see also Fig. 5) are:

- **Pleasure–Displeasure scale:** also known as the *Valence* scale, expresses how pleasant an emotion is. Fear is an unpleasant emotion, joy is a pleasant emotion.

- **Arousal–Nonarousal scale:** expresses the intensity of an emotion. Rage, fear and boredom are all unpleasant emotions, but rage has a higher arousal value, than fear and even more so than boredom.

- **Dominance–Submissiveness scale:** expresses how controlling or dominant the emotion makes you feel. Anger and fear are both unpleasant emotions, but anger is a dominant emotion, while fear is a submissive emotion.

3.3 How to measure mood

There are several ways to measure a person’s mood with different levels of accuracy and intrusiveness and probing different observable consequences of those moods. Emotion (and thus mood) can be measured by looking at overt behavior in your face, body or voice, by measuring physiological responses using electroencephalograms (EEG’s), heart rate monitors and galvanic skin response sensors, and also using self-reporting in (extensive) questionnaires or using self-assessment manikins \cite{Mesken2002}. To keep the practical relevance of this research as high as possible we chose to look for a way of measuring mood that allowed people to self-report it at home without spending a lot of time and without much effort (see section 5.2).

3.4 How to recommend based on mood

Once a recommender system has an estimate of what your mood will be after watching any movie, the question that arises then is how to recommend movies, using that information. This is largely a question of what the user is seeking in his entertainment at that very moment.

3.4.1 Motivational cues and desired mood

The general reason that *Requiem for a Dream* will not fit in every context is that people’s motivations for watching a movie changes with time and context. Tesser et al. \cite{Tesser1988} identify three major motivations people have for seeing movies: self-escape, entertainment, and self-development. These motivations can be different every time you watch a movie and will depend on your current context. So it seems logical that if a recommender maximizes the fulfillment of these motivations, this will yield the largest possible user satisfaction.

For this to work ideally, a recommender would need a way to reliably guess what motivation the user is currently seeking, so that it can tailor its recommendations to this particular motivation. If a user is seeking self-escape, a light comedy is probably more suitable than a long movie about people’s misery. On the other hand, when a user is seeking self-development a movie about the genocide in Rwanda is probably a better fit, than a slapstick comedy.

More generally though, it would be useful for a mood-based recommender to know what moods are desired when a user is seeking a certain motivation. This may be difficult to investigate, as it likely depends highly on a person’s personality. Generally users tend to pick activities or movies that maintain or maximize pleasure and minimize pain \cite{Zillmann}.
There are a number of contextual sources of information that could be used to discriminate the motivations users have in watching a movie and that may provide a clue in figuring out what mood is probably desired for a user. For example, who the user is watching the movie with and whether he is watching at home or in a theater may already discriminate some motivations. We postulate that one other clue that could provide some hints in this regard is people’s current mood since it is quite likely that people with the same mood may be having similar motivations in looking for entertainment.

3.4.2 What movie fits what mood?

To investigate this, one possible experiment would be to gather data about people’s movie watching behavior when they are in certain moods. Every time a person watches a movie, he or she likely made a deliberate decision to watch this movie. To help people with this decision it makes sense to first find out how people’s moods influence this decision when they make it themselves. Likely there will be a general trend steering toward certain movies in certain moods, but equally likely this will be slightly or significantly different in different groups of people. While not the focus of this research, our experiment did allow investigating this relation to some extent (see section 6.5). Any results have to be taken with a grain of salt however, since Caruso and Shahi (2006) showed that merely thinking about your mood (e.g. when participating in the experiment conducted in the present research) already influences your decision making in entertainment selection and thus may skew any visible relationships.

3.5 Previous research

The field of mood-based recommendations is relatively new and underdeveloped, but quite a bit of research has been carried out.

Gonzalez et al. (2007) for example tested ways of implicitly capturing emotional context of users through emotional intelligence, and used this in a recommender system for online courses.

Winoto and Tang (2010) studied the role of user mood in movie recommendations and found that users in a good mood rated certain movies significantly higher. This causes bias in ratings used in recommender systems. They also tested a CF system incorporating mood to predict ratings which they claim outperformed a normal CF recommender. Their experiment consisted of questionnaires in which they asked people to rate movies they had seen in the past, so the dynamics of mood and rating movies may be different in a real-world setting where people watch a movie in a certain mood and rate it immediately or soon afterwards, still being in the mood caused by the movie.

Wang et al. (2010) also tested a CF system that uses mood information, using the Moviepilot dataset released for the Challenge on Context-Aware Movie Recommendation (CAMRa2010). One major difference is that they did not use mood of users themselves in their system, but only affect-related information about movies (in the form of 16 different mood id’s) to find users that like the same moods in movies. Just like traditional recommenders, this system finds your average taste in movies (moods).

Adomavicius et al. (2005) consider the notion of context-aware recommenders by proposing a multidimensional approach in which they use contextual information as new dimensions in the recommender in addition to the user and movie dimensions. As in our research, they considered time, location and companions as extra dimensions. Watching a movie in a movie theater and watching a movie in the weekend turned out to be important factors to consider when predicting user ratings in their experiment. Perhaps moods may be influenced by these contexts in similar ways.

Ono et al. (2007) used a Bayesian network to incorporate user attributes, situation attributes and movie attributes into a recommender system. Their idea was that these three input sources cause “impressions” in the user, which in turn influence the rating a user gives a movie, a model similar to the conceptual framework we constructed for mood-based recommenders (Fig. 3).

In a more real-life setting, in 2005 a startup, called MovieProfiler, built a system in which a user could filter movies based on their affective content (Pesonen 2005). For example, you could search for movies with very high levels of fear, but low levels of disgust. The 8 different emotions used on the site were based on Robert Plutchik’s psychoevolutionary theory of basic emotions (Plutchik 1980). As mood levels for each movie had to be entered manually by generous users, interest waned and the website shut down a few years later.

4 Constructing affective profiles

Taking into account past efforts of research into mood and context-based recommenders, we postulate that building affective profiles for movies may be a useful first step in predicting user moods and generating useful recommendations. Thus, as we already mentioned in section 1.3, our first step in trying to predict people’s mood after watching a movie, is to construct these affective profiles. The information we use to create these profiles is metadata for each movie. The idea here is that the affective influence of a movie on a person’s mood may be predicted to some degree using affect-related information about a movie.
4.1 Metadata

To create affective profiles for each movie, we used 29 mood-related tags from Jinni and 22 genre tags from IMDb. Jinni is a movie recommendation engine that uses, what they call, the “Movie Genome” to generate content-based recommendations. This “Movie Genome” is a combination of tags related to the mood, tone and plot of the movie and was generated for each movie “using advanced machine-learning technology and Natural Language Processing” by “analyzing user reviews and metadata” (Glick, 2008). IMDb is the largest movie database on the internet and lists all the genres that apply to each movie. A list of all 51 mood tags and genres, listed alphabetically, can be found in Fig. 1.

Once all applicable tags for each movie are known, each movie can be pinpointed to a location in the 51-dimensional space spanned by these tags.

4.2 Mapping tags to PAD-space

Using 51 variables in a model makes it hard to generalize the model effectively, so we employed 3 different strategies to map these 51 dimensions into three dimensions more or less corresponding to the three affective dimensions in the PAD-model (section 3.2.1). These higher-order affective dimensions, signifying possible mood-related influences on viewers, were then used to find significant relationships between movies and mood (section 3), and can in later research be used as input for mood-based recommending.

4.2.1 ANEW-approach

In the first approach to mapping these 51 dimensions on just 3, we took affect values in PAD-space from the Affective Norms for English Words database (ANEW), constructed by Bradley and Lang (1999), for each of the 29 mood tags and 22 genre tags. The ANEW database was designed to provide emotional ratings for a large number of words in the English language in terms of pleasure, arousal and dominance. For each tag that was not directly present in the ANEW database, we used a thesaurus to make a list of possible synonyms of the tag, and averaged the affect values for all synonyms present in the database.

For the 8 out of 51 tags for which we could not find synonyms present in the database, we took normalized and rescaled values from the AffectButton approach (section 4.2.2). This meant that a tag one standard deviation away from the mean from the AffectButton approach, would also be one standard deviation away from the mean in the ANEW-approach. The following tags were treated in this way: atmospheric, bittersweet, feel good, offbeat, stylized, thought provoking, documentary, film-noir.

One drawback from this approach is that the words from ANEW were not rated in the context of movie content. This drawback is tackled in our next approach. Furthermore, as we already noted in section 3.2.1 groups of people may react differently to movies with the same affective profile (e.g. horror movies).

4.2.2 AffectButton approach

To ensure that affective ratings would be determined in the context of a movie’s content, we asked 22 people to rate all the mood-related or all the genre-related tags in the context of watching a movie, yielding 11 test subjects for each tag. Specifically, we asked: “Please indicate how you think an average person would feel after watching a movie tagged as being...” This rating was done using the AffectButton also used in the mood-gathering part of our experiment (see section 5.2). While the tags in our experiment were not rated by as many people as the tags in the ANEW database, in our experiment we specifically polled for effects on mood after watching a movie, which for some tags may yield a significant and interesting difference from the ANEW-based affective ratings.

As expected, the affective values for both approaches correlated quite well, with a correlation of 0.83 for pleasure, and 0.52 for both arousal and dominance. See Figure 6 for scatter plots showing the distribution of these tags in both approaches. There are a few outliers such as sentimental, emotional and thriller, which are all rated significantly higher on pleasure in the ANEW-database than in our AffectButton approach. While the cause of these outliers is uncertain, it seems for example quite plausible that in the context of movies people associate the term emotional with sad movies and thus low pleasure, even though the word itself may have a higher pleasure-rating in a more general context.

4.2.3 Similarity approach

A completely different approach we investigated is concerned with how often tags appear together in the tag lists of movies. We postulate that tags that frequently occur in the tag list of a movie are very similar to each other and will therefore yield similar emotional responses in viewers, while tags that hardly ever co-occur in a tag list are very different from each other. These similarities can be used to build a map of tags, wherein tags that are similar to each other will appear close to each other.

To investigate this, we downloaded the list of all movies present on IMDb, publicly accessible via FTP. To raise the success rate in downloading tags from the much smaller Jinni-database, we filtered the IMDb list so that only feature films with more than 100 votes remained. This resulted in a list of
Figure 6: Scatterplots showing ANEW vs. AffectButton-based emotional ratings, with linear fits (black) and 95% confidence intervals (gray region). There are a number of outliers, especially in the case of the pleasure axis. Here thriller, sentimental and emotional were all estimated in the AffectButton-based experiment to result in moods showing low pleasure, while the ANEW-emotional ratings seem to indicate a high pleasure rating.

\[ D = \tan \left( \frac{1 - MCC}{2} \cdot \frac{\pi}{2} \right), \]

which maps \([-1, 1]\) to a distance metric in the range \([0, \infty[.\] We then used multi-dimensional scaling (MDS) to build a 3-dimensional map of tags. From now on we will refer to this 3D-space as the MDS-space. We used 3 dimensions because we assumed this could potentially translate into the 3 dimensions of the PAD-model. While this dimensionality reduction has a sound basis, the three resulting dimensions, which correspond to principle component axes, are difficult to interpret.

### 4.3 Combining the techniques

Each of the three techniques mentioned has its disadvantages. The ANEW-approach uses experimental data not gathered in the context of movies, the data from the AffectButton approach is limited, so might be less accurate than the other two approaches, and the similarity approach results in dimensions that are not readily interpretable.

Therefore we investigated the possibility of mapping the MDS-space from the similarity approach onto the PAD-space from the two other approaches. Since it is unlikely that there are significant second-order relationships in these spaces, we decided to use generalized linear models (GLM) to transform MDS-space to PAD.
Figure 7: Combined PAD-values for each tag by averaging all three approaches from section 4.

Figure 8: 2D-projection of all 51 mood and genre tags after the three techniques from section 4 have been combined. The colour scale shows the Dominance dimension which is quite correlated with the Pleasure dimension.

We fitted the locations of the tags in MDS-space to averaged PAD-locations from the two other approaches (ANEW and AffectButton). The fitted values resulting from this linear model are then the PAD-values for the tags transformed from MDS to PAD. Finally, we scaled each of the three PAD-dimensions to a mean of 0 and standard deviation of 1 and then for each tag took the average PAD-location resulting from each of the three approaches:

\[
\text{PAD}_{\text{avg}} = \frac{\text{PAD}_{\text{ANEW}} + \text{PAD}_{\text{AffectButton}} + \text{PAD}_{\text{MDS}}}{3}
\]

The resulting positions of tags in PAD-space can be seen in Figure 7 and Figure 8.

4.3.1 Final affective profiles

Using these PAD-values for each of the 51 available tags, we can construct an affective profile for each movie by averaging the locations in PAD-space of the tags applicable to a movie. This gives an average location in PAD-space for each movie, which can then be used to predict mood.

In our experiment we not only used the combination of the three approaches, but also used the three approaches themselves, as well as a principle components analysis (PCA), carried out on the tags in the 9-dimensional space spanned by the three original approaches.

Thus for each tag, and each affective movie profile derived from these tags, we have five triplets of locations in five different spaces:

1. ANEW PAD-space (hereafter referred to as ANEW\(_P, A, \text{ and } D\))
2. AffectButton PAD-space (AffB\(_P, A, \text{ and } D\))
3. Similarity MDS-space (MDS\(_1, 2, \text{ and } 3\))
4. Combined PAD-space (Avg\(_P, A, \text{ and } D\))
5. Principle Components space (PCA\(_1, 2, \text{ and } 3\))
5 Predicting mood

Now that we can construct an affective profile for any movie present in Jinni’s and IMDb’s database, we can use this information to try to predict people’s mood after watching a movie. We suspect that we could make these predictions using a number of factors that influence your mood as mentioned in section [3.1]. To test this, we set up an experiment to gather data about people’s moods before and after watching a movie, as well as other contextual variables of interest such as location, time and companions.

5.1 Where to gather data?

We considered a number of ways in which we could gather this data. For example, we could have organized several events for students to watch movies together and ask people to submit their moods or we could have asked random people at a movie theater to help us out before and after they watch one of the new releases. Ultimately we chose to set up the experiment in a way that allowed people to submit their moods at their own convenience at home or any other location. Underneath you find the main advantages of this approach.

5.1.1 Uncontrolled environment

Gathering data about mood in an uncontrolled environment rather than a controlled location, cannot prevent unknown variables from influencing people’s moods, however this does increase the chance that any significant results arising from this research could be replicated in a recommender system operating in the real world. Furthermore since mood-based movie recommenders are the underlying purpose of this research, it made sense to try to replicate an environment in which such a recommender would actually be used (i.e. primarily by people watching a movie at home).

5.1.2 Balanced number of people and movies

Another reason to allow people to submit their moods at their own convenience, is that it ensures that we could gather data corresponding to many different movies. Organizing one big event with many people watching the same movie has the advantage that you can investigate the different ways in which people respond to the same movie, but any knowledge you gain from this is hard to generalize to other movies without repeating the same experiment with many other movies.

5.2 How to gather mood data?

As mentioned in section [3.3] there are a number of ways to measure people’s mood. We chose to use a technique that could be implemented in the living room with currently available technology and does not put too much burden on the user.

This not only ensured that the amount of submitted data was as high as possible, but also ensured that whatever conclusions we drew, could in principle be put into practice without needing expensive or intrusive equipment or techniques and materials that would currently be considered futuristic, such as heart rate monitors and sweat conductivity sensors in your TV’s remote control or movement sensors in your mobile phone.

Taking both the desired description of moods and the desired levels of intrusiveness and accuracy into account, we chose to measure moods using the Affect-Button developed in [Broekens and Brinkman 2009], which allows a person to self-report his or her mood in PAD-space with one mouse click or finger tap using a dynamic emoticon that can be tweaked with mouse or finger movement. One advantage of using the AffectButton is that it does not rely on the user’s interpretation of words, but uses visual interpretation of faces, in a similar way that self-assessment manikin’s have been used to self-report mood in PAD-space in the past (Bradley and Lang 1994).

The AffectButton has not yet been verified to work accurately in self-reporting mood, but has been verified to work accurately when judging affective value of standard stimuli (Broekens and Brinkman 2009). Whether the accuracy with which people use the AffectButton for the first time is as good as when they have been using the tool several times is a question we answer in section [6.1.1].

5.3 What other data to gather?

To control for any secondary effects and to find out which factors have the biggest influences on mood, we decided to not only gather information about people’s mood, but also the following information:

Basic demographic information: gender, age, country.

Ethnicity: Fischoff et al. (1998) found significant differences in movie preferences between different ethnicities, so movies might have different influences on mood for different ethnicities.

Time: time and date at which the person started watching the movie.

Title: obviously in order to construct an affective profile for the movie a user watched, we had to know what movie it is they watched.

Who chose the movie? Whether you chose to watch the movie yourself or whether somebody else decided to watch the movie and you went
along with it, might influence the way you react emotionally to the movie. More importantly, this variable allows us to filter out data points from people who did not choose the movie themselves. We need this information when we investigate the influence of mood on the movies you choose to watch (section 5.5). (Possible answers: “me, today”, “me, yesterday or before”, “somebody else”)

First time? Whether you are watching the movie for the first time or not may have an impact on your emotional response. (Possible answers: “yes”, “no”)

With whom? Your emotional response to a movie may be influenced by the people that accompany you. Intuitively it seems that watching a movie with your parents may be different than watching a movie with your friends. (Possible answers: “alone”, “with my significant other”, “with friends”, “with family”, “with colleagues”, “other”)

Where? Watching a movie at home may have a different influence on your mood than watching a movie at the theater or at a friend’s place. (Possible answers: “at home”, “at a friend’s place”, “in theater”, “somewhere else”)

(Expected) rating: Lastly before people started watching, we polled how much they expected to like the movie on a scale from 1 to 10. Afterwards, we polled how much they actually liked it. We asked this because we suspect that when you liked a movie, your mood will also have improved, not just from the movie itself, but also from the fact that you liked it.

5.4 Mobile web application

To put all this into practice, we designed a mobile web application located at http://inthemood.tv that allowed people to submit their mood than watching the AffectButton at their own convenience before and after watching a movie. No registration was required, as cookies were used to keep track of each user between submissions.

Using the AffectButton on a touch screen such as a smartphone or tablet is a bit less intuitive than using it with a mouse, so a small tutorial was shown before you entered your mood to ensure people were able to use the AffectButton correctly.

Each time a person went to the application to let us know he or she was about to watch a movie, we asked for their mood (using the AffectButton), their basic demographic information (age, gender, country, ethnicity), as well as the title of the movie, and the rating they expected to give the movie. After having watched the movie, when the user went back to our application, we again polled for their mood, and then also asked about the contextual clues: who they watched with, where they watched, who chose to watch the movie, whether it was their first time watching it, and finally how much they liked the movie.

We did put in some time constraints for submitting your mood after the movie. This was done to ensure that people were submitting their mood immediately after the movie ended and not several hours later when their assessment of their mood after the movie may not be as accurate anymore due to memory biases. Especially the ‘fading affect bias’, which causes emotions associated with unpleasant memories (bad movies) to fade more quickly from memory than pleasant memories, could have an effect here (Walker and Skowronski, 1996).

5.5 Running the experiment

We ran the experiment from Saturday February 23rd, 2013 until Sunday May 18th, 2013. We promoted the experiment with fellow students, friends and family, on several online film communities, and advertised on Facebook to people in the Netherlands with interest in movies and owning smartphones.

6 Experimental results

Here we describe our experimental results. First we discuss the calculated affective movie profiles, then we give an overview of the gathered data, discuss some anomalies and finally elaborate on the mood prediction results.
6.1 Affective profiles

Affective profiles for all movies watched by the people participating in our experiment were generated by averaging the locations of each of the movie’s mood and genre tags in PAD-space, using each of the three approaches mentioned in section 4. We then also calculated the locations in the combination of these three approaches, and using the principle components as explained in section 4.3.1.

Note that the movies watched during the experiment span the distribution of mood tags quite nicely, as can be seen in Figure 10. This means that the data from our experiment should be reasonably representative of all movies and any results should therefore be generalizable to a significant degree.

6.1.1 Repeated use of the AffectButton

To check if users had trouble using the AffectButton at first use and got better as they used it more often, we randomized the order in which mood tags had to be rated in the AffectButton approach (section 4.2.2) so that no tag was only rated early in the process or late in the process. This allowed us to check people’s “errors” in rating the tags as they grew more experienced. If it were found that at first use the variance in people’s ratings for each tag was much higher than after 20 ratings, it would mean that people the first use would be unreliable. The results can be seen in Figure 11.

Notice that the 95% confidence region of the line fit includes a horizontal line indicating 0 correlation. We also performed an ANOVA on this data using RMSE $\sim$ experience as the model, which yielded a p-value of 0.345, signifying no significant relation between experience and root-mean-squared error.

This is an important finding, since it strengthens the hypothesis that the AffectButton is already accurate at first use and users do not need a lot of experience using it. This means that data points from our mood gathering experiment are valid even if the person never used the AffectButton before.

6.2 Summary of the data

In total this experiment yielded 155 data points. 15 of these have been removed because no Jinni-tags were available for the movie or because it concerned a TV show rather than a movie. 140 good data points remained, corresponding to 80 different movies and coming from 45 different people. In Figure 12 distributions are shown of all data points in the most important variables. Note that in some variables a distinction was made between data points gathered during a special event, called the Oscarweekend, and other points (see section 6.3.3).

6.3 Anomalies in the data

Looking at the raw gathered data from Figure 12 it is obvious that there are a few variables that are not representative of the movie watching population as a whole. Therefore a few variables had to be discarded or were altered for the analysis because any results obtained from them would be difficult to generalize. This concerns the following variables:
Figure 12: First row: Histograms of PAD-values for people's mood before and after a movie, as well as the movies they watched (Avg P, A, and D). Second row: distributions of data points vs. day of the week, hour of the day, and movie appreciation. Third row: distribution of data points vs. location, the people watching along, and age. Fourth row: distribution of data points vs. whether they've seen the movie before, gender, country and ethnicity.
6.3.1 Ethnicity & country

Due to the nature our promotional channels, we were unable to gather a significant amount of data points from non-caucasian people. Therefore we decided not to take this variable into account even though it might play a role in the influence of movies on mood. For identical reasons the country variable was discarded as the data was dominated by people from the Netherlands and Belgium.

6.3.2 Watching with family

A person claimed to be watching a movie with family 28 times. Of these data points however, 22 are data points from a single family consisting of a mother and daughter that watched about 10 movies together over the span of the experiment. This anomaly can also be seen in the correlation matrix from Figure 16 where a newly created variable called Belgian.Family correlates strongly with With.Family and Gender, but also negatively with the Dominance dimension in moods before and after watching a movie. Since one family cannot be considered representative of all possible families, we chose to merge these data points with the 18 data points that were labeled as watching “with my significant other” and the 39 data points labeled as watching “alone”. This reduced the question of “Who did you watch with?/” to the question “Did you watch with friends, yes or no?”

6.3.3 Oscarweekend

We have to acknowledge that some of our variables are substantially correlated with one specific event that we used to kickstart this experiment, called the ‘Oscarweekend’. During this event people watch 5 Oscar-nominated movies in one day. As is evident from Figure 12, there are two variables correlated with the Oscarweekend during which we gathered 28 of our 140 data points:

- **In theaters:** while seemingly a valid variable, it has to be acknowledged that 28 out of 33 data points labeled as ‘in theaters’ came from the Oscarweekend and may therefore not be entirely representative of how you’d normally feel when going to a movie theater. Therefore we chose to merge all 56 data points that were not labeled as ‘at home’ (including the 28 data points from the Oscarweekend) and label them as ‘not at home’.

- **Time of the day:** Since people watch 5 movies in one day during the Oscarweekend, quite a few movies are watched at unusual times in the morning and early afternoon, causing most of the movies watched early in the day to be movies watched during the Oscarweekend. Thus this variable is not representative of all movies watched early in the day.

6.4 Deduced variables

Using the existing time data, two new variables were deduced, which we will call cos.week and sin.week. They are constructed by mapping the time variable for each data point onto a circle that makes one loop every week. In this circle 0° is Monday morning at 6am, 90° is Tuesday evening at midnight, 180° is Thursday evening at 6pm and 270° is Saturday at noon. cos.week is then the cosine of the angle at which a data point is located in this circle, while sin.week is the sine of this angle. The orientation of this circle was chosen such that the correlation between cos.week and arousal after the movie is maximized. Mapping time on a circle allows days of the week to be treated as a continuously oscillating variable. A graph of all data points in this ‘wheel of time’ can be seen in Figure 13.

![Figure 13: This graph shows all data points mapped onto a circle with Mondays on the right at roughly 2 o’clock, going counterclockwise to the other days of the week. They are color-coded using arousal after the movie A_after (red = low, green = high). Some radial jitter was added to prevent the dots from overlapping too much.](image)

6.5 What movie fits what mood?

To investigate the relationship between a user’s mood and the movie they decide to watch, we did analyses of variance on the 75 of 140 data points that were labeled as ‘chose the movie myself, today’. In this ANOVA, we find no direct relationships between any of the dimensions of user mood before a movie and the affective profile of the movie they decide to watch.
However, we find a significant \((p = 0.002)\) interaction effect between the pleasure and arousal dimensions of a user’s mood. It seems that if pleasure is high, but arousal is low or the other way around (i.e. people are “at ease” or “angry”, see Figure [5], people watch movies with lower pleasure, arousal and dominance than when they have 1) both high pleasure and high arousal or 2) low pleasure and low arousal (“sad” or “delighted”). This seems to be in contradiction with the findings we mentioned in Section 3.1.1 by Zillmann and Bryant [2002] and Zillmann [1988]. They reported that people tend to pick movies that maintain or maximize pleasure, in which case you would expect both “angry” and “sad” people to watch highly arousing and highly pleasant movies.

We did not have the time or the amount of data needed to investigate the causes, motivations or implications for this relationship much further. Nevertheless the relationship between a viewer’s mood and the movies he or she decides to watch is important to be aware of, because recommending movies similar to the movies he decides to watch themselves considering their current mood, can be very helpful. Therefore we leave this as an important point for future research.

### 6.6 Predicting mood

Before we could find a model to predict mood, we first had to select the features that had significant relationships with mood after watching a movie in either of the three PAD-dimensions. This ensures that any model we find to have good capabilities of predicting mood in our dataset, has potential to be generalizable to other datasets.

#### 6.6.1 Significant relationships

To do this selection, we started an Analysis of Variance investigation (ANOVA) into all variables that we had not discarded for the reasons mentioned above in section 6.3. This investigation allowed us to judge whether any correlations between these variables and people’s mood after watching a movie were significant and which variables had the best chance of being valuable in a predictive model.

To clarify our strategy, underneath you find the basic algorithm to determine which variables had significant relationships with (as an example) the Pleasure dimension of a person’s mood after watching a movie, \(P_{\text{after}}\). Basically we tried each available variable \(V_i\) as the independent variable, with the remaining variables as a set of blocking variables \(B\). This means that you calculate the significance of variable \(V_i\), if all blocking variables were to remain constant. Then we removed the variable \(V_{\text{high}}\) with the worst significance (i.e. highest p-value) from the list of variables, and iterated this procedure until the set of remaining variables \(V\) was empty.

For example, in order to calculate the significance of gender, one of the models you would try, is:

\[
P_{\text{after}} \sim \text{gender} + \text{(cos.week + ... + with.friends)},
\]

where \(P_{\text{after}}\) is the dependent variable, gender is the independent variable and \(\text{(cos.week + ... + with.friends)}\) are the blocking variables.

#### Algorithm 1 Calculate significant features for \(P_{\text{after}}\)

\[
\begin{align*}
\text{Y} & \leftarrow P_{\text{after}} \quad \text{(dependent variable)} \\
\text{V} & \leftarrow \text{set of all independent variables} \\
\text{repeat} & \quad \text{for all } V_i \in V \\
& \quad \quad \text{B} \leftarrow \emptyset \\
& \quad \quad \text{for all } j \neq i \quad \text{do} \\
& \quad \quad \quad \text{B} \leftarrow B \cap V_j \quad \text{(list of blocking variables)} \\
& \quad \quad \text{end for} \\
& \quad \text{model } M \leftarrow Y \sim V_i + B \\
& \quad \text{ANOVA with model } M \text{ gives you p-value of } V_i \\
& \quad \text{end for} \\
& \quad V_{\text{high}} \leftarrow V_i \text{ with the highest p-value} \\
& \quad V \leftarrow V \setminus V_{\text{high}} \\
& \quad \text{until } V = \emptyset
\end{align*}
\]

What you then end up with, is a list of variables with p-values, telling us how significant each variable is, keeping in mind that more significant variables already explain some of the variance. Thus each variable in this list with a p-value < 0.05 contains significant information in predicting mood that is not present in any of the other variables.

We performed this procedure for each PAD-dimension. To ensure that any results would not be influenced by the Oscarweekend or the single Belgian family, we repeated the procedure four times: once with all the data, once without data gathered during the Oscarweekend, once without data gathered by the Belgian family, and once without data gathered during the Oscarweekend or by the Belgian family. To be as conservative as possible, we then took the highest p-value (i.e. the worst significance) from these four configurations as our estimate of the significance of this variable in new datasets.

In Table II, you can find the resulting significance levels for each variable’s relationship with \(P_{\text{after}}, A_{\text{after}}\) and \(D_{\text{after}}\).

#### 6.6.2 Interesting observations

As is to be expected, \(P_{\text{before}}\) significantly influences \(P_{\text{after}}\) and \(A_{\text{before}}\) significantly influences \(A_{\text{after}}\). The connection between \(D_{\text{before}}\) and \(D_{\text{after}}\) is not significant. Looking at the list of variables related to the affective profiles of the movies, it is clear that the Arousal dimension has a highly significant relation with all three mood-dimensions. In other words, the
arousal value of the movie not only influences how aroused you get, but also how pleased you are and how dominant you feel. The other dimensions are less significant, although it seems the Principle Components are good candidates to predict Pleasure and Dominance.

This, as simply a significant correlation and regression does not imply causation.

Lastly, note that the score you give a movie is particularly correlated with how pleasant you feel after the movie. Whether the movie was better or worse than you expected, on the other hand, is mostly correlated with your arousal level. Thus, it appears that when you liked a movie, you feel pleasant; when a movie was better than you thought, you feel aroused.

### 6.6.3 Adding noise to ratings

Before we try to find models to predict mood using these variables, let us consider one particular variable: your rating given to the movie after seeing it, \( R_{\text{after}} \). In section 5.3 we mentioned that we polled people for their expected and real rating, i.e. how much they expected to like the movie and how much they really liked the movie. The reason we polled for this, is that recommender systems by their very nature already have an estimate for how much a person is going to like a product (or movie). So it makes sense to use this as a source of information in predicting mood.

In our experiment we did not have access to such a recommender system, so we used the ratings people actually gave to the movie. In order to judge the predictive power of our models, it would be cheating if we used this variable ‘as is’. We know exactly how high a person rated a movie, whereas a recommender only knows an estimate of this.

Therefore we decided to add random normalized noise to this variable to simulate the estimates of people’s ratings a recommender would have. We based the amount of noise on the performance of the winner of the Netflix Prize (Koren, 2009b). This entry achieved an RMSE of 0.8567, where the standard deviation in ratings (using a scale of 1 to 5) in the dataset is approximately equal to 1.1296 (Koenigstein et al., 2011). The standard deviation in ratings in our dataset (using a scale of 1 to 10) is 1.39. Thus to simulate Netflix Prize performance in estimating ratings, we added noise to the ratings with an RMSE of 0.8567/1.1296 · 1.39 = 1.05.

### 6.6.4 Finding a linear model

Now that we know which variables are significantly correlated with your mood after watching a movie, we added noise to the ratings, we searched for the linear models that best fit our data in order to predict your mood after the movie as best as possible.

To test which sets of variables are most suitable to put in our predictive model, we used the algorithm described in Algorithm 2. We started by testing which variable \( V_i \) from the list of variables in Table 1 has the best predictive ability, i.e. yields the lowest root-mean-squared error (RMSE).

To calculate the RMSE for each model \( M \), we used 10-fold cross validation, meaning that we divided the

<table>
<thead>
<tr>
<th>Class</th>
<th>( V_i )</th>
<th>( P_{\text{before}} )</th>
<th>( A_{\text{before}} )</th>
<th>( D_{\text{before}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>mood before</td>
<td>( P_{\text{before}} )</td>
<td>***</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>affective profiles</td>
<td>MDS2</td>
<td>***</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>time related</td>
<td>cos.week</td>
<td>*</td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td>ratings</td>
<td>( R_{\text{after}} )</td>
<td>**</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>context</td>
<td>at.friends</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Table showing significant relationships between independent variables \( V_i \) and people’s mood after a movie. The significance codes (stars) indicate the p-value: ‘***’ \( \Rightarrow 0.001 \), ‘**’ \( \Rightarrow 0.01 \), ‘*’ \( \Rightarrow 0.05 \), ‘.’ \( \Rightarrow 0.1 \). To clarify: \( R_{\text{after}} \) is the rating people gave to a movie after seeing it, \( \Delta R \) is the difference between the given rating and the rating they expected to give.
Comparison of tested models

<table>
<thead>
<tr>
<th>#</th>
<th>Model</th>
<th>RMSE</th>
<th>% of variance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mean $M_{after}$</td>
<td>0.440 0.757 0.500 0.565</td>
<td>1%</td>
</tr>
<tr>
<td>2</td>
<td>$M_{before}$</td>
<td>0.503 0.871 0.591 0.655</td>
<td>−33%</td>
</tr>
<tr>
<td>3</td>
<td>$M_{before} + \text{mean } \Delta M$</td>
<td>0.506 0.863 0.594 0.654</td>
<td>−33%</td>
</tr>
<tr>
<td>4</td>
<td>GLM (all)</td>
<td>0.383 0.657 0.464 0.501</td>
<td>22%</td>
</tr>
<tr>
<td>5</td>
<td>GLM (no ratings)</td>
<td>0.392 0.664 0.464 0.507</td>
<td>20%</td>
</tr>
<tr>
<td>6</td>
<td>GLM (no $M_{before}$)</td>
<td>0.398 0.688 0.464 0.517</td>
<td>17%</td>
</tr>
<tr>
<td>7</td>
<td>GLM (no affective info)</td>
<td>0.429 0.722 0.488 0.546</td>
<td>7%</td>
</tr>
</tbody>
</table>

Table 2: Results of the model predictions showing the mean RMSE on the test set for all three mood dimensions and the percentage of the variance in the dataset the model explains.

140 data points randomly in 10 sets of 14 points. Ten times we then used 9 sets as a training set to fit a linear model using the given variables and tested this fit on the remaining set of 14 points. This procedure was repeated 100 times to give a stable average RMSE.

This variable $V_{low}$ with the lowest RMSE then becomes the first variable in our set of variables $C$ for the model $M$. Then we iteratively add the variable that yields the lowest RMSE when combined with the variables $C$ we already had, until no variables remain or the RMSE does not improve anymore. This approach ensures that we find the simplest model with the best performance.

Algorithm 2 Calculate best set of variables to predict $P_{after}$

1. $Y \leftarrow P_{after}$ (dependent variable)
2. $C \leftarrow \emptyset$ (set of confirmed variables)
3. $V \leftarrow$ set of independent variables
4. $\text{RMSE}_{\text{best}} \leftarrow \infty$
5. repeat
   1. for all $V_i \in V$ do
      1. linear model $M \leftarrow Y \sim C + V_i$
      2. calculate RMSE for model $M$
   end for
6. $V_{low} \leftarrow V_i$ with the lowest RMSE
7. if $\text{RMSE}_{\text{low}} < \text{RMSE}_{\text{best}}$ then
       1. $\text{RMSE}_{\text{best}} \leftarrow \text{RMSE}_{\text{low}}$
       2. $C \leftarrow C \cap V_{low}$
       3. $V \leftarrow V \setminus V_{low}$
    else
       1. quit
    end if
8. until $V = \emptyset$

We did this procedure in a number of configurations, leaving out different variables, to test their respective predictive performances. The four GLM model configurations we tested are:

GLM (all): model using all variables from Table 1

GLM (no affective info): model using only information currently already available to recommenders: rating information and temporal information (day of the week). Thus this model has no personalized information about the user’s mood or affective profiles.

To put the results of these predictions into perspective, we also compared the results from these GLM’s with a few naive or obvious models that predict $M_{after}$ in slightly different ways. These are:

Model 1: Predict the average of all reported moods after a movie ($\text{mean } M_{after}$).

Model 2: Predict the same mood as before the movie ($M_{before}$).

Model 3: Predict the same mood as before the movie + the average mood change caused by watching a movie ($M_{before} + \text{mean } \Delta M$).

The mean values used in these predictions can be found in Table 3.

<table>
<thead>
<tr>
<th>#</th>
<th>Model</th>
<th>P</th>
<th>A</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mean $M_{after}$</td>
<td>0.38</td>
<td>−0.17</td>
<td>0.21</td>
</tr>
<tr>
<td>2</td>
<td>$M_{before}$</td>
<td>always different</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\text{mean } \Delta M$</td>
<td>0.04</td>
<td>0.16</td>
<td>−0.05</td>
</tr>
</tbody>
</table>

Table 3: Values used to predict mood after a movie using the naive models. Note that people generally get slightly more pleased, significantly more aroused and feel slightly less dominant after a movie.
Table 4: Results of the model predictions showing the mean RMSE on the test set for all three mood dimensions and the percentage of the variance in the dataset the model explains.

<table>
<thead>
<tr>
<th>#</th>
<th>Var.</th>
<th>RMSE</th>
<th>∆σ² expl.</th>
<th>Var.</th>
<th>RMSE</th>
<th>∆σ² expl.</th>
<th>Var.</th>
<th>RMSE</th>
<th>∆σ² expl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MDS₂</td>
<td>0.417</td>
<td>11.4%</td>
<td>Avgₐ</td>
<td>0.717</td>
<td>10.1%</td>
<td>Affₐ</td>
<td>0.469</td>
<td>12.7%</td>
</tr>
<tr>
<td>2</td>
<td>Pₚ_before</td>
<td>0.400</td>
<td>7.0%</td>
<td>Aₚ_before</td>
<td>0.686</td>
<td>7.7%</td>
<td>cos.week</td>
<td>0.464</td>
<td>2.1%</td>
</tr>
<tr>
<td>3</td>
<td>Rₐ_after</td>
<td>0.392</td>
<td>3.3%</td>
<td>cos.week</td>
<td>0.665</td>
<td>4.6%</td>
<td>Rₐ_after</td>
<td>0.657</td>
<td>1.9%</td>
</tr>
<tr>
<td>4</td>
<td>PCA₃</td>
<td>0.387</td>
<td>2.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PCA₂</td>
<td>0.383</td>
<td>1.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final:</td>
<td>0.383</td>
<td>25.3%</td>
<td>Final:</td>
<td>0.658</td>
<td>24.5%</td>
<td>Final:</td>
<td>0.463</td>
<td>14.8%</td>
<td></td>
</tr>
</tbody>
</table>

A comparison of the performance of each model can be found in Table 2 and Figure 14.

Figure 14: Graph showing the performance of the most important models in explaining the variance in mood after a movie. M4 = GLM (all), M5 = GLM (no ratings), M6 = GLM (no Mₚ_before), M7 = GLM (no affective info), M1 = mean Mₚ_after.

6.6.5 Linear model predictive performance

As can be seen in Table 2, our model using all variables shown to be significant is able to explain 22% of the variance in the dataset. That leaves 78% that is currently noise or unmodeled signal. If a mood-based recommender would decide to use affective profiles, but does not want to ask people how they are feeling, they can still estimate moods quite well compared to a model with Mₚ_before, since 17% of the variance can be explained. Not using any affective info and only using information that is currently already available to recommenders (day of the week and estimated ratings), only 7% of the variance can be explained, which is significantly worse.

To put things in perspective a bit more, let’s consider the Netflix dataset again which has been researched ad nauseam thanks to the Netflix Prize. The total variance in the Netflix dataset is 1.276, corresponding to an RMSE of 1.1296 (Koenigstein et al., 2011). After three years of multi-team efforts to reduce this RMSE, the Prize was won with an RMSE of 0.8567, corresponding to a variance of 0.7339. This means that 42% of the variance has been explained. This is almost double the variance in mood that we were able to explain in this research.

The winning solution of the Netflix Prize does rely heavily on personalization, although, whereas in this research the only form of personalization we used are the ratings (or in their noisy form, the estimated ratings) given to movies. The other variables in our models are either derived solely from the watched movie (affective profiles), the user (demographics), or time (day of the week), but not from any interactions between those variables. To make the comparison a little more fairly, one of the models we tested did not include rating information, so it only relies on biases in movies or items, not on interactions between them. This yielded 20% of explained variance. The best reported bias-model for the Netflix dataset reached an RMSE of 0.9278, thereby explaining 33% of the variance (Koren, 2009b).

Looking at the different types of variables we used in our models, it is clear that using the mood before the movie to predict people’s mood after a movie is a bad starting point as it performs much worse than just predicting the average Mₚ_after. This is not surprising when you look at the details of the best performing model in Table 4 and the performance in this model by variable type in Figure 15. Here it can be seen that the most important variables in predicting Mₚ_after are the variables from the affective movie profiles (MDS₂, Avgₐ, PCA₂, PCA₃ and Affₐ). Together these explain on average about 13% of the vari-
ance in each dimension, while $M_{before}$ explains about 5% on average.

Though our model only explains 22% of the variance in $M_{after}$, it seems evident that affective profiles of movies in particular and affect-related properties of movies in general seem to have a very large impact on people’s mood after watching a movie. Additionally, how much you like a movie ($R_{after}$) and what day of the week you watch a movie also have significant impacts. While our dataset was not varied enough and relied a bit too heavily on data gathered at the Oscarweekend, we do still suspect that the people you watch a movie with, and the location you watch a movie at, do have a significant impact on your mood. Further research would have to confirm or deny this.

![Performance by variable type](image)

**Figure 15:** Graph showing the performance of the most important models in explaining the variance in mood after a movie. $M1 = \text{GLM (all)}$, $M2 = \text{GLM (no ratings)}$, $M3 = \text{GLM (no } M_{before}\text{)}$, $M4 = \text{GLM (no affective info)}$, $M5 = \text{mean } M_{after}$.

### 6.6.6 Predictive neural network

We tested neural networks using the most significant variables from Table 1, but found no combination of network parameters that could outperform the linear models. Therefore we suspect that any nonlinear relationships in the data are too weak for a neural network to pick up or exploit. This might be a subject for future research.

### 7 Discussion

We set off to use affective movie signatures or ‘profiles’ to predict people’s mood after watching a movie. We succeeded to the extent that:

**Finding 1:** we were able to explain 22% of the variance in people’s moods thanks to a linear model using the affective movie profiles, people’s moods before the movie, their (simulated) estimated ratings and the day of the week.

While this model is obviously not perfect, and it is hard to compare with existing literature since mood prediction in the field of movies has not been attempted before, it seems to be a good first attempt at modeling user mood. Our performance in predicting moods using no personalization was able to get two thirds of the way towards the best reported bias-model for predicting Netflix movie ratings.

As our dataset had limited variety in some variables (e.g. country, ethnicity, watching with family and watching in theaters), future research may be able to get closer using more diverse data, so that more contextual clues potentially influencing moods may be used. Despite the fact that we found no contextual clues to be helpful in building our model, we suspect that more data could still unearth some useful clues in this regard.

Aside from contextual clues, we have found that:

**Finding 2:** people’s mood after watching a movie is largely influenced by the affective profiles of the movies they watch, and only to a limited extent by their mood before the movie.

In fact, the Dominance dimension of people’s mood after a movie was not significantly correlated with any of the ‘before’-dimensions. Resultingly,

**Finding 3:** using affective profiles for movies is very important in trying to predict people’s mood.

Also we found an interesting correlation between days of the week and people’s arousal after a movie. Even though their arousal before the movie is not correlated with the day of the week, on Wednesdays, Thursdays and Fridays, people are significantly less aroused after a movie, than on Sundays and Mondays. It is not clear why this is so.

Lastly, given the success of using these affective movie profiles in predicting mood, we suggest that more research should be done in mood-based recommenders in general and motivational and mood-based clues for recommending movies in particular. Affective movie profiles could help bootstrap mood prediction systems and alleviate cold start problems, while a collaborative filtering system could form a core engine.

We suspect that a solid mood prediction system, combined with a traditional movie recommender system and using the results of future research into the behaviors and desires of people in different moods
and contexts could significantly improve movie recommendations. In turn this will make it easier for people to choose a movie to watch and raise their enjoyment.

References


Figure 16: Correlations between the most important variables available in our analysis. Bigger circles indicate stronger correlations, blue circles indicate positive correlations, red circles indicate negative correlations.