

Make Me Laugh: Recommending Humoristic Content on the WWW

Daniel Buschek, Ingo Just, Benjamin Fritzsche, Florian Alt

University of Munich (LMU) – LFE Medieninformatik

Abstract

Humoristic content is an inherent part of the World Wide Web and increasingly consumed for micro-entertainment. However, humor is often highly individual and depends on background knowledge and context. This paper presents an approach to recommend humoristic content fitting each individual user's taste and interests. In a field study with 150 participants over four weeks, users rated content with a 0-10 scale on a humor website. Based on this data, we train and apply a Collaborative Filtering (CF) algorithm to assess individual humor and recommend fitting content. Our study shows that users rate recommended content 22.6% higher than randomly chosen content.

1 Introduction

Micro-entertainment, content tailored to engage users for brief time spans, has become an integral part of our life: We watch short video clips while waiting for the bus, we read news snippets as we cue for lunch, and we look at cartoons during coffee breaks. However, providers of such content rarely respect the viewer's personal taste and context (Figure 1). For users, presenting tailored content can be more interesting and enjoyable; for content providers, improved user experience may result in longer visits and higher return rates.

Previous work has adapted micro-entertainment content by fitting its length to the expected time available to users in certain contexts. For example, Alt et al. derived waiting times in front of traffic lights from GPS data to find and present short video clips to drivers (Alt et al. 2010). Similarly, loading times on mobile phones have been used to present small chunks of information to the user (Alt et al. 2012). While we also address enhancing micro-entertainment, we do not use waiting times but rather aim to improve the fit of the content to the user's interests, thus expecting more extensive engagement with the content.

In particular, our work focuses on entertaining, humorous content. Thousands of websites offer funny pictures, videos or jokes. However, users may have to browse such pages for quite some time to stumble upon content matching their taste, a problem similar to browsing



Figure 1: Screenshot of humorous content on our prototype website. Users rate such cartoons, video clips, and jokes with a 0-10 scale slider, shown at the bottom. We apply a collaborative filtering algorithm to the resulting user data: For an individual user, our approach recommends new content, which received high ratings from users with overall similar rating behavior. This approach allows us to address individual taste and interests without knowing more about the users' backgrounds.

news websites as covered in related work (Liu et al. 2010). The *distinct humor phenomenon* (Veatch 1998) explains that humor depends on the individual’s knowledge about the topic of the content. For example, not everybody is familiar with Schrödinger’s Cat or Star Wars in enough detail to laugh about related jokes. Unfitting humorous topics may only be of marginal interest to an individual, hence resulting in less time and engagement with the content.

In this paper, we leverage recommender systems to address this challenge and provide fitting humorous content to each individual user. In particular, we propose and evaluate collaborative filtering (CF) to suggest content to each user. Our approach is motivated by the success of CF in current recommendation systems popular on the World Wide Web. CF is widely used in e-commerce – but has also successfully advanced movie recommendations (cf. (Miller et al. 2013) or Netflix Prize¹). Prior work also suggested this concept to be applied to (text-based) newsgroups (Konstan et al. 1997), including different types of content. Our work is different in that it targets websites exclusively focusing on humorous content (including text, images, and videos) and in that it compares the concept to non-targeted content.

We set up a website which compares each user’s humor profile with all other users. Users are considered similar if they rate the same content similarly. Instead of having to know the exact topic of the content, we can thus simply rely on a rating system. In particular, we present a 0-10 rating scale below each humorous content (Figure 1). Following the CF approach, we build user profiles and identify users with the same background knowledge by taking into account all their ratings. For each user, we can then recommend yet unknown content with high ratings by similar users.

¹ Netflix Prize: <http://www.netflixprize.com>

This paper contributes a user study in the wild, exploring the usefulness and acceptance of recommendation methods for personalized humoristic content presentation. Analyses show that users rated content recommended by our CF approach 22.6% higher than randomly presented content.

2 Concept

According to Veatch (Veatch 1998), three conditions have to be met for humor to occur: First, perceivers have the “normal” situation in mind. Second, they also see a violation of this norm. Third, both understandings meet simultaneously in the mind.

In general, a website is not capable of predicting the occurrence of these conditions. By relying on user similarities instead, our concept assumes that these conditions are likely met across individuals, if they have rated similar content similarly in the past. Hence, this assumption introduces a simplified perspective on *user-specific* taste and context, which can render aspects of personal taste and interest in humor accessible for websites.

In summary, our concept creates user profiles based on content ratings. Users with a similar profile are considered as neighbors. An individual’s neighbors provide the foundation for recommending new content to this user.

The remainder of this section describes our concept in four parts: 1) rating content, 2) defining user profiles and similarity, 3) collecting initial ratings, and finally 4) recommending content to individual users.

2.1 Rating Content

Each content item has a slider (Figure 1) to rate it on a scale from 0 (not funny at all) to 10 (very funny). We chose this rating scale to capture detailed user ratings. Our algorithm considers every single one of these ratings.

2.2 Defining User Profiles and Similarity

In our application of a memory-based collaborative filtering approach, the profile of a specific user u is a vector containing all of u ’s ratings for all items.

The similarity of two users can then be computed as a function of their profile vectors. We further set a similarity threshold to define neighbors: If u ’s humor profile matches the profile of another user v with a minimum similarity of 50%, v will be promoted to a neighbor of u . The low threshold of 50% was chosen to enable recommendations even in a potentially smaller user base. It is possible to have multiple neighbours.

2.3 Collecting Initial Ratings

Collaborative filtering systems struggle with the “cold-start” problem: New users haven't rated any content yet. Hence, no neighbors and thus no recommendations can be computed. To address this issue, we define a fixed set of 30 items called “humor check”. All new users are shown this set to collect initial ratings. The choice of items in this “humor check” was informed by a pre-study, selecting the 30 items which had received the most controversial ratings (i.e. highest variance across users), and at least 5 ratings. Such controversial content seems especially suitable to quickly assess a user's general taste of humor and interests. The check does not contain videos to keep it fast to complete.

2.4 Recommended Content

Finally, we can recommend content to individual users. As a prerequisite for receiving recommendations, a user u needs to have at least one neighbor v , who rated at least one item with a score of 7 or higher, which u has not rated so far.

We assume a rating of 7 to be considerably high and thus to provide an adequate threshold for recommendations likely perceived as relevant by the receiving user.

3 Evaluation

3.1 Apparatus

We implement a custom humor website with appr. 2000 items, which presents content in an “infinite scrolling” view, similar to popular humor sites such as *9gag*². Infinite scrolling automatically appends new content to the page if the user scrolls close to the bottom.

Our backend system stores user accounts and ratings, and implements a user-based CF algorithm provided by the *vogoo*³ library for PHP. Therein, user similarity is based on the mean squared difference of users' ratings. The user profile is updated after each successful rating.

The frontend consists of a view for the “humor check” showing the 30 predefined items, and a recommendation view presenting content based on the users' humor profile and neighbors.

To measure if recommended content receives better ratings than randomly chosen content, we display both in equal shares. In particular, our infinite scroll view alternates between adding one recommended and one random item. The website advertises its customized humor recommendations, but shows no indicators whether specific items were recommended or not. Hence, users will believe that every item was recommended.

² 9gag website: <http://9gag.com>

³ Vogoo website: <http://sourceforge.net/projects/vogoo/>

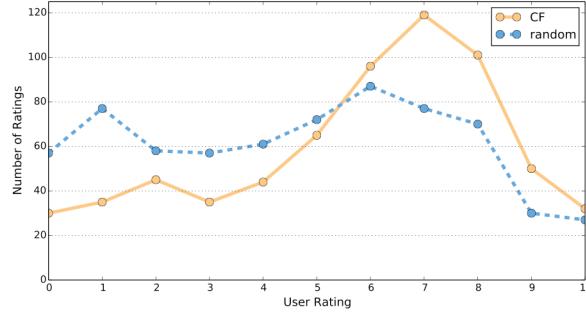


Figure 2: Number of user ratings for content recommended with our CF approach, and for randomly chosen content. This figure shows that individually recommended content received higher ratings than random content. The difference is statistically significant ($p < 0.001$).

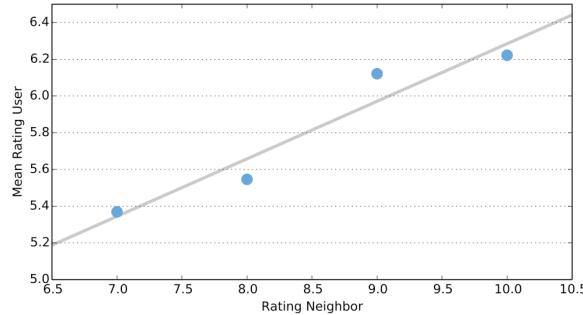


Figure 3: Mean user rating depending on the neighbor's rating (the four dots) with regression line. The correlation is significant ($p < 0.05$). These results show that our applied concept of comparing users based on their rating behavior can render users' taste or interest in humorous content more predictable.

For the purpose of this study, we log all ratings in the recommendation view. A rating implies that the user engaged with the content. Each record includes the user's id and rating. For recommended items, we also log id and rating of the neighbor on which the recommendation was based. The similarity between both users at the time of the rating is recorded as well.

3.2 Procedure

To evaluate whether individual humor can be assessed with our concept, we conducted a field study using the described humor website.

In total, 150 individual users registered on our website. Of those, 54 rated recommended content while others simply perceived the content. Participants were between 18 and 35 years old. They had to complete the “humor check” upon registration on our website. They were then forwarded to the main view, showing recommendations and randomly selected content in the infinite scrolling list. We logged users' ratings over a period of 4 weeks.

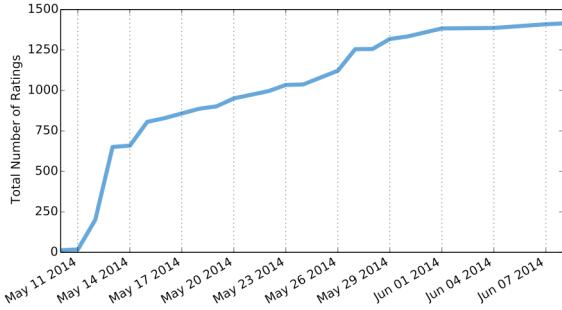


Figure 4: Number of user ratings over the course of the study. At the beginning, many users registered and rated after we had advertised the study. After this initial phase, many users still returned to view and rate recommended content throughout the month.

4 Results

Overall, we gathered 1325 records: 652 ratings of recommended content and 673 ratings of randomly selected items. We use the Shapiro-Wilk test to determine whether or not these ratings follow a normal distribution. Based on the result (not normally distributed), the Wilcoxon signed rank test is used to determine statistical significance.

4.1 Quality of Recommendations

The mean rating for all items is 5.18. The average rating for recommended items is 5.7, and 4.65 for randomly chosen content. Hence, our CF approach leads to a significantly higher average rating ($p < 0.001$). Figure 2 visually compares the distributions of user ratings.

Furthermore, Figure 3 shows the users' mean ratings for recommended content split by the neighbor's rating. These results show that users tend to rate content higher, if it also received high ratings from the neighboring user. The correlation is statistically significant ($r=0.96$, $p<0.05$). This supports the finding that our approach of modeling and comparing users based on their rating behavior leads to the desired outcome, namely suitable recommendations which better match the taste and interest of individual users.

4.2 Behavior Ratings over Time

Additionally, we examine rating behavior over time: Figure 4 shows the development of the number of user ratings collected over the course of the study (about one month). Ratings grow roughly linearly after an initial “kick-off”, following the advertisement of the study. Moreover, Figure 5 visualizes the distribution of rating times over the day. Here, we observe peaks, which could coincide with commuting times in the morning, lunch and coffee breaks, and relaxing in the evenings. Therefore, these results indicate that our website was used for micro-entertainment at several different times throughout the day.

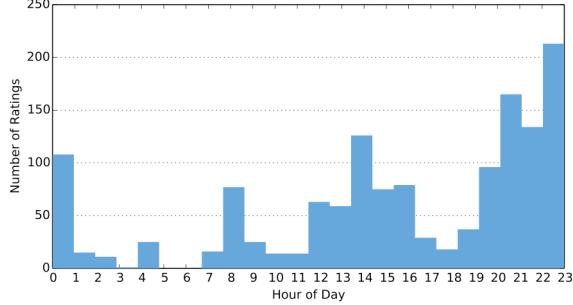


Figure 5: Number of user ratings over the day. This plot indicates that our website was used for micro-entertainment throughout the day: the observed peaks possibly coincide with commuting times in the morning, lunch and coffee breaks, and relaxing in the evenings.

4.3 Recommendation Graph

Each recommendation is based on a rating by the neighbor with the highest similarity. Therefore, we can analyze which item is recommended by whom to whom. Figure 6 visualizes these neighbor-to-user connections as a graph. Therein, each node represents one user and each edge between two nodes shows one recommendation.

Overall, we observe few nodes with many edges, and many nodes with a relatively small number of edges. This reveals two main groups of participants: “power-users”, who visit the website often, rating many items, and “passers-by”, who stop using it after a period of time. Although a popular humor site may have many more visits than our study website, we can expect similar types of user behavior for larger sites as well, leading to interesting networks. Future studies could further investigate these structures.

5 Limitations

The occurrence of humor and laughter is an individual phenomenon. It can occur in special situations and has several variations, like laughing out loud or only judging something as funny (Warren and McGraw 2014). We focus on user similarity in terms of humor as expressed by their content ratings, but we cannot assess their current situation or reactions.

Related work has analyzed different aspects of collaborative filtering algorithms, like prediction quality, performance, learning speed or the minimum size of datasets (Breese et al. 1998, Sarwar et al. 2000). The selection and weighting of neighbors is the main characteristic to ensure high prediction quality of a collaborative filtering algorithm (Bellogin et al. 2014). We did not rigorously optimize these aspects here. Our study shows that a CF algorithm is able to recommend humorous content, but we do not claim to present the best possible CF method in this paper.

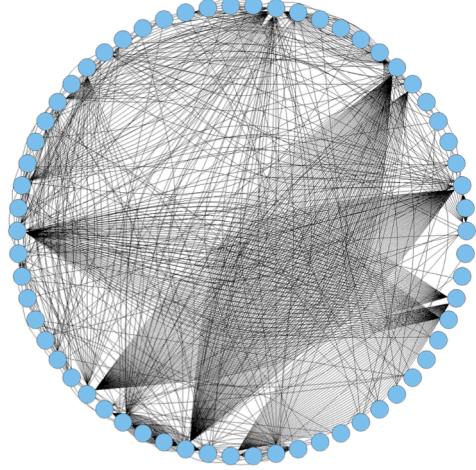


Figure 6: Recommendation graph: Each edge represents a successfully recommended content between two users (nodes), meaning that the recommendation was rated by the receiving user.}

Other rating systems exist and may have an impact on the resulting quality of recommendations. In particular, binary systems based on up and down votes are a popular choice on the web (e.g. used by *reddit*⁴, *9gag*, and *imgur*⁵).

6 Conclusion and Future Work

Many users today seek micro-entertainment content throughout their day, for example to bridge waiting times. Fitting content to the user's individual taste and interests can be expected to be more engaging and enjoyable, and may result in longer visits and higher return rates. However, humor is a complex and individual phenomenon, and highly depends on background knowledge and context, which renders humor prediction a challenging task.

In this paper, we have addressed this challenge with a collaborative filtering approach. Content predictions are based on user similarity derived from personal content ratings. In a field study with a humor website, we have explored the quality and acceptance of these recommendations. Our analyses of user ratings and the resulting recommendation graph show two main results: 1) some participants acted as “power-users”, making extensive use of our website and the rating system; and 2) users rated content recommended by our CF system significantly higher (+22.6%) than random content.

⁴ Reddit website: <http://www.reddit.com>

⁵ Imgur website: <http://imgur.com>

Future work could examine different variations and parameter settings of collaborative filtering algorithms. We also plan to investigate the influence of different rating-systems, binary up/down voting in particular.

References

- Alt, F., Kern, D., Schulte, F., Pfleging, B., Shirazi, A. & Schmidt, A. Enabling Micro-Entertainment in Vehicles based on Context Information. In *Proceedings of the Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI'10*, ACM (New York, NY, USA, 2010), 117–124.
- Alt, F., Shirazi, A., Schmidt, A. & Atterer, R. Bridging Waiting Times on Web Pages. In *Proceedings of the Fourteenth ACM SIGCHI's International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI'12, ACM (New York, NY, USA, 2012).
- Bellogin, A., Castells, P. & Cantador, I. Neighbor selection and weighting in user-based collaborative filtering: a performance prediction approach. *ACM Trans. on the Web (TWEB)* 8, 2 (2014), 12.
- Breese, J. S., Heckerman, D. & Kadie, C. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc. (1998), 43–52.
- Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L. & Riedl, J. GroupLens: applying collaborative filtering to Usenet news. *Communications of the ACM*, ACM 40, 3 (New York, NY, USA, 1997), 77–87.
- Liu, J., Dolan, P. & Pedersen, E. R. Personalized news recommendation based on click behavior. In *Proceedings of the 15th International Conference on Intelligent User Interfaces, IUI '10*, ACM (New York, NY, USA, 2010), 31–40.
- Miller, B. N., Albert, I., Lam, S. K., Konstan, J. A. & Riedl, J. MovieLens unplugged: Experiences with an occasionally connected recommender system. In *Proceedings of the 8th International Conference on Intelligent User Interfaces, IUI '03*, ACM (New York, NY, USA, 2003), 263–266.
- Sarwar, B., Karypis, G., Konstan, J. & Riedl, J. Analysis of recommendation algorithms for e-commerce. In *Proceedings of the 2nd Conference on Electronic commerce*, ACM (New York, NY, USA, 2000), 158–167.
- Veatch, T. C. *A theory of humor*. *Humor* 11 (1998), 161–215.
- Warren, C. & McGraw, P. Appreciation of Humor. In *Encyclopedia of Humor Studies*. SAGE (2014), 52–55.

Contact Information

Daniel Buschek, Florian Alt
 University of Munich (LMU) – LFE Medieninformatik
 Amalienstraße 17, 80333 München, Germany
 daniel.buschek@ifi.lmu.de, florian.alt@ifi.lmu.de