Emotion Centric Content Classification

Evelina Bakhturina
NYU Center for Data Science
e.bakhturina@nyu.edu

Michael Higgins
NYU Center for Data Science
mch529@nyu.edu

Matthew Dunn
NYU Center for Data Science
mtd368@nyu.edu

***Abstract*—Most recommendation systems fail to utilize the emotional response users have in relation to various pieces of content. These systems either assume content has an inherent emotional state independent of whoever is viewing it, or simply classify content in relation to clusters of users. These approaches fail to recognize the unique emotional relationships between users and content. We attempt to define a recommendation system that suggests content to users based on their specific emotional reaction to a specific piece of content. We then attempt to cluster users along this emotion dimension with the goal of creating a recommendation system that allows users to seek out an instance of content that will have a high probability of eliciting in them a specific emotional response. We discuss the progress of this work and lastly set forth areas of possible future work.**

***Index Terms*—Emotion, Content, Recommendation System**

I. Introduction

Current approaches to recommendation systems fail to utilize the emotional response users have in relation to various pieces of content [15]. Certain online platforms are recognizing the value of understanding their content along this dimension. Facebook has added emotion-based reactions to its timeline beyond the binary "Like" button. Nonetheless, most large online platforms assume content has an inherent emotional valence independent of whoever is viewing it.

The issue with this approach is that users of these systems are diverse and a single piece of content may elicit a wide variety of emotional responses from different users ([1], [2], [3]). A timely example is political content. A picture of Donald Trump may elicit positive feelings for one user while conversely cause very negative reactions in another. Large online platforms seek to mitigate this issue by collaboratively filtering content for various groups using the reactions of group members to determine which content will be acceptable to other group members.

Research has shown that current approaches to content selection often times lead to creating an echo chamber" for the user, devoid of opposing viewpoints [5]. As a result, this leaves users with a stream of content that is acceptable to them, generated by people with similar viewpoints, but that is unpredictable in terms of the emotional response it may or may not elicit in them [9]. Without an emotional dimension to content selection, users are therefore unable to seek out specific pieces of content that will provide a desired emotional experience.

To address this issue, research has been pursued that seeks to predict the emotional response that a certain piece of content will elicit from users. An example of this work focuses on classifying video content based strictly on the video [16], [17]. Further, there has been extensive research in *the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language*[6], but research in the area of how users emotionally relate to content holds great promise.

Based off work by [9], further studies have attempted to understand how image content impacts users emotional state [10]. This research is promising as it seeks to understand how users are being emotionally impacted by the content they consume on social media sites. Nonetheless, we believe these approaches to understanding content in relation to emotion leave much room for improvement - we seek to extend this research by building a recommendation system that classifies content in relation to the users anticipated emotional reaction.

II. Emotion Centric Recommendation System

One of the inherent challenges of building an emotion-centric recommendation system is that an emotion-centric data set or content classification system does not currently exist. As mentioned above, there have been steps in this direction made by Facebook, which introduced emotion-based reaction buttons in users' posts, and by Netflix, which suggests additional content based on users' previous viewing patterns.

Nonetheless, neither approach is able to understand how a person reacts to a specific piece of content and then allows users to access content along that dimension. As a result, we set out to build a proof of concept that would allow us to capture data on users' reactions to a piece of content, while also allowing users to seek content along a requested emotional dimension. The content returned via our recommendation system would have a high probability of eliciting the emotional response that users sought.

A. User Experience and Algorithm Design

The first step in this direction was to build a system that would allow users to view videos while capturing their emotional reactions. A significant learning process presented by this project was the challenge of designing a recommendation algorithm based on user data while actively designing the user interface, given that the data available for our recommendation algorithms is dependent on how we design the user interface. Specifically, what we confronted was an endless conditional feedback loop, in which a seemingly trivial user interface design decision resulted in a significant shift in our ability to implement one algorithm versus another to generate recommendations for users. Conversely, when we considered one recommendation algorithm, we would find that the choice of algorithm may impact the design of the user interface.

The endless combinations of user interface designs and recommendation algorithms led us to approach the problem incrementally. With so many unknowns about how users will interact with the system and how one recommendation algorithm will perform over another, we sought to get the system in front of users as quickly as possible so that we could begin iterating over the design of the user interface and recommendation algorithms.

One illustrative instance of this challenge was the decision to give to users the option to select the next video they would watch from a selection of the top six recommended videos, which the recommendation algorithm had selected, as shown in Figure 1. The challenge with design decision was that if we provided users with the option to select a video, then the top video proposed by the recommendation algorithm may not be selected by the user, and consequently the learning rate of the recommendation algorithm would be hindered. On the other hand, this approach struck a good balance between training the algorithm (show videos that are most likely to improve our model) and creating an enjoyable user experience.

B. Content Selection

Another challenge similar to the User Experience and Algorithm Design problem was the selection of content to initially populate the website. In order to increase the number of data points we had on users more quickly, one decision we made was to initially only use very short pieces of content. Most videos were under one minute in length, with the intention being that users would then watch a greater number of videos in a single sitting and that consequently our recommendation algorithm would be able to learn more quickly.

A second issue that came up when initially populating the website with videos was the need to select only videos that were not offensive to users. As a result, we had multiple team discussions during which we would review questionable pieces of content and decide what types of content had a very low likelihood of offending users. As a result, we eliminated a large number of videos that could have been considered even remotely offensive by users.

III. Solution Stack

To build an emotion centric content classification system, we had to engineer the prototype from the ground up. Fortunately, we were able to find example code from the Affectiva Developer portal[1]. The example open-source code by Affectiva was a demo using Javascript and Boostrap that allowed users to select content from the Youtube API in order to watch and observe their emotional reaction to the videos. We extended this code in several significant dimensions.

[1]http://www.affectiva.com/

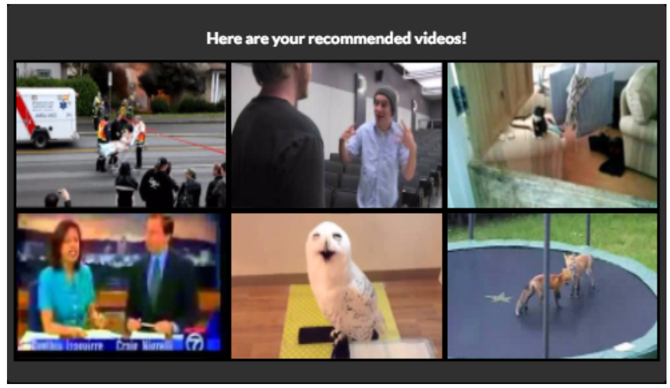


Fig. 1. User Recommended Videos Selection Box

First we set up a web-server using web framework Flask[2] and HTTP server NGINX[3] so that we could request user email addresses, control the videos served, and adjust the behavior of the application to allow users to watch multiple videos in a short period of time. Additionally, we built out a framework to record users' emotional responses from the Affectiva Emotion API as a JSON entry in the NoSQL Cloudant[4] database. Figure 2 details the components of our solution stack. The final working system can be found at *www.conten.me*.

With the emotional responses saved in Cloudant we were then able to run queries on these emotional responses using python scripts. The python scripts handled modeling and generating recommended YouTube videos for individual users, which would in-turn be loaded via the Youtube API and presented to the user. Lastly, it should be noted, we deployed both a production and development instance of the website with corresponding databases so that we could continue to develop and test new iterations of the solution as we collected data from users.

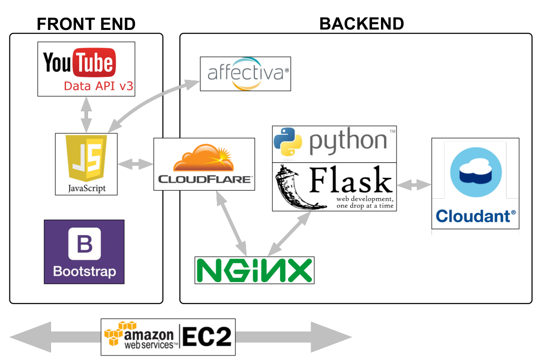


Fig. 2. Solution Stack for Prototype

A. Database

From the outset, a challenge of designing and building a completely novel solution to capture users' emotional reactions

[2]http://flask.pocoo.org/
[3]https://www.nginx.com/
[4]https://cloudant.com/

to content would require an iterative approach. As a result we chose to utilize a NoSQL database solution provided by Cloudant. By utilizing a NoSQL solution database we were able to add new fields on the fly without having to restructure the database relationships or re-index, as we would have needed to do using a relational database solution such as SQL.

The choice of a NoSQL database proved to be correct due to the fact that at one point in the project, while working on the recommendation algorithm using graph impact analysis, we found that we needed to add a field to the user database to mark users whose recommendations list needed an update. We were able to accommodate this minor change quickly and with minimal downtime.

1) content - stores content and related meta-data:
   - video id, view count, title and description etc.
2) users - stores user related information:
   - user id, recommendations, up-to-date recommendation flag, view ids
3) views - stores information about the user's response to content:
   - estimated age, sex, ethnicity; emotions: joy, angry, engagement, surprise; emojis: relaxed, smirk, laughing.
4) useritem - stores the relationship between a user and a piece of content
   - user id, video id

IV. Model

A. Impact Analysis

Active learning is a special case of semi-supervised learning in which an algorithm is able to interactively query the user to obtain the desired outputs at new data points [18], [19]. In our case the active learning algorithm chooses which videos to present to which users. Our label is the measured emotional response using Affectivas Emotion API. By intelligently choosing which content to show a given user(s) we can drastically cut down on the amount of data required to build an accurate model. Because the data we get back from Affectiva API is so rich (10 emotions, estimated age, sex, facial expressions, emoticons every second ...) our "target emotional response" is undefined.

For instance when trying to measure the extent to which a user found a video humorous, we could use any combination of happiness, laughter, joy and engagement. Which combination to use, or even if there is a combination that is universally applicable to all users, is unclear. Furthermore, because the users will eventually be able to choose which emotional state they would like to experience, there is no single target.

Because of these difficulties we chose to implement an active learning strategy called impact analysis that it is ambivalent to the choice of the optimization target or algorithm. The key insight behind the algorithm is that in order to make a recommendation for a user there must be a four-node path (that goes through the user in question) in the bipartite user-video graph - the more of these four-node paths there are the more accurate recommendations you will be able to make.

As a simple example consider the case of two users, Alice and Bob, who have been presented a disjoint set of content (Figure 3). By showing Alice a single piece of content that Bob has seen, we are able to improve our estimates of Alice's preference for all of the other videos that Bob has seen but Alice has not and vice versa for Bob. Impact analysis seeks to maximize the number of these four node paths.

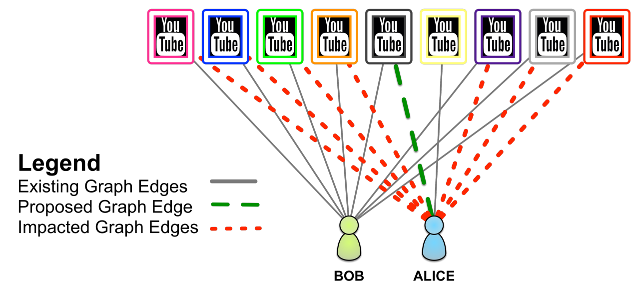


Fig. 3. Impact Graph Example

We implemented a slight variant to the algorithm whereby the top six most impactful recommendations are given to the user as opposed to the user being forced to watch any particular video. This strategy naturally led to a diverse set of user-item pairs that will allow us to make useful predictions regardless of the choice of the target function and is shown in Figure 4 where we visualize user content pairs.

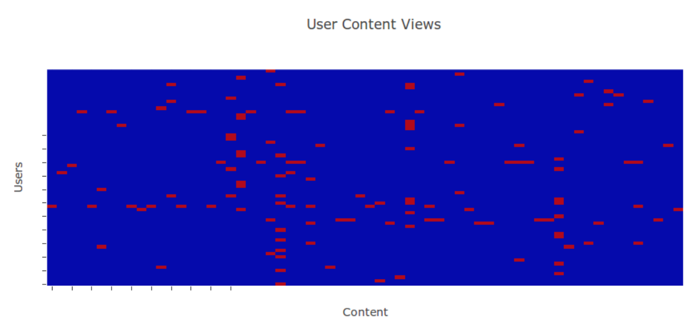


Fig. 4. User Item Views

V. Results

The data collected during the development of the prototype seems to confirm our belief that every user is unique and the emotional response to a particular piece of content can differ drastically from one user to another. Concrete evidence of this is shown in Appendix 1, which illustrates how the same piece of content can elicit joy in some and contempt in others, thus confirming our underlying hypothesis that individual users may have dramatically different responses to the exact same piece of content. Though the user base is small, the results indicate that utilizing the emotional response of the users could significantly impact an individual user's experience.

To further investigate our results data we applied t-distributed stochastic neighbor embedding (t-SNE) to reduce dimensionality of our data. Appendix 2 demonstrates that the gathered user-content data let us classify the video content based on the most strongly elicited emotion by a particular video. As Appendix 2 shows, the videos that elicit similar reactions are clustered together: for example, videos that evoke surprise or contempt are clustered together.

An interesting point of note is that many of the videos we selected to insert in the system were of a joyful or funny nature and that many users experienced this emotion when viewing the videos. This is interesting for two reasons. First, that we were able to effectively elicit a targeted emotion from our users, and, second, that inside the emotion ”joy,” users may be experiencing various sub-emotions. This suggests that future work may want to explore how to deconstruct the high-level emotion of ”joy” into smaller sub-categories. It may be interesting to understand how a video classified as 'joyful' can elicit a variety of feelings. For instance, some people are interested in watching a joyful video which also has an element of surprise, while others may enjoy joyful and silly videos. All these fine-grained emotions should be a part of the final recommendation system.

Unfortunately, our project did not achieve a large enough data set to answer some of these questions. As shown in Table 1, we were relatively unsuccessful in securing a large set of users. Many of our conclusions are therefore tentative. Nonetheless, our design approach to maximize the number of views that a single user has was successful, as can be seen by the ratio of Number of Users to Views. We can see that each User watched a large number of videos and this was in part likely due to our choice of very short videos, which allowed users to watch a large number of videos in a short period of time.

Statistic	**Total Count**	**Production Mode**
Number of Users	84	41
Views	677	191
Video Content	68	100

TABLE I
STATISTICS

VI. Future Work

A. Addressing Bandwidth Limitations

A current limitation to our approach is the inherent bandwidth limitations needed to send user videos to Affectiva's Emotion API. A possible area of research would be to build distributions for each type of content so that the system would only send users' facial data captured at the points of the content viewing with the highest probability of eliciting an emotional response from users. An alternative approach to addressing this issue is simply to handle the emotion classification from user videos locally by deploying a trained model directly onto mobile devices. Work has started in this area but further research is required [20].

B. Addressing Privacy Concerns

Several of our users expressed concern about being filmed and about how their images and data would be used. As an attempt to alleviate their hesitancy we added a notification at the beginning of every video reassuring the user that we would not be storing any images of the user but would store only their emotional responses in order to give them better recommendations. Paradoxically a better solution could be to make the users' data more open. It was very entertaining for users to be able to watch their emotional reactions in real time; in fact from a user experience perspective this was the only novel aspect of the site. By showing them how other users emotionally responded to a piece of content (using interactive visualizations) we can create a fun and unique experience as well as habituate users to the idea of being recorded.

C. Mobile Prototype

Lastly, another area for future work would be to build the prototype on an Android or iOS mobile device. Initially, we did not build a mobile prototype and, as a result, we were unable to secure the necessary viewership required to generate enough data that would allow us to execute meaningful analysis of our model. We intentionally chose to build a desktop version of the project because we wanted to minimize the technical challenges associated with launching a minimum viable product. In hindsight, we should have anticipated the challenge of asking users to watch only from a desktop device when the vast majority of content is accessed via mobile devices.

VII. Conclusion

An emotion-based content classification system could significantly improve the current state of the art recommendation systems by increasing the dimensionality used to build predictions for users. Designing, building, and deploying an emotion-centric classification system is not trivial and requires significant design and technical resources. We overcame many of these challenges on a short timeline. Future work should focus on experimenting with our general framework for active learning and utilize a mobile first approach. The capacity of algorithms to recommend content to users will be greatly enhanced by an emotion-centric approach and future work should continue to explore the designing and building of recommendation systems using this information.

References

[1] Gonzalez, Gustavo, et al. ”Embedding emotional context in recommender systems.” Data Engineering Workshop, 2007 IEEE 23rd International Conference on. IEEE, 2007.

[2] Tkalcic, Marko, Giovanni Semeraro, and Marco de Gemmis. ”Personality and Emotions in Decision Making and Recommender Systems.” DMRS. 2014.

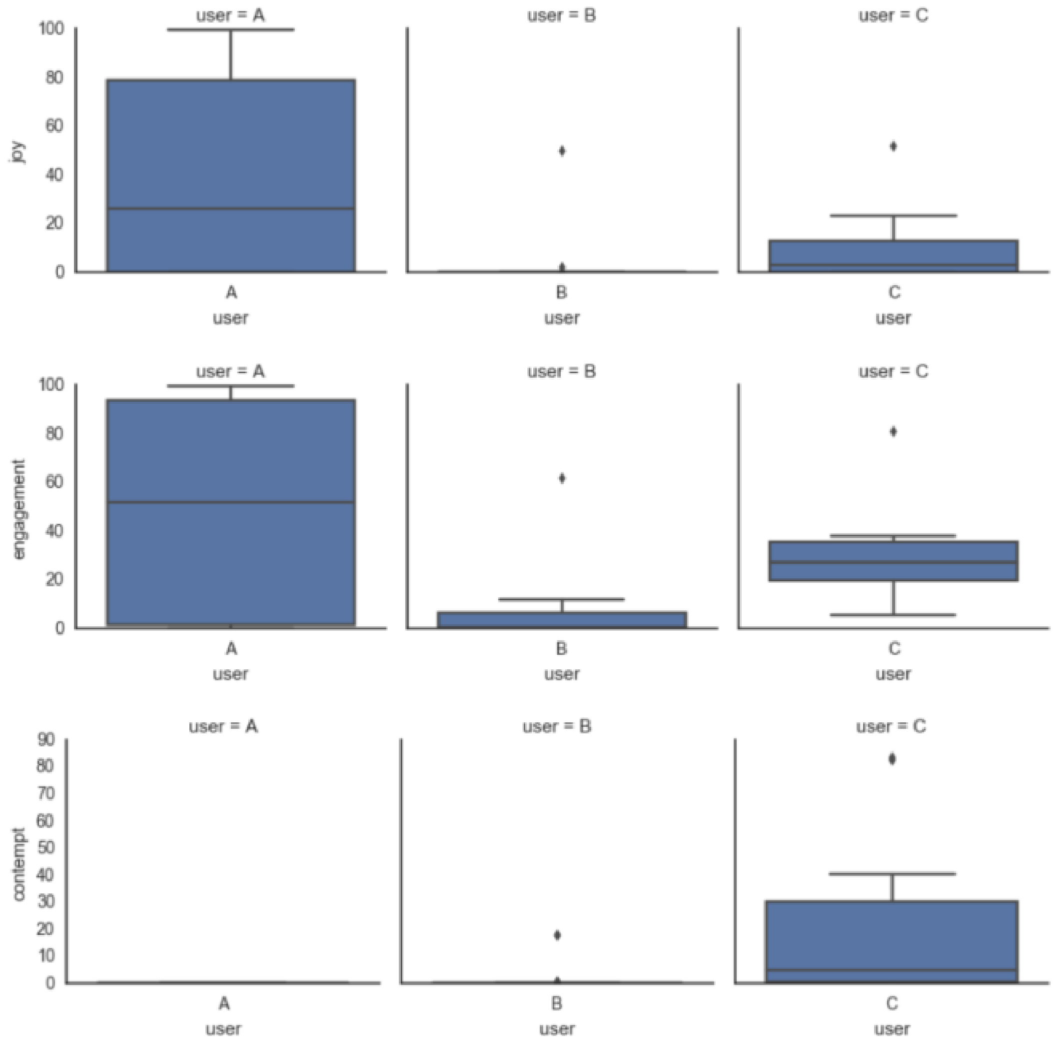
[3] Polignano, Marco. ”A framework for emotion-aware recommender systems supporting decision making.” Proceedings of the 6th Symposium on Future Directions in Information Access. British Computer Society, 2015.

[4] Hill, Ronald P., and Michael B. Mazis. ”Measuring emotional responses to advertising.” NA-Advances in Consumer Research Volume 13 (1986).

[5] Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. ”Exposure to ideologically diverse news and opinion on Facebook.” Science 348.6239 (2015): 1130-1132.
[6] Liu, Bing. ”Sentiment analysis and opinion mining.” Synthesis lectures on human language technologies 5.1 (2012): 1-167.
[7] Socher, Richard, et al. ”Recursive deep models for semantic compositionality over a sentiment treebank.” Proceedings of the conference on empirical methods in natural language processing (EMNLP). Vol. 1631. 2013.
[8] Cambria, Erik, et al. ”New avenues in opinion mining and sentiment analysis.” IEEE Intelligent Systems 28.2 (2013): 15-21.
[9] Kramer, Adam DI. ”The spread of emotion via Facebook.” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2012.
[10] Wang, Xiaohui, et al. ”Modeling emotion influence in image social networks.” IEEE Transactions on Affective Computing 6.3 (2015): 286-297.
[11] Yang, Yang, et al. ”Social Role-Aware Emotion Contagion in Image Social Networks.” Thirtieth AAAI Conference on Artificial Intelligence. 2016.
[12] Soleymani, Mohammad, Maja Pantic, and Thierry Pun. ”Multimodal emotion recognition in response to videos.” IEEE transactions on affective computing 3.2 (2012): 211-223.
[13] Berger, Jonah, and Katherine L. Milkman. ”What makes online content viral?.” Journal of marketing research 49.2 (2012): 192-205.
[14] Sharma, Ankur Satyendrakumar, and Mohamed Elidrisi. ”Classification of multi-media content (videos on youtube) using tags and focal points.” Unpublished manuscript (2008): 152-170.
[15] Zheng, Yong, Bamshad Mobasher, and Robin D. Burke. ”The Role of Emotions in Context-aware Recommendation.” Decisions@ RecSys 2013 (2013): 21-28.
[16] Ramalingam, Soodamani. ”Metadata Extraction and Classification of YouTube Videos Using Sentiment Analysis.” IEEE International Carnahan Conference on Security Technology. IEEE 978-1-5090, 2016.
[17] Filippova, Katja, and Keith B. Hall. ”Improved video categorization from text metadata and user comments.” Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. ACM, 2011.
[18] Rubens, Neil, et al. ”Active learning in recommender systems.” Recommender Systems Handbook. Springer US, 2015. 809-846.
[19] Settles, Burr. ”Active learning literature survey.” University of Wisconsin, Madison 52.55-66 (2010): 11.
[20] Kim, Yong-Deok, et al. ”Compression of deep convolutional neural networks for fast and low power mobile applications.” arXiv preprint arXiv:1511.06530 (2015).

Appendix

Appendix 1: Reaction of Different Users to a single Piece of Content



Appendix 2: Video Classification Based on Users' Emotional Response