A Mood-based Genre Classification of Television Content

Humberto Jesús Corona Pampín
Insight Centre for Data Analytics
University College Dublin, Ireland
humberto.corona@insight-centre.org

Michael P. O'Mahony
Insight Centre for Data Analytics
University College Dublin, Ireland
michael.omahony@insight-centre.org

ABSTRACT

The classification of television content helps users organise and navigate through the large list of channels and programs now available and enables personalised program recommendations. In this paper, we address the problem of television content classification by exploiting text information extracted from program transcriptions. We present an analysis which adapts a model for sentiment that has been widely and successfully applied in other fields such as music or blog posts. We use a real-world dataset obtained from the Boxfish API to compare the performance of classifiers trained on a number of different feature sets. Our experiments show that, over a large collection of television programs, genres can be represented in a three-dimensional space of valence, arousal and dominance, and that promising classification results can be achieved using features based on this representation. This motivates the use of the proposed representation of television content as a feature space for similarity computation and recommendation generation.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information Filtering

General Terms

Measurement, Experimentation

Keywords

Mood analysis, Text classification, Genre classification

1. INTRODUCTION

The problem of choice overload in television is well known. For example, in 2009 there were circa 2218 television broadcast stations in the U.S.[1], making it very difficult for users to manually organise, browse or decide what content is relevant or best suited for them. Thus, there is a need for new ways to classify television content that can be applied by intelligent systems to enable content discovery.

Most of the research on television content classification is based on audio and visual features, focusing on genre classification [3, 6] and on the relationship between content and industry, audience, and culture [8]. However, in this paper we focus on the analysis of text extracted from television program scripts for genre classification. We use metadata obtained from the Boxfish API[2] to build a textual representation of television program and channel content. This allows us to explore content in a three-dimensional space of affect defined by valence, arousal and dominance. To this end, we follow the approach presented in [2] and expand it using the arousal and dominance dimensions applying it to the television domain, and more specifically to the genre classification task.

The remainder of this paper is organised as follows. First, Section 2 describes the related work. Then, Section 3 introduces the datasets used in this work. Section 4 describes the a feature analysis of a year of television content, and showcases and example of mood analysis for a news channel. Section 5 introduces the classification approach studied in this paper, and discusses the results. Finally, in Section 6, conclusions and future work are presented.

2. RELATED WORK

The moods associated with multimedia content such as television programs or songs are difficult to infer: people perceive them differently [12] and they are culture dependent [7]. For example, songs like *Bohemian Rhapsody* by *Queen* or movies like *Forest Gump* are non-trivial to classify.

Several ontologies for mood classification rely on models developed in the psychology field, *Rusells model of affect* [9] being one of the most widely used. This model represents each mood into a two-dimensional space defined by valence v (which measures the good–bad dimension of sentiment) and arousal a (which measures the active–passive dimension of sentiment). Thus, each mood $m \in M = \{v, a\}$ can be represented by a vector in this two-dimensional space. The model is based on the evidence that the affective dimensions are built in a *highly systematic fashion*, instead of being independent dimensions.

Dodds et al. [2] uses features extracted from lyrics and the Affective Norms for English Words (ANEW) dataset [1]

[1]List of countries by number of television broadcast stations (19/12/2013). In Wikipedia, The Free Encyclopaedia.

[2]http://boxfish.com/api

ACM Workshop on Recommendation Systems for Television and Online Video '2014 Foster City, California USA

to measure the average happiness (valence) of songs, blogs and State of the Union presidential speeches. The aim of the work is to quantify the evolution of the overall happiness in different contexts. The approach calculates the average valence of each document (song, blog post or speech) by counting the number of times each of the terms in the ANEW dataset appears in the document, and multiplying it by its associated mean valence value. The results show that, for example, valence can help distinguish between music genres, when a large number of songs are considered, and that interesting trends in presidential speeches are revealed.

Eggnik et al. [3] perform a large scale experiment for mood classification of television programs from the BBC channel using a live user study. Participants were asked to watch short clips from television programs and assign mood labels. The results obtained showed that there was consensus on the mood labels applied, and an automatic classification based on the data obtain a 90% accuracy for certain programs. Moreover, the study performed a principal component analysis, finding two main components in the mood of television content: one related to the seriousness of the program and the second related to the perceived pace.

A mood-based similarity metric to exploit movie mood similarities for context-aware recommendations is presented in [10]. Here, the proposed metric is used in a joint matrix factorisation model, obtaining results that lead to better recommendations than other evaluated mood-based movie similarities (in the context of mood-based recommendations).

From the related work it is clear that the analysis of television content using a multidimensional mood space is an interesting problem. Thus, in this paper we consider a text-based classification of television content using features based on the dimensions that define Rusell's model, and also consider the dominance (or control) dimension. Moreover, we follow the approach proposed in [2] expanding the study to the valence and dominance dimensions, (in line with [5]).

3. DATASETS

The *ANEW* dataset [1] is a collection of 2,476 words annotated with emotional ratings in three dimensions — *valence*, *arousal* and *dominance*. The dataset was created using human assessment, and it aims to provide a *set of normative emotional ratings* for the words included. For each dimension, the dataset contains the mean and standard deviation of the ratings values obtained for each word. Here, we normalise the original scale from $[1-9]$ to $[0-1]$.

All the television content information was obtained through the Boxfish API, which provides the electronic programming guide for each channel, text entities based on program transcriptions and a genre division on a program by program level. We describe the data obtained in detail below.

- We selected eight different channels which capture a broad range of programs and genres; news (FOX News, CNN, MSNBC), general entertainment (FOX, E!), Science Fiction (SyFy), educational (Discovery Channel) and children (Cartoon Network). For each channel, we obtain the electronic programming guide (EPG), which contains the program schedule for each of the selected channels, including program title, showtime and program genre.
- The Boxfish API provides a list of approximately 45 different genres, from which we selected a representative selection, including: *reality, documentary, animated, newscast* and *horror.*
- The entities derived from program transcriptions were also obtained using the API. For each item (program or channel) considered, we queried the total number of occurrences of each entity over a period of time. On average, there were 283.2 and 1,569.4 distinct entities per program and channel, respectively. Overall, 2,034 distinct entities were obtained for all programs and channels considered, out of the 2,476 included in the ANEW dataset.

4. FEATURE ANALYSIS

In this section, we analyse the content of different television channels and their relationship with the valence, arousal and dominance dimensions. We use the approach proposed in [2] to calculate the average valence, and we expand it considering also the arousal and dominance dimensions. Thus, a television program is defined by its mood $m \in M = \{v, a, d\}$ in this three-dimensional valence-arousal-dominance space. With the proposed approach we expect to understand how these dimensions can be used to distinguish television content considering different granularities.

4.1 Methodology

We perform the analysis over a year of television content, from January to December 2013. We obtain the count of ANEW terms for each week and for each of the selected channels. The average valence (arousal and dominance) values are calculated by multiplying the number of times a term occurs by its associated valence (arousal and dominance) value in the ANEW dataset, as in previous work [2].

4.2 Results

Figure 1 presents, for each channel, the average valence, arousal and dominance values for each of the thirteen periods considered (each period lasting four weeks[3]). From the results we can infer that the valence and dominance dimensions, in particular, can potentially help to classify television content by channel (at least for the channels considered here). For example, compared to other channels, *E! Entretaniment* has on average a much higher valence (~ 0.684) — it can be considered a very happy channel — and dominance (~ 0.608), which correlates with the nature of this channel's content (mainly focused on general entertainment and reality television). Moreover, all the news channels are clustered together in this space, showing the lowest average valence (~ 0.639) and lowest average dominance (~ 0.577) values. These values also appear to be well correlated to the typical content of news channels and the language used. Finally, it can be seen that the rest of the entertainment channels are also clustered together in the space.

It is also important to analyse the deviation over the mean values shown in Table 1. For example, in the *Cartoon Network (CN)* channel, the average standard deviation of valence over the 52 weeks is 0.179. This relatively high variation is due to the fact that the channel broadcasts a wide range of shows, intended for different audiences groups. For

[3]The clustering over periods of four weeks is done for clarity in the representation. Moreover, these periods correspond to circa a month of television content.

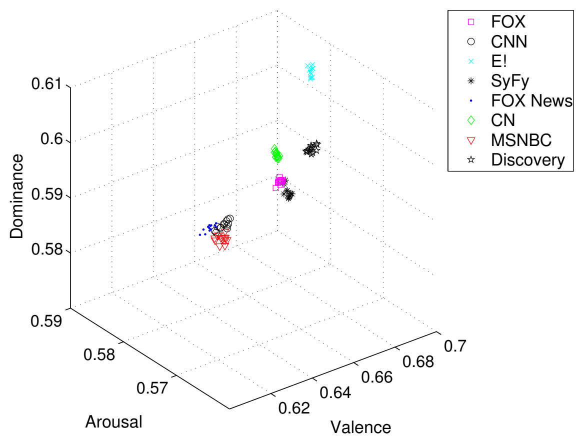

Figure 1: Valence, arousal and dominance for a year of television broadcasting per channel.

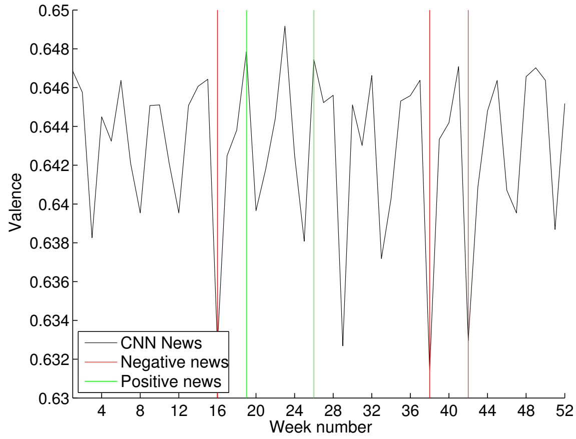

Figure 2: Valence per week of *CNN* content in 2013.

example, a particular episode of *The Amazing World of Gum Ball* has an average valence of ~ 0.61, which is lower than average for this channel. However, this is understandable as this show contains some references to sex and nudity, violence and profanity[4]. On the other hand, an episode of *Grobjad*, which contains no references to mature content[5], has an average valence of ~ 0.69. Moreover, there is no intersection between the top-6 most similar programs to each of the two examples considered, based on data obtained from IMDB. Thus, while the average valence, arousal and dominance values computed over all programs broadcast by a channel appear to be discriminative, classifying individual programs by channel (i.e. by genre) may be problematic.

Channel	Valence	Arousal	Dominance
E!	0.684 (0.168)	0.577 (0.098)	0.608 (0.100)
Discovery	0.665 (0.169)	0.570 (0.098)	0.601 (0.103)
CN	0.663 (0.179)	0.575 (0.101)	0.597 (0.105)
SyFy	0.654 (0.175)	0.570 (0.099)	0.595 (0.108)
FOX	0.658 (0.171)	0.573 (0.009)	0.594 (0.105)
FOX News	0.639 (0.177)	0.579 (0.097)	0.585 (0.109)
CNN	0.643 (0.176)	0.577 (0.098)	0.586 (0.108)
MSNBC	0.635 (0.177)	0.575 (0.098)	0.586 (0.110)

Table 1: Valence, arousal and dominance mean and standard deviation (in parentheses) values per television channel.

Finally, in Figure 2, we present a valence analysis of CNN news channel content by week over one year (2013). The vertical markers highlight some of the top news events[6] from the year. Most of the events are clearly correlated with changes in the channel content valence. For example, low valence is evident for very negative events such as the Boston Marathon bombings (week 16), the Navy Yard Shooting (week 38) or the collapse of Obamacare (week 42). We also found valence correlations with positive news events such as the liberation of three women kidnaped in Ohio (week 19) and the Supreme Court ruling on the DOMA and California Marriage Equality Bill (week 26), all of which correspond to content with relatively high valence values.

The results presented in this analysis show that the valence and dominance dimensions, in particular, can be used to distinguish between the content broadcast by different channels. However, as mentioned previously, classifying the genre of individual television programs (for use in a recommender system) based on these dimensions may present challenges (given the variance observed within genres), an analysis of which is considered in the next section.

5. PROGRAM GENRE CLASSIFICATION

In this section, we analyse the performance of a single-label supervised classification approach for television genre classification using the selected genre categories derived from the Boxfish API (Section 3). We use an instance-based representation of each television program based on statistical features derived from the valence, arousal and dominance dimensions described above. In particular, we consider an early-fusion ensemble approach [11] in which all these meta-features are combined into a single feature space. We compare this approach against a standard *vector space model (VSM)* (based on derived entities) approach.

5.1 Classification Approach

Feature-based instances for each television program are created as follows. First, we select all the television programs from each of the genres and channels described in Section 3. For each program, we extract the valence values (likewise mean arousal and dominance) of all entity terms contained in the ANEW dataset. Thus, for each dimension, the the following attributes are computed: *minimum value, maximum value, mean, standard deviation* and *median*. We also include a set of simple word count features: *number of total words* and *number of unique ANEW words.*

5.2 Experimental Methodology

In total we obtain 343 television programs, queried over the period of two weeks in February 2014. The program genres of instances are distributed as follows: *animated (120)*, *documentary (65)*, *horror (24)*, *newscast (41)* and *reality (93)*. The classification was performed using the Weka machine learning framework [4]. In each case, a standard 5-fold

[4]http://www.imdb.com/title/tt1942683/parentalguide
[5]http://www.imdb.com/title/tt2406986/parentalguide
[6]Obtained from the website http://www.infoplease.com.

cross validation approach was used to evaluate performance, expressed in terms of precision and recall over each class. For the meta-features based classification we used a *random forest* classifier with 100 trees and 10 features, as in preliminary experiments this configuration proved to make the best use of the very restricted feature space available. For the VSM approach, we used a Naïve Bayes classifier and *term frequency* weighting as this approach proved to be the best performing classifier in our previous experiments.

5.3 Results

The precision and recall performance provided by each classifier is shown in Table 2. In general, VSM performed better than the meta-features based approach. The exception to this trend was the *newscast* genre, where the meta-features classifier obtained a precision of 0.853 compared to 0.826 for VSM, albeit with reduced recall. This finding is inline with that shown in Figure 2, where the valence dimension proved to be discriminate well between news content.

Genre	**Classifier**	
	VSM (Precision - Recall)	Metafeatures (Precision - Recall)
Animated	1.000 - 0.925	0.792 - 0.825
Documentary	0.925 - 0.75	0.623 - 0.508
Horror	0.792 - 0.792	0.522 - 0.500
Newscast	0.826 - 0.927	0.853 - 0.707
Reality	0.800 - 0.903	0.685 - 0.796
Average	**0.896 - 0.889**	**0.719 - 0.720**

Table 2: Genre classification performance.

There are differences in performance among the different genres. For example, VSM provided poorest performance for the *reality* genre, something expected due to the broad and informal language associated with this particular genre. On the other hand, the precision obtained for genre *animated* using this approach is perfect, indicating a distinctive vocabulary is associated with this genre.

Finally, while generally outperformed by VSM, the meta-features based approach shows promising results and offers a new representation for this type of data. In particular, the meta-features classification results are different in terms of their relative ordering (by genre) according to accuracy, indicating that the two approaches work on different aspects of the data. Thus, an ensemble technique, combining both types of features, may provide enhanced performance; an analysis of which is left to future work.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a study of television content classification, relying solely on textual features as a source of information. From the feature analysis described in Section 4, it is clear that television content, at least at a high (i.e. channel) level, can be discriminated by the proposed three-dimensional space of affect. While classifying the genres of individual television programs using this approach in general did not outperform a traditional VSM based classifier, nevertheless there is evidence to suggest that meta-features based on valence, arousal and dominance values have the potential to contribute to enhanced classification performance, particularly if used in combination with other feature types. Moreover, we note that the meta-features used in this work were based on ANEW values computed over entities derived from the Boxfish API; since not all of these entities were present in the ANEW dataset (only 82%), better performance may be achieved using the full program transcription text. In future work, we will also consider using the model of affect in a personalised content-based recommendation approach, as well as conducting live user studies to understand how individuals respond to mood-based recommendations.

7. ACKNOWLEDGEMENTS

This work is supported by Science Foundation Ireland through the CLARITY Centre for Sensor Web Technologies under grant number 07/CE/I1147 and through the Insight Centre for Data Analytics under grant SFI/12/RC/2289. We gratefully acknowledge Boxfish for providing access and support to their API.

8. REFERENCES

[1] Bradley, M. M., and Lang, P. J. Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings.

[2] Dodds, P. S., and Danforth, C. M. Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents. *Journal of Happiness Studies 11*, 4 (July 2009), 441–456.

[3] Eggink, J., and Bland, D. A Large Scale Experiment for Mood-Based Classification of TV Programmes. *2012 IEEE International Conference on Multimedia and Expo* (July 2012), 140–145.

[4] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. The WEKA data mining software, 2009.

[5] Hu, X., and Downie, J. S. Improving mood classification in music digital libraries by combining lyrics and audio. *Proceedings of the 10th annual joint conference on Digital libraries - JCDL '10* (2010), 159.

[6] Jasinschi, R. S., and Louie, J. Automatic TV program genre classification based on audio patterns. In *Conference Proceedings of the EUROMICRO* (2001), pp. 370–375.

[7] Kosta, K., Song, Y., Fazekas, G., and Sandler, M. A Study of Cultural Dependence of Perceived Mood in Greek Music. *ppgia.pucpr.br* (2013).

[8] Mittell, J. A Cultural Approach to Television Genre Theory. *Cinema Journal 40*, 3 (2001), 3–24.

[9] Russell, J. A. A circumplex model of affect., 1980.

[10] Shi, Y., Larson, M., and Hanjalic, A. Mining mood-specific movie similarity with matrix factorization for context-aware recommendation. *Proceedings of the Workshop on Context-Aware Movie Recommendation - CAMRa '10* (2010), 34–40.

[11] Snoek, C. G. M., Worring, M., and Smeulders, A. W. M. Early versus late fusion in semantic video analysis. *Proceedings of the 13th annual ACM international conference on Multimedia - MULTIMEDIA '05* (2005), 399.

[12] Song, Y., Dixon, S., Pearce, M., and Halpern, A. Do Online Social Tags Predict Percieved or Induced Emotial Responses to Music? *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR 2013)* (2013).