

Know Your Data: Understanding Implicit Usage versus Explicit Action in Video Content Classification

Jude Yew^a and David A. Shamma^a

^aYahoo! Research, 4301 Great America Parkway, Santa Clara, USA;

ABSTRACT

In this paper, we present a method for video category classification using only social metadata from websites like YouTube. In place of content analysis, we utilize communicative and social contexts surrounding videos as a means to determine a categorical genre, e.g. Comedy, Music. We hypothesize that video clips belonging to different genre categories would have distinct signatures and patterns that are reflected in their collected metadata. In particular, we define and describe social metadata as *usage* or *action* to aid in classification. We trained a Naive Bayes classifier to predict categories from a sample of 1,740 YouTube videos representing the top five genre categories. Using just a small number of the available metadata features, we compare the classifications produced by our Naive Bayes classifier with those provided by the uploader of that particular video. Compared to random predictions with the YouTube data (21% accurate), our classifier attained a mediocre 33% accuracy in predicting video genres. However, we found that the accuracy of our classifier significantly improves by *nominal factoring* of the explicit data features. By factoring the ratings of the videos in the dataset, the classifier was able to accurately predict the genres of 75% of the videos. We argue that the patterns of social activity found in the metadata are not just meaningful in their own right, but are indicative of the meaning of the shared video content. The results presented by this project represents a first step in investigating the potential meaning and significance of social metadata and its relation to the media experience.

Keywords: Video, Social, Categorization, YouTube, Naive Bayes, Classification

1. INTRODUCTION

Socially sharing videos has become a very popular activity on websites like YouTube, Vimeo and Flickr. Everyday, thousands of users not only upload videos onto these websites*, they also interact with each other asynchronously by leaving comments and rating content. These interactions, which surround each shared video, leave behind large amounts of contextual metadata that we can use to better understand how people explicitly communicate, share and experience media with each other. But more than that, we believe that this metadata, generated through social sharing, can also provide insights into the characteristics of the media content itself. When aggregated and distilled, social sharing metadata (or social metadata for short) are not just representations of social actions but are also indicative of the underlying patterns of media consumption. In this paper, we present a method to harness the patterns of social interaction found in the social metadata of YouTube videos to provide insights into the nature and characteristics of the content being shared. In particular, we use a Bayesian approach to determine the genre category of YouTube videos using only their contextual metadata.

Historically the methods used to classify the genre of web videos have so far focused on the machine recognition of content within the videos. While this content-based approach has enabled classification of a wide variety and large quantities of online video, it also presents several problems. On its own, it can be technically complex and prior work have mainly focused on classification tasks where the genres are well-defined and commonly recognized.¹ Instead of focussing solely on understanding the content of a video, we argue the task of genre classification can be made more accurate by taking into account the social interaction surrounding online video content—in effect, how the video is discussed and interpreted superceeds its physical context.² Present methods

Further author information: (Send correspondence to David A. Shamma)

David A. Shamma: E-mail: ayman@acm.org, Telephone: 1 408 349 2684

Jude Yew: E-mail: jyew@umich.edu, Telephone: 1 408 349 2860

*YouTube reports that every minute twenty-four hours worth of video is uploaded onto its website: http://www.youtube.com/t/fact_sheet (Accessed July 2010)

of genre classification do not consider the social effects of sharing as a defining context. How a video is used, interacted with, and commented is often indicative of the nature of its content. For instance, the social interactions surrounding a Music video will be quite different from those surrounding a Comedy clip; users are more likely to scrub to particular points of interest in a Comedy clip and are more likely to let a Music video play in its entirety. When taken in aggregate across all users, we hypothesize that different video genre categories would have different patterns and signatures of contextual interaction surrounding them.³ By identifying these patterns in the social metadata, we can then predict a video’s specific genre category based on particular signatures of social activity.

We test our hypothesis using a Naive Bayes classifier to predict the genre classifications of 1,740 videos crawled from the YouTube website. We trained the classifier on a small number of features in the social metadata of the videos—features such as the number of views and the ratings the video received. The results produced by the classifier were then compared with the genre category originally provided by the video’s uploader. In summary, our classifier performed well, achieving up to 75.5% accuracy in predicting genre categories for the videos in our dataset. However, this accuracy is only achieved when we *factor* the ratings feature provided to the bayes classifier. It turns out that *factoring* a variable produces more accurate genre predictions from the classifier because it clusters the values and allows the Naive Bayes to better detect patterns of associations between the ratings score of a video and its genre category. In the following sections of this paper, we will detail our method in using social metadata to determine the genre classifications of YouTube videos and discuss the surprising finding that the ratings score of a video is somehow indicative of its genre category.

2. RELATED WORK

The popularity of social media websites can be partly attributed to the functionality and tools for interaction found on these websites. These tools allow users to not just share media content with each other, but to experience that media content with a greater degree of interactivity and sociality. Video sharing websites like YouTube have tools that let users leave comments and even express their opinion about the content being shared. The aim, perhaps, of socially sharing videos on sites like YouTube is to replicate the experience of watching television together in a virtual environment. Media is seldom experienced alone and a large part of that social experience involves backchannel conversations and frequent starts/stops to review key moments in the media content. The activities and behavior surrounding the social experience of media content has, until now, been ephemeral and difficult to capture. Through the tools provisioned by social media websites, we are able to capture the *digital traces* of these contextual interactions in the form of server logs and large-scale databases.⁴ These digital traces not only provides us with insight into the users’ behavior; when aggregated and distilled, we can also harness the “wisdom of the masses”⁵ by identifying patterns in the metadata to generate novel insights.⁶ An example of this approach can be seen in Crane and Sornette’s³ use of YouTube metadata to identify “quality” content amongst the 40 million uploaded videos. Their project utilizes the viewing histories of YouTube videos, specifically the “relaxation signatures” in the view counts after a video has attained the maximum number of watches. Differences between the “relaxation signatures” enable the researchers to filter out junk from those with high quality content that viewers want to watch.

On a similar note, Szabo and Huberman⁷ analyzed the view counts on YouTube and voting histories of stories on Digg.com in order to forecast a video’s/story’s popularity thirty days ahead. The authors found that they had to follow the “attention” paid to YouTube videos for much longer (10 days) than stories on Digg (2 hours) in order to make their popularity forecasts. The authors cite the difference as the nature of the content—Digg stories become outdated very quickly while YouTube videos can still be accessed on the site long after they are uploaded, hence content with longer life cycles are prone to larger statistical errors than those with more dynamic lifespans.

Predicting popularity based on social data is a type of classification. We are particularly interested in socially motivated content categorization. We wish to identify “signatures” in the social metadata and to make genre category predictions of the videos. Naive Bayes classifiers are well suited for this task. From a set of observed “features” on an object or an event, they can make accurate probability predictions of the object’s class. For instance, given a set of observed features like size, shape and color, a Naive Bayes classifier will assign the most likely class of an object belongs to. One reason for utilizing Naive Bayes classifiers is their robust performance

in identifying hidden patterns in the data for exploratory and predictive analysis.⁸ When compared with other methods, like decision trees, Naive Bayes classifiers performed better in terms of efficacy and accuracy.^{9,10} Naive Bayes classifiers can also perform well with small amounts of training data and they perform well even when the assumptions of independence in the input features are not met.¹¹ In a study of data characteristics that impact the performance of Naive Bayes, it was found that the assumption of feature independence was not a good predictor of accuracy. Rather, accuracy also improved when features were functionally dependent and when there is minimal loss of the mutual information between the features.¹² These attributes of the Naive Bayes classifier emerged as particularly well suited for our task of video genre classification, especially in contrast to other methods such as network decision trees.

3. CLASSIFICATION STUDY

Our project adopted a multi-phased process of data collection, exploratory data analysis and training/testing the Naive Bayes algorithm to make predictions about YouTube video genre categories. This includes crawling YouTube to obtain a corpus. Identifying the *type* of each collected feature in the metadata, to determine if we should nominalize that set. Finally we train several classifiers against the collected raw and factored data.

3.1 Definitions

For this study, we make a distinction in the social metadata collected between implicit *usage* and explicit *action*. Implicit usage includes data like number of page loads, video length in seconds, and other data which is generated by the nature of simply visiting, viewing, reading and not participating. Explicit action can be thought of as any write to disk operation which requires knowing who a person is and recording that a particular activity has taken place. In our case, this involves rating a video, favoriting, adding a comment, and so forth.

Notice, usage and action are two forms of social metadata. In fact, socially motivated methods, like collaborative filtering¹³ can be done with usage data alone. More so, on a site like YouTube, we find the community tends to converge on its usage. The distribution of clip duration, for example, tends to be within a general range depending on the community and site restrictions or guidelines. In our corpus collected (detailed in the following section), the average clip length was 229.9 seconds ($IQR = 74$). Explicit action, we hypothesize, converges differently. Ratings and number of comments vary depending on a video’s function, purpose, and viral popularity, and not necessarily on the video itself. Examples of this include videos responding to a trending meme, like Double Rainbow or Laughing Baby, which elicit a challenge or a call and response to action.¹⁴

3.2 Sampling a corpus

To collect our data, we crawled YouTube using the website’s application programming interface[†]. The adopted crawl method retrieves the top videos from each of the 15 genre categories listed on the YouTube website. From this list, the crawler then retrieves the related videos that are listed on each video’s page on YouTube. This process is iterated a few more times to get at the videos that are at least two degrees of separation from the initial list of top videos.

In total our crawl retrieved the unique identifiers and the associated metadata of 3,379 videos. The data collected by our crawler consisted of a video identifier, the video’s title, its published date, its description, the genre category the uploader used, the tags the video was labelled with, the video’s duration and the rating score it attained. For this project, we paid attention to attributes of the video like its duration, the number of views and the average rating score. Not all of these features in the metadata emerged as salient to the accuracy of predicting genre categories—given the results discussed in the next section of this paper, we argue that social metadata that captures the active participation of the user, rather than automatically generated log data, will provide the classifier with more predictive information about a video’s genre.

We also collected metadata about the genre category that the video uploader selected—this was an important aspect of our project as the genre assigned by the uploader functions as a form of “ground truth” with which to compare the predictions produced by our classifier. However, it is important to highlight that YouTube only provides a constrained list of 15 genre categories for video uploaders to choose from. There are a number of issues

[†]YouTube data API: <http://gdata.youtube.com/feeds/api/> (Accessed June 2010)

Table 1. Distribution of videos across genre categories. For this study, we examined the top 5 categories, representing 51.4% of the overall crawl.

<i>Category</i>	<i>Number of Sampled Videos</i>
Comedy	373
Entertainment	424
Film	242
Music	469
People	232
Other Categories	1,639
Total Crawl	3,379
Removing <i>Other</i> categories	-1,639
Total Study Sample	1,740

with relying on user-selected categories as a basis for “ground truth”, but we will save the discussion about the reliability of user assigned genre categories to the discussion section of this paper.

In the exploratory phase of the project, we conducted a series of descriptive statistical analysis to better understand the data and to determine the next steps of our project. One of the decisions that resulted from this phase of the project was the choice of using only the Top 5 genre categories represented in our dataset. The reason for this was that the number of videos were unevenly distributed across all the YouTube genre categories and some genres did not emerge as numerically salient. As can be seen in Table 1, we only paid attention to a subset of 1,740 videos from our entire crawled dataset.

3.3 Naive Bayes Classification

We initially trained the Naive Bayes classifier to determine the genre category of a YouTube video using four particular attributes found in the social metadata—the tags a video received, the duration of the video, the number of views and the rating score the video attained. We chose these four attributes because they were most reflective of user activity and interaction in the YouTube metadata. However, as we carried out our tests with the classifier, the tags attribute was simply too noisy to use. Because there was no limit on the number of tags that can be associated with a video and the different languages used for the tags, it was difficult to gain meaningful results from using tags as a feature variable for the classifier. Thus, we limited the number of features our classifier utilized to just the video’s duration, it’s view history and it’s average rating.

We trained on our classifier on different sample sizes ranging from 20%, 40%, 60% to 80% of our total dataset’s size. This is to ensure that our classifier is not unnecessarily biased by the size and amount of variance in our data. In addition to the classifier, we produced a random prediction, based on the known category distribution in the corpus, for comparison.

Using all the YouTube social metadata raw, without any manipulation, produces a poor prediction rate of 14.6% video categories correctly identified. This means that given all fifteen YouTube genre categories and all three features to work with, our classifier is able to produce accurate genre classifications on only 15% of the videos. This result is especially low when compared to the 23% accuracy performance by random guessing (see Table 2).

However, our classifier’s prediction accuracy is improved to 32.4% when we constrain our data set to only the top 5 genre categories represented in our dataset. While the accuracy rates are still very much below par, there is a doubling of the prediction accuracy when we only focus our classification task on the most represented categories in the data. One explanation for this improvement could be by providing our Bayes classifier with well populated genre categories we are providing it with adequate information accurately detect patterns in the social metadata. Quite simply put: noise reduction is effective but overlooks the less well-represented categories.

Table 2. Naive Bayes and Random Chance predictions on Video Categories across data sets. Chance prediction was based on random guess based on the corpus distribution of categories. Bayes predictions based on 80% training sample size.

<i>Random chance prediction accuracy</i>	
23.0%	Correct by Chance
<i>Naive Bayes prediction accuracy</i>	
14.6%	All YouTube features and all categories
32.4%	All YouTube features Using just Top 5 Categories Only
51.8%	Top 5 categories with factoring only “Duration” feature
66.9%	Top 5 categories with factoring only “Views” feature
75.5%	Top 5 categories with factoring only “Ratings” feature

Table 3. Naive Bayes predictions on Video Categories across different training sample sizes.

Training Set Sample Size	Prediction accuracy
20%	28.8%
40%	52.6%
60%	66.0%
80%	72.1%

3.4 Nominal Factoring

In summary, we found that our classifier performed well, but only when we “factored” the features used to train the classifier. In the following section we will discuss the results and performance of our classifier in detail, and in particular highlight the importance of factoring as a method in attaining more accurate predictions of genre categories for the videos in our dataset.

The results presented in Table 2 highlights the impact “factoring” has on the result produced by our classifier. Factoring is a process that transforms a list of variables into a set of nominal values by identifying all the unique values ($1 \dots k$; where k is the number of unique values in the nominal variable) and grouping all the variables in that list to that unique value. In other words, by making a list of variables into a factor, we are basically creating a vector of integers that has grouped all the similar unique values together. For instance, by factoring the ratings feature in our dataset, we are turning the list of 1,740 values in that attribute into a vector of unique integers, that has mapped to it, the number of times a particular value occurs. In our dataset, 178, or 10%, videos had a 5.0 rating. The rest of each nominal rating class represented $< 1\%$ of the distribution and were groups of 9 or less videos. Factoring ratings turns that attribute into a vector of “binned” values. When used as a feature for the Naive Bayes classification task, factoring an attribute improves the accuracy of the results tremendously. This generally asserts videos with a 3 star rating differ from videos with a 5 star rating, but makes no interval or linear assumption. In other words, we group explicit action data but make no assumption aside from associating videos with the same aggregated explicit action together. Additionally, the several classifiers were built on various training set sizes, see Table 3. With our relatively small sample, at least 60% of the corpus was needed to train an effective classifier.

Factoring implicit usage features (number of views, duration of the video) from the social metadata does not produce as high an accuracy rate as when we factored the ratings feature, see Table 2. We suspect that accuracy rates may decline with features where there is much more entropy in the data. Unlike the ratings variable which is constrained to a top rating score of 5, the views and duration attributes contain a much greater spread in their data.

Naive Bayes classifiers can perform better when explicit social action features are factorized. In this study, the explicit action feature, video rating, proved most predictive of a video’s genre classification. A Naive Bayes classifier trained on view count and duration features (both implicit usage from YouTube) and the factored ratings produces a 75.5% prediction accuracy. We found that individually factoring duration and ratings shows a small loss of performance (-4.3%) and individually factoring views and ratings showed a small gain (1.9%). Factoring all three features individually showed a 2.7% increase.

4. DISCUSSION

One conclusion we can make about our genre classification method is that using social metadata to determine a video clip’s genre can yield relatively accurate results. However, it is important to note that care should be taken when selecting the appropriate metadata to use for the classification task. As our results have highlighted (see Table 2), not all the social metadata were highly predictive of a video’s genre classification. We believe that features which are more reflective of the user’s social actions, rather than automatically generated usage statistics, are likely to provide more meaningful information to the classifier, and hence be more predictive about the characteristics of the shared content. An illustration of this point can be seen in the Szabor and Huberman study⁷ discussed earlier in Section 2. In that study, the authors found that they were able to more accurately predict the future popularity of Digg stories over YouTube videos. While we do not dispute the authors’ insights about their results, we feel that another explanation for the different results can be found in the contextual metadata used in their analysis. There is a qualitative difference between YouTube viewing histories and Digg story vote counts; voting requires explicit and conscious action on the part of the user while view counts are automatically accumulated data by the website. We believe that a strong influence on the accuracy of predictions made by Szabor and Huberman is the whether the metadata used captured conscious and intentional action on the part of the user. In the case of popular Digg stories, voting for an article requires conscious action on the part of the user, and is thus a stronger predictor of longer-term popularity than the viewing history. Likewise, we believe that explicit action data will be more predictive of a the content being shared because the data is representative of user intentionality—in the case of ratings, these metadata when aggregated highlight the users taste and preference in their sharing and consumption of videos.

This difference between implicit usage metadata versus metadata that captures a user’s explicit action is an important distinction to make and figures strongly in our classification task. Features used for classification should be selected on the basis of how reflective they are of the activity taking place around socially sharing videos. Metadata that captures the users’ explicit social actions provide the classifier with more information to accurately map the patterns in the data to genre classifications. Of the three features that we used in this project, the ratings attribute proved to be the most predictive in our classification task because the act of rating a video is a conscious user action expressing his/her opinion about the video clip. This intentionality of the users, expressed through the ratings feature, when factored and combined with the other features found in the metadata enables our Naive Bayes classify to more accurately predict the genre of a shared video. Additionally, it is also important to select contextual metadata that lend themselves to the nominal factoring process. Out of the three features we used with the classifier, ratings were the most amenable to be factored. We believe that the factoring process made the classifier more accurate because it reduced entropy in the data by clustering all the values in the ratings variable along this five point scale. The Naive Bayes algorithm produced more accurate predictions with the addition of the factoring step on particular variables. Thus, it is important to be knowledgeable of the social metadata and be selective about which features to use in our Bayesian method of genre classification.

5. CONCLUSION

Media content does not exist in a vacuum, by ignoring the social experience of sharing and consuming media on the internet, we are overlooking information that allow us to better understand why users engage in social sharing in the first place. But more than that, this project has shown that the actions surrounding socially sharing videos are also reflective of the properties of the content itself. Given the popularity of social media websites, we are presented with the opportunity to capture social sharing metadata on a large-scale to derive greater insight into questions about the taste, opinion and cultural preferences of users today. Our project represents a first step in this direction; by harnessing the patterns and signatures found in the metadata of social video sharing, we can determine, with relative accuracy, the genre category of the content.

While we are encouraged by the results from our classifier trained on YouTube data, some lingering issues still remain from this project. For instance, we make the distinction between implicit usage and explicit action in the collected metadata, and argue that metadata which captures a user’s intentional actions tend to be more predictive of genres than automatically collected usage data. In this paper, we’ve highlighted the use of the ratings feature, which we contend is a much more informative feature to use in the Bayesian classification task.

In the next iteration of this project we would like to investigate what other kinds of social metadata captures a user's conscious and intentional action in the social sharing of content. To do this, we have to look at other online spaces that may have different ways in which users can interact with each other and with the shared content. One such avenue to continue our investigations is the area of synchronous video sharing, where a video clip is shared in real-time over a chat or Instant messaging session. We speculate that highly social interactions, such as synchronous Instant messaging or chatting while sharing a video clip will ostensibly provide us with greater and more detailed data about the ebb and flow of conversation and the experience of sharing a video clip in real-time.

Another open issue uncovered by our project is the imperfect nature of classification systems and their ontologies. To verify the accuracy of our classifier, we conducted a quick and dirty content analysis of some of the videos that were classified. What this task revealed was that, one, the YouTube categorization scheme was highly problematic and two, the user's choice of genre categories for their uploaded videos was often entirely subjective and flawed. For instance, the genre of "Entertainment" in YouTube was often a catch-all category of all kinds of videos ranging from clips of TV variety shows to music videos. The problems associated with classification systems and the task of assigning objects to their appropriate category is not new. In their examination of the far-reaching impact of classification systems, Bowker and Star¹⁵ noted that no classification system is perfect and that users of classification systems often ignore the boundaries between categories or even routinely mix them up. This is especially evident in the user-generated classifications of the videos on YouTube. As more media content on the Internet rely on user-generated genre categories, the genre classifications associated with these content will become increasingly unreliable and difficult to sort out. In the next iteration of this project, we would like to look more carefully at the accuracy of our method of classifying video genres by not only comparing it with the uploader's categories, but also by comparing the classifier's results with the opinions of test subjects in a controlled classification task. The main goal of doing this would be to sort out the issues of classifying the genre of online videos via the use of social metadata, as opposed to user generated classifications.

We have highlighted a viable method for future systems to determine the genre categories of online media content. While there are still several important questions left unanswered, we believe that the utilization of social metadata that captures explicit user action will be a fruitful avenue for future research work.

ACKNOWLEDGMENTS

We like to thank and give a shout to Elizabeth F. Churchill, Bryan Pardo, Eytan Bakshy and Matt Cooper for their advisement and support. Also, thanks goes to Justin Weisz for lending us his crawler.

REFERENCES

- [1] Roach, M. and Mason, J., "Recent trends in video analysis: A taxonomy of video classification problems," in [*In Proceedings of the International Conference on Internet and Multimedia Systems and Applications, IASTED*], 348–354 (2002).
- [2] Shamma, D. A., Shaw, R., Shafton, P. L., and Liu, Y., "Watch what i watch: using community activity to understand content," in [*MIR '07: Proceedings of the international workshop on Workshop on multimedia information retrieval*], 275–284, ACM, New York, NY, USA (2007).
- [3] Crane, R. and Sornette, D., "Viral, quality, and junk videos on youtube: Separating content from noise in an information-rich environment," in [*Proc. of AAAI symposium on Social Information Processing, Menlo Park, CA*], (2008).
- [4] Wesler, H. T., Smith, M., Fisher, D., and Gleave, E., "Distilling digital traces: Computational social science approaches to studying the internet," in [*The Sage handbook of online research methods*], Fielding, N., Lee, R. M., and Blank, G., eds., 116 – 140, Sage, London (2008).
- [5] Surowiecki, J., [*The Wisdom of Crowds*], Anchor (August 2005).
- [6] Segaran, T., [*Programming Collective Intelligence: Building Smart Web 2.0 Applications*], O'Reilly Media, 1 ed. (August 2007).
- [7] Szabo, G. and Huberman, B. A., "Predicting the popularity of online content," *Commun. ACM* **53**(8), 80–88 (2010).

- [8] Larsen, K., “Generalized naive bayes classifiers,” *SIGKDD Explor. Newsl.* **7**(1), 76–81 (2005).
- [9] Lowd, D. and Domingos, P., “Naive bayes models for probability estimation,” in [*ICML '05: Proceedings of the 22nd international conference on Machine learning*], 529–536, ACM, New York, NY, USA (2005).
- [10] Caruana, R. and Mizil, A. N., “An empirical comparison of supervised learning algorithms,” in [*ICML '06: Proceedings of the 23rd international conference on Machine learning*], 161–168, ACM, New York, NY, USA (2006).
- [11] Domingos, P. and Pazzani, M., “On the optimality of the simple bayesian classifier under zero-one loss,” *Mach. Learn.* **29**(2-3), 103–130 (1997).
- [12] Rish, I., Hellerstein, J., and Jayram, T. S., “An analysis of data characteristics that affect naive Bayes performance,” Tech. Rep. RC21993, IBM (2001).
- [13] Herlocker, J., Konstan, J., and Riedl, J., “Explaining collaborative filtering recommendations,” in [*ACM 2000 Conference on Computer Supported Cooperative Work*], 241–250, Association of Computing Machinery, Association of Computing Machinery (12/2000 2000). [ipjIn proceedingsi/pj.](#)
- [14] Arceneaux, K., “The remix as cultural critique: The urban contemporary music video,” *Popular Music and Society* **16**(3), 109–124 (1992).
- [15] Bowker, G. C. and Star, S. L., [*Sorting Things Out: Classification and Its Consequences*], Inside technology, The MIT Press (October 1999).