

Knowing Funny: Genre Perception and Categorization in Social Video Sharing

Jude Yew
School of Information
University of Michigan
Ann Arbor, MI, USA
jyew@umich.edu

David A. Shamma
Internet Experiences Group
Yahoo! Research
Santa Clara, CA, USA
aymans@acm.org

Elizabeth F. Churchill
Internet Experiences Group
Yahoo! Research
Santa Clara, CA, USA
churchill@acm.org

ABSTRACT

Categorization of online videos is often treated as a tag suggestion task; tags can be generated by individuals or by machine classification. In this paper, we suggest categorization can be determined socially, based on people's interactions around media content without recourse to metadata that are intrinsic to the media object itself. This work bridges the gap between the human perception of genre and automatic categorization of genre in classifying online videos. We present findings from two internet surveys and from follow-up interviews where we address how people determine genre classification for videos and how social framing of video content can alter the perception and categorization of that content. From these findings, we train a Naive Bayes classifier to predict genre categories. The trained classifier achieved 82% accuracy using only social action data, without the use of content or media-specific metadata. We conclude with implications on how we categorize and organize media online as well as what our findings mean for designing and building future tools and interaction experiences.

Author Keywords

Video, social, classification, categorization, YouTube, Naive Bayes, genre, interview, survey

ACM Classification Keywords

H.5.1 Information Interfaces and Presentation: Multimedia Information Systems; J.4 Computer Applications: Social And Behavioral Sciences

General Terms

Human Factors, Experimentation

INTRODUCTION

In an episode of *The Simpsons*, Homer lectures Bart about shoplifting, "Why do you think I took you to all those Police Academy movies? For fun? Well, I didn't hear anybody laughin', did you?" [18] Here, the genre of the comedy, *Police Academy*, is recategorized by Homer based on the audi-

ence's humorless reception of the movie. Although clearly a joke in itself, Homer's perception of *Police Academy* illustrates how the social consumption of media can alter the way content is perceived and categorized. Media content that we believe to fit a particular genre is both constituted by, and constitutive of, the changing social contexts in which that content is produced, shared and consumed [1]; genres are socially constructed. This article presents a study of genre categorization and demonstrates how the analysis of social consumption and sharing behaviors can reveal the nature and characteristics of online video content. Genre engenders recognizable patterns of social action and we can infer the genre of a video by looking at what people do with that content.

Genre categories have both epistemological and functional dimensions; they are first and foremost ways of organizing and defining content, but also ways of organizing *social actions* [1]. For instance, the category Comedy helps us identify and expect a particular kind of video content, such as a humorous narrative and funny characters. At the same time, Comedy also presents content that specifies a type of individual behavior and social interaction that would be different, or even inappropriate, if one were to watch a Documentary. Besides the dual nature of genres, they can also be understood as "socially constituted systems" [5] that shape the context and social activity surrounding them. Genres are social constructs which specify particular interaction patterns and social activity, which in turn, determine how we define and interpret those genre categories [14].

Current automated tag-based classification techniques ignore these aspects of social construction. In this paper, we focus on the categorization of videos that are socially shared on the Internet. The online social sharing of media can take place asynchronously, through social media websites like YouTube, or synchronously, with services like <http://justin.tv> or Yahoo! Zync. In addition to simply sharing videos, people interact around the video by leaving comments, rating content or, in the case of synchronous sharing, by chatting. These interactions leave traces which are either explicit, publicly viewable annotations in the form of comments and ratings, or are implicit, logged usage data; both can be mined to reveal patterns of interaction and engagement with the video content. More importantly, the data resulting from social actions around the video content can also provide insights in to the nature, characteristics and genre classification of the video content itself.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

In this paper, we address people’s understanding of genre categories from a social construction perspective. For the purposes of our analysis we focus on Comedy as a genre. We develop a model of video genre classification using data captured from the social consumption and sharing of online videos. The predictive model is built from 5 features sampled from a dataset of 9,364 videos taken from 2,188 sharing sessions. The model classifies videos by genre at an accuracy greater than 82%.

In the following sections, we review the online video categorization task from a human and computational perspective. To gain further insight into the human perspective of comedy, we present findings from two online surveys and eleven 1-hour, semi-structured interviews with a sample of the survey participants. The findings from the qualitative exploration are used to inform the training features used for automated genre prediction. Finally, we conclude by suggesting what this model means for current recommendation systems as well as for the interaction design of future systems.

CATEGORIZATION

[categorization is] not a social act simply because it takes place in some social context; it is social because it is at work in shaping the very context within which it functions [1]

Social action can transform how a media event is received by an audience [1]. For example, overlaid commentary can transform a serious event into something comedic. The cult TV series *Mystery Science Theater 3000* ran from 1988 to 1999 based entirely on this premise, the main protagonists providing a running commentary over a series of science fiction B-movies and transforming what had initially been created as “serious” content into comedy. Given social action can thus transform the reception of a media event, we can conclude that without social context, categorization is a difficult task to do accurately, and that representing content by classification must account for contextual and social uses. Bowker and Star explain this difficulty, “people (and the information systems they build) routinely conflate formal and informal, prototypical and Aristotelian aspects of classification. There is no such thing as an unambiguous uniform classification system.” [2] In reality, things cannot be sorted easily into mutually exclusive and complete categories. Online videos are no exception; they can be categorized on several levels such as, the semantic or narrative content of the video, the inherent properties of the video itself like its format, by stylistic features like mood, and so on [16].

Methods used to classify the genre of web videos have traditionally focused on the video’s content or metadata. Classification of online videos is computationally treated as tag suggestion from a pre-defined set [16]. However, we argue the task of genre classification can be made more accurate via an examination of the social interaction surrounding online video—in effect, how the video is discussed and interpreted. How a video is consumed, interacted with, and commented on is indicative of the nature of its content. In the following two subsections, we detail arguments made from related

work towards the issue of classifying videos from a social action point of view.

Categorizing on social data

Online tools for content viewing support many forms of social interaction [3, 23] such as link sharing/forwarding, ratings, “liking” as a public statement of endorsement, explicit recommendation, commenting, synchronous chatting while watching, and so on. In synchronous video sharing contexts, people also start/stop videos to review key moments and discuss them. These activities leave digital traces in the form of server logs and large-scale databases [24]. These digital traces are used to assess the popularity of content, and to create promotional mechanisms (such as leaderboards) to further raise the profile of the content for recommendation. Activity data are also mined, aggregated and analyzed to develop models of user consumption patterns. As well as understanding people’s consumption practices around content, these activity data can be used to categorize the content itself. Crane and Sornette [6] identified “signatures” in the metadata of YouTube videos in order to identify “quality” content. Whitman et al. [25] utilized textual data from online forums to classify and infer similarity in music content. However, while actions on or around a media object by individuals have been explored to classify content, social actions have not been explored. In our work we analyze patterns of activity between people over or around that media object as a means of classifying content.

Categorization as Tagging and Recommendation

Historically, in computing and the social media, categorization is a collaborative filtering task for tag suggestion. Computationally, finding a category for a video can be the equivalent of suggesting the single best tag for that video. Automatic tags and tag suggestions are often content-based either from the video’s meta-data or the visual media’s content [8]. Recently, Zhang, Zhang, and Tang [27] proposed a hybrid method based on video meta-data and social-graph distance. Similar web-graph systems have been proposed for image classification [13]. These richer tag sets have been shown to improve the performance of web video categorization [4]. Similar approaches use tag metadata co-occurrence using machine learning (ML) or term-ranking techniques [22, 15].

From the start of tagging’s popularity, social navigation, or recommendation, has been an ongoing research investigation [19]. Likely the most publicized example is the Netflix Prize contest. From a dataset of 100 million movie ratings, the open contest challenged people to find a 10% improvement over the current state of the art. The solutions used a variety of information types, from ratings data to the movie rental’s ZIP-code location to the renter’s age; no content was used. In general, the actual video, image, or audio signal/content is believed to be unreliable for recommendation. Instead, ML researchers look for explicit distinct actions, like 5-star rating scores, to find similarity [21] and categorization. For example, Yew and Shamma [26] trained a Naive Bayes classifier to predict YouTube categories to an accuracy of 75.5% using only three features of YouTube Metadata: video duration, view count, and 5-star rating. In

particular, their finding showed that nominal factorization of explicit data, in their case: ratings, improves performance without the need to increase the training set size. They attribute the rating's dominance to "explicit social action" being a higher fidelity signal than implicitly aggregated usage data. From our findings in this study, we model "implicit social sharing behaviors" as the construction of genre.

Research Questions

Is there a correlation between genre and watching behavior? If so, could watching behavior be more effective than 5-star ratings when used for categorization? This paper aims at identifying this connection as guided by two primary research questions:

[Research Question 1] How do people socially consume, perceive, and categorize videos (in particular, comedy)?

[Research Question 2] Can a predictive model be built to automatically categorize media using social traces (such as conversational and interactional metadata) and without the use of explicit annotations or video content?

METHOD

This work employs a mixed-method approach to explore the issues of categorization of shared online videos, answering RQ1 qualitatively, which in turn should advise a quantitative model to answer RQ2. This approach consists of three phases combining multiple data collection and analysis methods. The first phase conducts two surveys, a pilot survey with 43 respondents administered internally and a larger survey conducted via Mechanical Turk with 69 valid responses, in order to assess the difficulties associated with categorizing online videos by humans. In phase two, the surveys are followed up with 11 interviews from the phase one survey respondents. The interviews involved questions about the respondents' mental processes when categorizing online videos, in particular videos that they labeled as "comedy." Finally, we use the findings of the surveys and the interviews to develop a classifier for categorizing video genres using only metadata from conversational and interaction activity. The main aim of adopting a machine learning classifier is to devise a method of genre classification based on how we watch videos, and not what is in them. The results from all three phases will then be used to explicate on the issue of why the categorization task, especially for comedy videos, is so difficult. Each method complements the others and allows for a stronger generalization of genre categorization from various human-computer interaction perspectives [7].

Finding a Conversational Video Dataset

For this study, we need to investigate how people judge the category of a set of videos and compare that judgement to a classifier. Since we will be training the classifier on conversation activity, the metadata that arises from crawling a site like YouTube is insufficient. Conversational activity surrounding video sharing can be found online in various live-stream rooms like JustinTV¹ and video-on-demand chat and

¹<http://justin.tv> (Accessed 9/2010)

synchronous sharing tools like CBS Social TV² or Meebo Rooms³, Yahoo! Zync⁴ [20] which lets people host viewing parties with their broadcast TV content.

We acquired a 24-hour sample of the Zync event log for Christmas Day 2009. Zync allows two people in an instant-message session to watch a video together; the video's playback stays in sync across both participants and both participants share playback control. This provides a set of watched videos from YouTube as well as conversational activity rich enough to train a classifier. The dataset records several features: anonymous user id hashes, session start/stop events, the session duration, the number of play commands, the number of pause commands, the number of scrubs (fast forwards or rewinds), and the number of chat lines typed as a character and word count. For the chat lines, the dataset contained no actual text content, only the aggregate count of characters and words. The only text that is collected is video URLs and emoticons. Each item/activity collected is a line in the dataset which records the time of the event and the playback time on the video [12].

In total, the dataset contained 2,188 dyadic sessions that shared a sum of 9,364 YouTube videos. Not all sessions appeared complete and were missing the start/stop events, was missing a participant, or other critical information; these sessions were discarded. We then took each remaining YouTube video identifiers and retrieved the related metadata associated with each video using YouTube's Application Programming Interface⁵. Some videos were no longer available on YouTube due to copyright infringement or owner deletion; these videos and their respective sessions were also discarded. The final test sample consisted of 1,732 videos with valid metadata. The data collected from YouTube 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 5-star rating score it attained. Of these data, we only use the video's genre/category, of which YouTube provides a constrained list of 18 genre categories. The YouTube video's category is specified by the person who uploaded the video, which is *required* at the time of upload. For this article, it functions as the predictive category for our surveys and classifier and allows us to investigate if there's a match between the uploaded category and the social actions.

Within our sample of 1,732 videos and associated sessions, not all categories were well represented. However, the category distribution was unevenly distributed. To ensure we had enough data to accurately perform this study, we retained only the Top 5 video categories for our investigation; this discarded 13 categories, 11 of which contained fewer than 20 videos of that category watched. The final categories from the remaining 1,580 videos were, in order of frequency, Mu-

²http://www.cbs.com/social_tv/ (Accessed 9/2010)

³<https://meebo.com/rooms/> (Accessed 9/2010)

⁴<http://sandbox.yahoo.com/Zync> (Accessed 9/2010)

⁵<http://gdata.youtube.com/feeds/api/> (Accessed 9/2010)



Figure 1. Screenshot of classification task in the pilot survey

sic (896), Entertainment (232), Comedy (212), Film (121), and People (119).

Surveys

Our investigation begins with two surveys: 1) a pilot survey and 2) a larger survey utilizing Amazon’s Mechanical Turk service. In both surveys we presented each respondent with a set of 20 videos, selected from our corpus of retrieved videos. The primary task in both surveys is to view each video and carry out a categorization task.

Pilot Survey

In the pilot, we assess how individuals categorize shared online videos from websites like YouTube. The online survey involved presenting each respondent with one video at a time randomly from our video corpus (Figure 1). The respondent would watch or scan through the video and then make a judgement about the genre category of the video by selecting one of the Top 5 sampled genre categories: Music, Film, Comedy, Entertainment and People and leave an optional comment. An agreement between the survey respondent’s genre selection and the video uploader’s original genre designation signified an accurate judgement. The survey call was posted to 3 internal mailing lists within our organization. A total of 43 valid and complete responses were gathered from this survey. Respondents were able to correctly categorize 60% of the videos with the same label as the original uploader’s categorization. Looking at Comedy specifically, the success rate was 50%. The comments left by the respondents on the incorrectly labeled comedy videos surfaced a lack of distinction between what is comedy and what the respondent believed was funny. One respondent illustrated this confusion in their comment as, “Comedy is the category...but I guess that doesn’t require it to be funny.” The respondents knew what was funny to them, but were less readily able to classify videos based on whether it belongs to the Comedy category. The pilot survey’s results prompted us to ask, why are people so bad at classifying Comedy genre videos from YouTube?

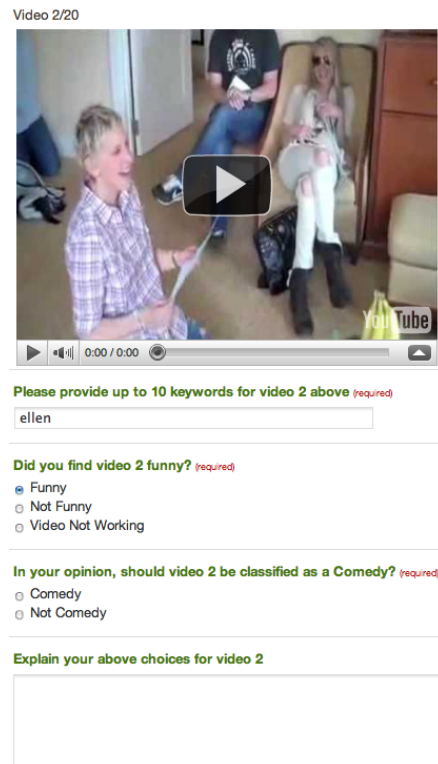


Figure 2. Screenshot of classification task on Crowd Sourced survey

Crowd Sourced Survey

We designed a second survey to address responses to what is funny and what is comedy. First, we performed a content analysis of the videos which surfaced the following media-types: cartoons, stand-up comedy, music videos, foreign film clips, and video blogs. In the second survey, each survey respondent was also presented a set of 20 videos at random. However, this time the survey presented each respondent with the same set of 20 videos. The videos were selected as a distribution of 10 Comedy videos and 10 Non-Comedy; the latter encompassed videos from the 4 other categories. Additionally, the second survey asked a different set of questions (see Figure 2).

This second survey asked “Comedy or Not” and “Funny or Not,” which differs from our first survey which asked for an explicit category label. Discussions with the pilot survey respondents indicated that category selection is dependent on the number of choices and more choices result in wider distribution of selections: i.e., consensus goes down.

The second survey was administered externally using the Crowdfunder *crowdsourcing service*⁶ to recruit respondents from Amazon’s Mechanical Turk⁷ (AMT). “Crowdsourcing”

⁶Crowdfunder assists with “crowdsourcing” respondents for tasks from several online “labor on demand” pools, see <http://crowdfunder.com/> (Accessed 9/2010)

⁷Amazon’s Mechanical Turk service is an crowdsourcing marketplace that enables users to recruit large numbers of workers for tasks that computers do not do well. See <https://www.mturk.com/mturk/welcome> (Accessed 9/2010)

is a quick and inexpensive method to gather responses for our survey which has grown in recent popularity within research communities. We followed the suggestions of Kit-tur et. al [9] to ensure data quality when using crowdsourcing techniques. Crowdfunder’s service pulls from a pool of *trusted* turkers as respondents for a survey. Our survey gathered 103 responses of which 69 of them were valid and complete responses.

For the questions “Is this funny” and “Is this a comedy?,” we test for inter-rater reliability using Krippendorff’s alpha [10] in order to determine if there was a consensus amongst the respondents about which videos were funny and which videos should be classified as a comedy. Similar to our finding in the pilot survey, respondents showed little agreement about Comedy videos or if they were funny (Funny: $\alpha = 0.374$, Comedy: $\alpha = 0.259$). Higher reliability scores for the non-comedy videos highlight more agreement, something is obviously funny or not funny (Funny: $\alpha = 0.687$, Comedy: $\alpha = 0.464$).

The inter-rater reliability test provides only a general insight about how the videos were perceived and classified. For a more detailed understanding, we conducted a text analysis of the keywords and comments solicited for each video. Table 1 lists the frequency of the keywords “comedy” and “funny” provided by the respondents for each video. Not all the comedy videos (videos 1–10) were tagged as “comedy” by the survey respondents. In particular, videos 8 and 9 are more frequently tagged as “funny” rather than “comedy”. To better understand why these two videos are tagged more frequently as “funny,” we reviewed the content of both these videos. These two popular, viral videos share the following characteristics: both have generated millions of views, are short, are easy to understand, and are humorous and “cute”. Video 8 is *Laughing Baby*⁸, a video where a swedish baby boy bursts into giggles whenever his father makes the “bing!” sound. Video 9 is *Sneezing Panda*⁹, it depicts a tiny infant panda, who is sleeping at the feet of its mother, suddenly sneezes about halfway into the video giving his mother a huge fright. For our respondents, both these videos exemplify videos that are “funny,” but don’t necessarily fit the definition of a comedy genre video. Some of the comments provided by the survey respondents about these two videos include:

This is very funny video of a baby laughing. Not sure it should be categorized as a comedy. (Respondent 16)

Its funny but its only an animal getting startled by her sneezing baby. It is not comedy because the actions were not specifically done to make us laugh. (Respondent 9)

For the respondents, not all funny videos should be classified as comedy. So what exactly are the characteristics of a comedy video? Just by looking at the frequency of the

⁸<http://www.youtube.com/watch?v=5P6UU6m3cqk> (Accessed 9/2010)

⁹<http://www.youtube.com/watch?v=FzRH3iTQPrk> (Accessed 9/2010)

Video Category	Keyword Count		
	Video	Comedy	Funny
Comedy	1	6	18
	2	34	12
	3	43	16
	4	10	50
	5	7	31
	6	52	25
	7	2	7
	8	0	29
	9	3	33
	10	26	5
Non Comedy	11	9	13
	12	1	21
	13	1	27
	14	1	4
	15	0	5
	16	2	8
	17	0	1
	18	1	4
	19	2	5
	20	1	3

Table 1. Frequency of the terms “comedy” and “funny” appearing in the keywords solicited from the survey respondents.

comedy	act	club	comedian
stand	standup	actor	monologue
shirt	crowd	audience	mic
talking	guy	laughter	laugh
performance			

Table 2. Terms associated with the keyword ‘comedy’ across all videos

tag ‘comedy’ used, there is no video in our sample where there is a clear consensus in describing it as a comedy. To better understand what our respondents typically view as comedy, we carried out a word association analysis on the tags provided for each video. Specifically, we transformed all the tag/keywords responses into a term-document matrix in order to find terms that correlate with the keywords “comedy,” essentially the term-frequency portion of a *tf · idf* model [17]. The results of the word association analysis provide us with some insight into how the respondents characterize videos tagged with the keyword “comedy”. Table 2 shows the strong association between the keyword “comedy” with descriptions of comedic structure or stylistic features such as the presence of audience laughter (or a laugh track) or if the video depicts the performance of a stand-up comedy routine.

The same word association analysis for the keyword “funny” was not as meaningful as the analysis of “comedy”. The analysis retrieved a total of 109 terms associated with the keyword “funny”.

Survey Results

One major finding from the 2 surveys show that people have difficulty in classifying videos with the genre “comedy” but they are more readily able to identify whether the content of

a video is funny. One reason why comedy videos are hard to classify is because people have very set structural and stylistic norms they look for—such as the presence of a laugh track or a standup comedy routine. However, a cursory scan of the comedy category on YouTube will reveal that most, if not the majority, of the videos there do not conform to this form of stylistic classification. In fact, many highly popular and viral videos, such as *Laughing baby* and *Sneezing panda*, fall between genre categories and are often difficult to place within existing categorization schemes.

Interviews

We followed-up the surveys with 11 interviews from the survey respondents. Namely we wanted to find their rationale and mental process for classifying and distinguishing between videos that belong to the comedy genre and videos that are funny. The interview questions and process involved having the interviewees explain their responses in the survey by showing them specific videos from their survey. One of the main goals of the interviews is to inform the selection of metadata features to use in the next stage of our project. All the interview participants were contacted via email and interviews were conducted face-to-face whenever possible.

The interviews highlighted the fact that the task of classifying online videos is much harder than the classification of movies. According to Participant 5:

Films in general tend to fit categories more narrowly. . . if Netflix says that something is a screwball comedy, I know what to expect. I think on YouTube the range of possibilities for the content of the videos is much less constrained. Cause it might literally be a segment from a film or it could something shot on a cheap digital camera. (Participant 5)

The variability in content makes categorization more complicated. Especially so with the Comedy genre, where the format of the content aids with the understanding of whether something was meant to be funny or not. For instance, it would be much easier to categorize a video as a comedy if the content were presented as a short film. If that same content were to be presented as in the format of a home movie video-clip, it would not be so easy to conclude that the video was a comedy because the home movie format is more ambiguous with regards to intentionality. The issue of intentionality is paramount for determining whether a video is to be classified as comedy:

Comedy videos are harder to classify. . . did it make me laugh? Was the intention of the video to make me laugh? An excerpt from *The Office* or standup would be “Comedy”. It gets hazier when the intention of the clip isn’t to make you laugh—but the content is funny. (Participant 2)

Participant 2’s comment highlights the problem of categorizing comedy videos found in the surveys; it is harder to classify a video as comedy than it is to categorize something as funny, especially when the intention of the video’s producer is hard to discern. This issue relates to why videos like *Laughing baby* and *Sneezing panda* are hard to categorize

as comedies. Without a priori knowledge of the video’s uploaded intentionality, what are some of the other cues that indicate a video is a comedy? According to two interview participants:

If I ever hear the laughter track, it doesn’t even have to be funny, it’s intended to be Comedy. The style of it. . . even if its a music video. (Participant 3)

Anything comedy is always impressed upon you with laughter in the background or some funny accompanying music. . . those contextual cues. (Participant 4)

Both participants 3 and 4 corroborate our earlier survey result that the genre of comedy was dependent on a few universally accepted features, such as a laugh track, even if the content wasn’t even funny. A reason for this is because stylistic features such as audience laughter relate back to the earlier notion of intentionality—background laughter signals to the viewer that the content is *meant to be funny* and should be categorized as Comedy. These are social indicators of *intended* comedy.

Apart from some of these contextual cues, Comedy, as a genre, shares little else in common. Our interview subjects had particular difficulty in categorizing videos that were labelled as *comedy* and often disagreed with the classification selected by the original uploader for these videos. In fact, some of the interview subjects have found that genre categories are sometimes used as prescriptions for how to appropriately interact with the content being presented:

They (the original uploaders of a video on YouTube) are uploading things and categorizing it as “Comedy” because they are proposing that there’s something funny in it. Even though the content itself may not be “Comedy”. And so, if you don’t watch to the end, you won’t pick up the intended funny part. (Participant 1)

This echoes our study’s premise—if people have little agreement in classifying comedy videos, what else, besides intention and contextual cues, can we use to help identify and categorize these videos? In the view of some of our interview subjects, one approach towards identifying whether a shared video is a comedy or not, is by studying the social actions and interactions surrounding the video:

There’s a context that you need to have for something to be identified as comedy. . . when I see a video that I have no context for I don’t know whether to identify it as funny. But if other people are interacting with it in a way that makes me believe that it’s funny. Same thing for the wedding dance. . . my interaction with it is, people are saying that this is funny. (Participant 5, referring to the virally popular video *JK Wedding dance*¹⁰ video.)

The above quote echoes the illustration from *The Simpsons* cited at the start of this paper. For Participant 5, and for Homer, the interactional context surrounding a video/movie,

¹⁰<http://www.youtube.com/watch?v=4-94JhLEiN0> (Accessed 9/2010)

how others are perceiving the content and how they are interacting with each other around the content, helps the both them to categorize the content and interact with it in a contextually appropriate manner. Genres are not just organizational tools, but also underpin social behaviors that surround the content of the shared videos. Towards our automated classification method, we will need to look for metadata that similarly highlights the behaviors and activities of the users surrounding shared videos to utilize as features for the classifier. In our case, the metadata from the Zync tool, detailed in the next section, consists of logs of video sharing sessions where users can chat, and at the same time, control the playback of a video. The interaction between the pair of users and their activity controlling the video constitutes as the social and interactional context that can be used to indicate the nature and characteristics of the video content.

It is important to note that not all social actions are indicative of a video's genre. Social media websites afford a variety of ways to interact, communicate and participate around the sharing of media content. Indiscriminate use of metadata surrounding socially shared videos for automated classification may not yield meaningful results. So we need to identify metadata that is representative of a person's relationship with the content and with others. As Participant 6 noted, utilizing metadata from individuals who are strongly related to each other, or who resemble each other, in some way will yield meaningful cues about the content of a video.

When consuming a video, I trust other consumers more than I trust the producer. So I would be interested in what other people who think like me, say about this video. . . Probably out of all the users on YouTube, people who look like certain niches have affinities to certain videos. (Participant 6)

This speaks towards how a video could be treated by various people, perhaps in different ways by different niches. Therefore, one important use of social metadata is that it can provide contextually and socially appropriate information to determine how the perception of a video's content by a particular group of individuals.

Another indicator of the viewer's relationship to a video can be found in the commentary and surrounding discussion.

Sending a photo of your dog and expecting them to comment on it is one thing. Sending a 15min video of your dog and expecting them to comment on it. . . they are two different things. . . the latter makes them commit to something much longer. (Participant 3)

As suggested by the interviewee above, commentary and discussion surrounding a video requires much greater effort and commitment from the users. Because of this, metadata derived from these contextual comments and discussions can provide a more robust signal than 5-star rating data that is typically used in contemporary ML approaches. Comments and conversational activity surrounding a video not only requires a greater commitment from the user, but is also more indicative of their opinion and relationship to the video content and other people.

Given the comments provided by the interviewees, our automated method of video classification will similarly utilize social metadata that is reflective of the social and interactional activity surrounding shared videos. In particular, we pay close attention to the use of social metadata that captures the conversational and interactional activity between the users and the shared video.

Automated Classification

Having elaborated the issues our participants raised, we now turn to automated classification of genre. From the surveys and interviews findings, and congruent with current ML research, content alone is insufficient for genre classification. Traditionally, asynchronous annotation data, such as 5-star ratings, produce the meaningful signals for classification, but do not fully account for the social construction of genre in practice. Activity data such how many times the video was watched in a single viewing session, the amount of conversational exchange (for example in chat) and use of video controls (such as fast forward, pause, etc) signal how the video functions within the social interaction between people.

Features for Training

For training a classifier, we aggregate all the events in each session into a feature vector. In our dataset of 1,580 video sessions, 80 sessions contained more than one video. These sessions are split into separate sessions. For example, if Bob and Mary have a 12 minute session where they shared a Music video for 8 minutes then a News video for 4 minutes, we create two unique sessions one 8 minute and one 4 minute. This ensures every session has one video to classify. This brings 1,660 sessions in total after multi-video sessions have been subdivided.

Each session yielded a large amount of data from feature use. For the purposes of this research, our focus was on social actions, specifically social interaction and shared control features. Data, such as load counts, emoticons, the video's event timestamps, and the event time were not considered. Social interaction features included for the classification were session duration, the number of play commands, the number of pause commands, the number of scrubs (fast forwards or rewinds), and the number of chat lines typed as a character and word count. Our goal is to predict categories accurately from social behavior *using as little data possible*. The final feature vector consists of two parts: the activity counts from the session's inviter and the activity counts of the receiver. Each video's feature vector is an aggregated count of activities from each person in the dyadic social exchange. The original YouTube tube video category, provided by the uploader, is used as the predictive category class variable for the independent feature vector. In effect, we are training the classifier to match non-content, conversation activity patterns to the source video's uploaded category.

Classification Accuracy and Nominal Factorization

As 1,660 feature vectors is a relatively small sample, we begin by using a generous training set, one that is 80% of the size of our feature set, a Naive Bayes classifier predicts a

Training Set Size	Prediction Accuracy	
	Raw Data	Factored
20%	26.33%	52.22%
40%	53.48%	67.59%
60%	54.30%	75.32%
80%	54.30%	82.34%

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

video’s genre poorly at roughly 53% accuracy. For comparison, people in our pilot study predicted categories at 60.9%. In Yew and Shamma’s YouTube study, they reported significant improvements from nominal factoring of features which are resultant of “explicit social action,” in their case, the 5-star ratings [26]. In their study, nominal factoring of the ratings bucketed numerical, interval data into discrete unordered bins (all the 4.0 videos in one set, the 4.1s in another, and the 4.3 into another and so on). This conversion from an interval to factors improved the classifier’s training and overall performance.

The Zync data does not represent explicit social annotation; it is implicit behavior observed from using a tool. Yew and Shamma’s findings demonstrated little to no performance gain from the nominal factoring of implicit usage data from YouTube. However, Zync’s data comes from two people sharing and conversing around the media, thus socially constructing its genre. As we hypothesized these data to be *purposeful* and *definitional* with regards to genre construction, we believe it to be equally predictive and explicit as Yew and Shamma defined. After nominal factorization, the Naive Bayes classifier predicted category genre with an accuracy of 82.34% (using a 80% training corpus), a performance increase of 30%. A 60% training set predicted with an accuracy of 75.32%. A full list of prediction accuracies by training set size is found on in Table 3. Category prediction based on implicit conversational data increases accuracy by 6.8% when compared to Yew and Shamma’s [26] YouTube dataset findings with an 80% training set. Furthermore, our model trained on a 60% sample (996 videos) predicts with the same accuracy as Yew and Shamma’s model trained on an 80% corpus (1,392 videos). See Table 4.

DISCUSSION

In our results and findings we have addressed two research questions: RQ1 to better understand what individual and social processes people use to categorize online videos as Comedy, and RQ2, to develop an improved method for automatically classifying online videos, taking into account social and contextual features.

Why are people so bad at classifying Comedy?

Triangulating between the results from the three phases of our study, we find that people find it hard to classify videos within the category ‘Comedy’. It is clear that what is or is not considered comedic is subjective, contextual and somewhat culturally specific. However, we also found that we were able to discern stylistic and contextual cues that signal

Accuracy	Features and Factorization
<i>Random Chance Predictions</i>	
23.0%	Sampled from YouTube Crawl
37.2%	Sampled from Zync Dataset
<i>Human Predictions</i>	
60.9%	Average from pilot survey
<i>YouTube Features/Naive Bayes predictions</i>	
14.6%	All YouTube features and categories
32.4%	All YouTube features/Top 5 Categories
75.5%	Factoring “Ratings” feature/ Top 5 categories
<i>Zync Features/Naive Bayes predictions</i>	
52.9%	All Zync Features and Categories
53.9%	All Zync Features/Top 5 Categories
82.3%	Factoring Zync features/ Top 5 categories

Table 4. Using nominally factored conversational data for classification shows a 6.8% increase in accuracy over traditional, asynchronous social media metadata and a 21.4% increase over human judgement. The chance prediction was based on random guess based on the corpus distribution of categories. Bayes predictions based on 80% training sample size. YouTube predictions reported are from Yew and Shamma [26].

something as potentially Comedy. For example, cues like background laughter functioned as signaling mechanisms, indicating that a video is a Comedy—even if for the viewer, the content was not deemed to be that funny. Given our finding that people can articulate what factors make something a Comedy (even if they personally did not find the content funny), we posed the question: why did our respondents fare so badly in classifying comedy videos in our surveys?

One explanation is that many of the videos shared on sites like YouTube do not, in fact, conform to the more culturally accepted conventions and characteristics of the Comedy genre. An informal review of videos classified as Comedy on YouTube revealed why there was far more agreement in our survey results about whether a video was funny or not than whether the video fitted in the category Comedy. People were much more likely to agree that something was funny, even if the original creator had not intended it to be funny. For example, in the course of the study, two specific videos, *Laughing baby* and *Sneezing panda*, emerged from our data as examples of videos where there was some ambiguity between the intentions of the video producer and the user who had uploaded the video. These are videos that people find funny but that were not intentionally created to be comedic; people therefore did not think they necessarily belong in the category Comedy. To be categorized as Comedy requires an assumption of comedic intention on the part of the creator of the video itself.

Use of classifications... in the wild

Unlike genre categorizations for other media channels (e.g., TV, radio), socially shared online videos are more diverse and varied in their form and content. The videos in our sample range from edited excerpts from longer films, to music videos, personal home movies, and video blogs. This variance in styles and formats makes it much harder for indi-

viduals to categorize videos from sharing sites like YouTube using definitive classifications like Comedy.

So why do individual uploaders classify their videos as Comedy? One explanation is the use of genres as a signaling mechanism for hoped-for audience orientation to the content. By categorizing a video clip as comedy, uploaders signal to potential audiences the appropriate or desired way to experience and interact around the content. A video clip that does not appear to be a comedy, but that has funny elements within it for the audience to enjoy, may be classified as Comedy, signaling that it is intended to be taken as such. The uploader of the video is using the genre classification to shape the social experience, shaping how the media content is perceived and classified. This is something that occurs during conversations or through gestural or postural cues when we watch things together, in a co-present setting.

Evolving Categories

Genres are more than just static organizational retrieval tools. Our results counter the view of genres that suggests “things come in well-defined kinds, that the kinds are characterized by shared properties, and that there is one right taxonomy of the kinds.” [11] Rather, our results are consistent with Bowker and Star’s [2] view that no classification scheme or genre category is perfect and users of classification systems often ignore the boundaries between categories or even routinely mix them up.

Our findings support the view that genres like Comedy are constantly evolving, shaped by the behavior of others and by our interactions with others around and through media content. These social interactions in turn affect our social behaviors and activities with and around that content. Viral videos like *Sneezing panda* were not intentionally created to be comedic. But having been constituted as funny, framed and shared with labels like the Comedy classification to indicate they should be read as such, drive them to be seen as humorous, and shape their trajectory through social networks as they spread and humorous memes. Even if the original intention of a video clip is not meant to be comedy, how it is viewed, interacted with, and shared shapes its classification and expands its genre. Our study supports the view that genres are social constructs, contextually defined and evolving out of people’s pragmatic interactions and activities. Therefore, for automated classification, implicit interaction which represents how media can transform context, should be utilized.

On Machine Classification

Addressing RQ2, we have found that a predictive model of genre identification can be built by using only implicit social traces that result from synchronous sharing, without using video content, metadata, or annotations. This means that social behavioral traces may be stronger and thus more predictive signals on which to train predictive models. This is a departure from the current ML approach, where explicit annotations are considered to be the most powerful features on which to train classification systems. Our results indicate that conversational activity features prove to be very effective,

Async		Sync	
Implicit	Explicit	Implicit	Explicit
View Counts	5-Star Rating	Play/Pause	Chat
Upload Count	Comments	Scrubbing	Emoticons
Video Duration	Tags	Session Duration	VoIP

Table 5. Types of data for social multimedia classification. Our classifier trained on implicit synchronous features outperformed those trained on asynchronous explicit features. Current ML techniques hold asynchronous explicit actions to be the more dominate features.

demonstrating more generally the potential for predictive classification of uncategorized content based on social interactional features; our classifier performed well with a small training set after nominal factorization of the implicit features. Table 5 shows a list of the social multimedia features which generate social and interactional traces for modeling.

Our approach also suggests diversity in classification. Not everything that is Comedy should be primarily ranked as Comedy for everyone. When building recommendation systems and new interaction experiences, there is no one-size-fits-all solution. Sharing media, not only can, but should and indeed will transform how we perceive, understand and ultimately categorize it. Therefore there is a strong role for systems that fluidly emerge categories for classification on the basis of social action and interaction.

CONCLUSION

This work contributes to a developing understanding of the experience of sharing/consuming media content online. In this article, we have shown how genre, specifically Comedy, can be modeled predictably when features are drawn from a qualitative understanding of media classification. We have demonstrated that there is more to genre and classification than just flat or hierarchical organization as embodied in post-hoc tagging by people or machines. Genre in practice is as much about how something is shared and communicated as how it fits into a pre-existing categorical schema, or genre, emerges from how we share and communicate around media. Our classifier out performs previous solutions underscoring the importance of social action and interaction in categorization, but also suggests we have uncovered a viable method for classifying online video for more effective organization and retrieval. With this approach, we can help people find things which others have found to be funny rather than things that have been prescriptively tagged as such.

ACKNOWLEDGEMENTS

We thank Judd Antin, Bryan Pardo, Lyndon Kennedy and Matt Cooper for their technical advice and comments.

REFERENCES

1. Bawarshi, A. The Ecology of Genre. *Ecocomposition: theoretical and pedagogical approaches* (2001), 69–80.
2. Bowker, G. C., and Star, S. L. *Sorting Things Out: Classification and Its Consequences*. Inside technology. The MIT Press, October 1999.

3. Cesar, P., Geerts, D., and Chorianopoulos, K. *Social Interactive Television: Immersive Shared Experiences and Perspectives*. Information Science Reference, 2009.
4. Chen, Z., Cao, J., Song, Y., Guo, J., Zhang, Y., and Li, J. Context-oriented web video tag recommendation. In *WWW '10: Proceedings of the 19th international conference on World wide web*, ACM (New York, NY, USA, 2010), 1079–1080.
5. Cooper, M. M. The ecology of writing. *College English* 48, 4 (1986), 364–375.
6. 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).
7. Erzberger, C., and Kelle, U. Making inferences in mixed methods: The rules of integration. *Handbook of mixed methods in social and behavioral research* (2003), 457–488.
8. Hölbling, G., Thalhammer, A., and Kosch, H. Content-based tag generation to enable a tag-based collaborative tv-recommendation system. In *EuroITV '10: Proceedings of the 8th international interactive conference on Interactive TV & Video*, ACM (New York, NY, USA, 2010), 273–282.
9. Kittur, A., Chi, E. H., and Suh, B. Crowdsourcing user studies with mechanical turk. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ACM (New York, NY, USA, 2008), 453–456.
10. Krippendorff, K. *Content analysis: An introduction to its methodology*. Sage Publications, Inc, 2004.
11. Lakoff, G. *Women, Fire, and Dangerous Things*. University Of Chicago Press, April 1990.
12. Liu, Y., Shamma, D. A., Shafton, P., and Yang, J. Zync: the design of synchronized video sharing. In *DUX 2007: Proceeding of the 3rd conference on Designing for User Experience* (Chicago, IL, USA, November 2007).
13. Mahajan, D., and Slaney, M. Image classification using the web graph. In *Proceedings of the International Conference on Multi-Media*, ACM (2010).
14. Miller, C. Genre as social action. *Quarterly journal of speech* 70, 2 (1984), 151–167.
15. Naaman, M., and Nair, R. Zonetag's collaborative tag suggestions: What is this person doing in my phone? *IEEE MultiMedia* 15, 3 (2008), 34–40.
16. 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* (2002), 348–354.
17. Salton, G., Wong, A., and Yang, C. S. A vector space model for automatic indexing. *Communications of the ACM* 18 (1975), 613–620.
18. Scully, M. Marge be not proud. In *The Simpsons*, 3F07. FOX, December 1995.
19. Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F. M., and Riedl, J. tagging, communities, vocabulary, evolution. In *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, ACM (New York, NY, USA, 2006), 181–190.
20. Shamma, D. A., and Liu, Y. *Social Interactive Television: Immersive Shared Experiences and Perspectives*. Information Science Publishing, Hershey, PA, USA, 2009, ch. Zync with Me: Synchronized Sharing of Video through Instant Messaging, 273–288.
21. Slaney, M., and White, W. Similarity based on rating data. In *Proc. of Int. Symposium on Music Information Retrieval*, Citeseer (2007).
22. Weinberger, K. Q., Slaney, M., and Van Zwol, R. Resolving tag ambiguity. In *MM '08: Proceeding of the 16th ACM international conference on Multimedia*, ACM (New York, NY, USA, 2008), 111–120.
23. Weisz, J. D., Kiesler, S., Zhang, H., Ren, Y., Kraut, R. E., and Konstan, J. A. Watching together: integrating text chat with video. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press (New York, NY, USA, 2007), 877–886.
24. 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*, N. Fielding, R. M. Lee, and G. Blank, Eds. Sage, London, 2008, 116 – 140.
25. Whitman, B., and Ellis, D. Automatic record reviews. In *ISMIR* (2004).
26. Yew, J., and Shamma, D. A. Know your data: Understanding implicit usage versus explicit action in video content classification. In *IS&T/SPIE Electronic Imaging* (January 2011).
27. Zhang, N., Zhang, Y., and Tang, J. A tag recommendation system for folksonomy. In *SWSM '09: Proceeding of the 2nd ACM workshop on Social web search and mining*, ACM (New York, NY, USA, 2009), 9–16.