Classification and Detection of Micro-Level Impact of Issue-Focused Documentary Films based on Reviews

Rezvaneh Rezapour
School of Information Sciences
University of Illinois Urbana-Champaign
rezapou2@illinois.edu

Jana Diesner
School of Information Sciences
University of Illinois Urbana-Champaign
jdiesner@illinois.edu

ABSTRACT
We present novel research at the intersection of review mining and impact assessment of issue-focused information products, namely documentary films. We develop and evaluate a theoretically grounded classification schema, related codebook, corpus annotation, and prediction model for detecting multiple types of impact that documentaries can have on individuals, such as change versus reaffirmation of behavior, cognition, and emotions, based on user-generated content, i.e., reviews. This work broadens the scope of review mining tasks, which typically comprise the prediction of ratings, helpfulness, and opinions. Our results suggest that documentaries can change or reinforce peoples' conception of an issue. We perform supervised learning to predict impact on the sentence level by using data driven as well as predefined linguistic, lexical, and psychological features; achieving an accuracy rate of 81% (F1) when using a Random Forest classifier, and 73% with a Support Vector Machine.

Author Keywords
Social Impact Assessment; Micro-level Impact; Review Mining; Natural Language Processing; Supervised Machine Learning;

ACM Classification Keywords
I.2.7 Natural Language Processing; H.3.1 Content Analysis and Indexing

INTRODUCTION
A recent improvement and advancement in the field of impact assessment is the consideration of the engagement of users with information products, such as scholarly publications [33, 38], and media content [31, 43]. We contribute to this line of work by developing a theoretically grounded classification schema for assessing the impact of media products on individuals. We bring this task to the domain of issue-focused documentaries, where funders and makers of films are interested in knowing how their products engage communities, impact society, and raise awareness for the issues addressed in their films [9, 22]. Social impact assessment (SIA) has been practiced for more than five decades in different sectors [2, 40]. There is also a long lineage of work on measuring impact and public opinion in academia [4, 29]. In psychology, social impact is defined as the effect of an individual or group on other people [25]. Also, impact assessment has a long tradition in the fields of environmental and political science [2, 40]. While the definition and naming of the concept of SIA may vary across fields and application domains, the goal with SIA is typically to measure, understand, and anticipate the consequences of information or events on individuals, groups, or society [25].

There are various sources of stimuli which can trigger a change or confirmation of a person's behavior, mindset, or emotions. Examples of these sources from the area of information products entail books, TV series, and films, including documentaries. Documentary films aim not only to tell a compelling story [35], but also to engage the public as well as to raise awareness about social justice issues, among other goals [9, 31]. According to George Stoney, a film pioneer and professor at New York University, "fifty percent of the documentary filmmaker's job is making the movie, and fifty percent is figuring out what its impact can be and how it can move audiences to action" [22]. Researchers at the University of Ohio conducted a case study where they compared the knowledge gained between two groups of students who watched a motion picture film versus a documentary about the same issue [31]. They found that increased awareness and knowledge were higher among the participants who saw the documentary.

The practical relevance of understanding the effects of information as represented in media products on people has motivated researchers from different fields to identify and measure the types and magnitude of these effects by using qualitative and quantitative methods [3, 4, 29, 31]. Access to user-generated as well as professionally-generated reflections on media products in the form of reviews has provided new opportunities for strategically generating and disseminating information, reaching people even in remote areas, and mapping the public opinion on various topics. With impact assessment becoming an increasingly important step for monitoring the post-production evolution

CSCW '17, February 25-March 01, 2017, Portland, OR, USA
© 2017 ACM. ISBN 978-1-4503-4335-0/17/03…$15.00.
DOI: http://dx.doi.org/10.1145/2998181.2998201

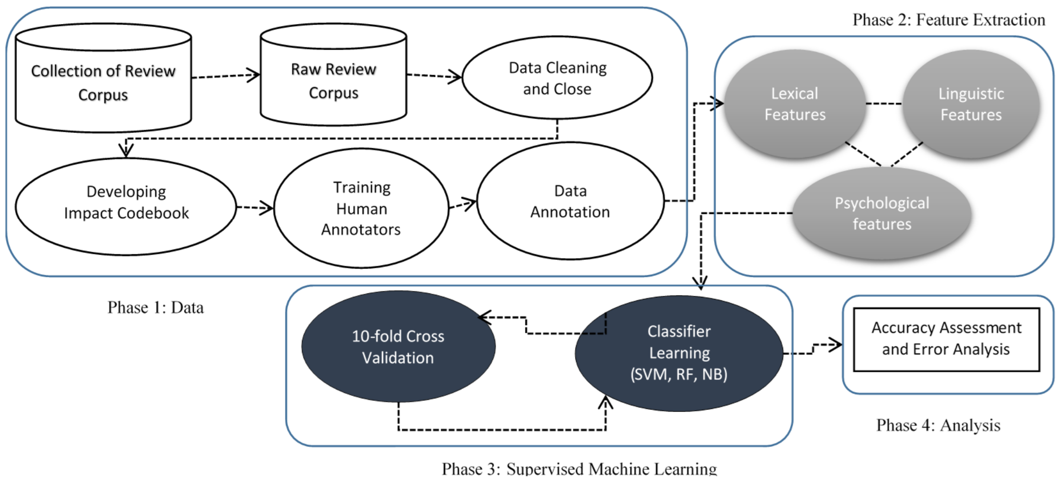

Figure 1: Project workflow and experimental design

of films [1, 9, 29], scholars and practitioners have been developing strategies for increasing the engagement of individuals (micro-level), groups and organization (meso-level), and society (macro-level) with themes and stakeholders, and measuring the effectiveness of this process. Several (normative) frameworks for this process have been proposed, but practical implementations are lagging behind [1, 4, 29]. In this paper, we address this gap by developing a computational solution for discovering impact from user-generated reviews.

In previous studies, researchers have tried to capture the impact of films on society by using techniques from network analysis and natural language processing [13, 20]. We extend this work by turning our attention to the micro-level impact of documentaries. People express their opinions on review sites, and these reviews are valuable sources of information for both commercial and research applications. Reviews may demonstrate different types of impact on a person. In fact, the act of writing a review is already an indicator of impact. Since social movements often start with engagement on the personal level, understanding the type and magnitude of this type of impact can also contribute to better model macro-level effects. In this paper, we leverage this idea and present a novel classification schema and method for measuring the impact of documentaries on people by using review data.

Traditionally, micro-level impact has been measured by conducting surveys and closed-group interviews [20, 26, 43]. These methods are limited to small groups of people as study populations. In addition, based on the nature of surveys and interviews, the questions are sometimes not broad enough or limited to closed question responses, which can lead to biased results, and may lack the explanatory details necessary to capture different levels of impact. If collected and used ethically, large-scale corpora of written accounts of user perceptions from online sources and databases allow us to overcome these limitations. As a necessary precondition for our study, we first obtained permission for collecting a corpus of user-generated documentary reviews (no personally identifiable information was collected). Using online sources also gives us more opportunities to gather data from users from different locations, ethnicities, and educational backgrounds; potentially resulting in a diverse set of opinions considered for analysis.

With the work presented in this paper, we have made the following contributions: First, we defined a categorization schema for micro-level impact based on a systematic review of different applicable bodies of literature (psychology, media studies), and close readings of samples from our data. Second, we developed a codebook for annotating reviews for these categories, and trained two individuals to apply the codebook to the data. Third, we analyzed the annotated data, selected features for training a classifier that predicts the defined impact categories (guided by prior work in review mining and our data analysis), and conducted experiments to evaluate the predictability of our impact types. We performed a detailed analysis on both results from human annotators and automatic prediction. We found that the sentence structure and tone in reviews are suitable features.

The knowledge gained with this work may be informative for future studies of impact in different fields as it allows researchers to focus on tagging and measuring micro-level impact efficiently, even for large corpora, and with relatively high accuracy. Our work might also inspire new research in review mining, which has been traditionally focused on sentiment analysis and opinion extraction, predicting ratings and helpfulness, and text summarization. Finally, the gained insights may be useful to filmmakers, funders, and outreach teams for understanding individual impact on a more fine-grained level.

The remainder of this paper is organized as follows. In the background section, we synthesize related work on both impact assessment and review mining. In section 3, we discuss the selection of data type and source used in this study, codebook creation, annotator training, data labeling, and difficulties as presented from the human coders' perspective. The methods section is then focused on dealing with imbalanced instances of classes, feature extraction, and the employed supervised learning methods. In the results section, we present the performance assessment of the classifiers, an error analysis, and statistical data analyses. Finally, we discuss our findings, limitations, and point to future directions of this research. Figure 1 summarizes the workflow of this study.

BACKGROUND

Review Mining

The large body of work in this area can be classified into three categories; 1) rating and helpfulness prediction [16], 2) summarization [16, 19, 47], and 3) opinion and polarity extraction [10-12, 28, 39, 42]. Our application (impact detection) is marginally related to the third category, i.e., opinion and sentiment analysis. In that field, researchers have tried to identify the users' opinion about (specific features of) products, and categorized the users' sentiment about an object as being for example positive, negative, or neutral [27]. Different methods have been used in this area, such as supervised and unsupervised learning techniques [16, 42], sometimes combined with ontology-based approaches [46].

Prior research on rating and helpfulness prediction has identified subjectivity or objectivity of the reviews as a useful feature for these tasks [16]. Other typically helpful characteristics for prediction include text meta-data features, e.g., the average length of sentences, lexical features, e.g., top tf-idf unigram and N-grams, and syntactic features, e.g., counts of part of speech and parse tree constituents [23, 45]. In addition, Ng and colleagues found that using top unigrams is a prominent feature for separating reviews from other types of texts [23].

The work in this paper leverages prior insights on features and training algorithms from review mining, but differs from previous studies in that we aim to detect and classify the impact of films on peoples' cognition, emotions, and behavior.

Impact Assessment and Media Effects

Impact assessment (IA) of media focuses on the influence of information on people. Prior IA of documentaries has used a variety of methods, e.g., conducting surveys, analyzing screening metrics, and applying text-mining methods to user-generated and professionally-generated reflections on films [20, 26, 43].

For example, Leiserowitz studied the impact of a Hollywood film about climate change ("The Day After Tomorrow") by quantitatively analyzing news articles before and after the release of the film, surveys, and interviews. His results showed both an impact on individuals' risk perception and an increase in the number of news articles by a factor of ten [26].

Whiteman analyzed the relationship between films and social movements [43]: He proposed a coalition model to assess the political impact of activist films and their role in social movements and public discourse by studying three successful films using interviews, participant-observation, and content analysis. His findings suggest that the new model broadens the range of impact after release.

Researchers from the John S. and James L. Knight Foundation published a report in which they made the case for using content analysis and sentiment analysis to analyze reviews written by attendees of screenings [22]. They also developed and used a new metric called "key indicator points" (KIP), which considers and employs factors such as audience, content, sustainability, and social media by monitoring websites to measure the impact of media [8].

In another study, a new set of metrics to measure reach, impact, the influence of media and engagement of the audience both online and offline was developed [21]. Researchers also used online surveys to measure the amount of knowledge that each audience could absorb [21, 37].

The Norman Lear Center in collaboration with the University of South California and the Knight Foundation are among the active research centers for finding new methods and metrics to evaluate the impact of different kinds of media. For instance, they have conducted an impact assessment study of a well-known, Oscar-nominated documentary film, "Food Inc.", where they used a combination of quantitative and qualitative methods. Based on their report, they compared two groups of individuals as viewers and non-viewers, and conducted a survey with some open-ended questions [3]. Their findings showed that the group of viewers gained knowledge and intended to change their behavior as the result of the film's message. Beside quantitative analysis, they used the answers to open-ended questions to conduct a qualitative analysis by using open coding for each answer, and reported the ratio of the perception of the viewers around the main concept of the film. The result of this study indicates the capability of films in changing people and improving societal knowledge.

As mentioned in several of these reports, website traffic data is insufficient to show or measure users' attitudes. Therefore in addition to quantitative data and techniques, there is a need for using qualitative methods and other data-mining techniques to identify the different types of impact of information products. Overall, IA of documentaries is a young and quickly evolving field. So far, basic text analysis techniques have been explored, but we argue that advanced data analytics can help to gain a more deep and comprehensive understanding of the influence of an information product on the micro, meso, and macro level [13].

DATA

Our choice of data type, i.e., reviews, is driven by our goal, i.e., measuring the impact of documentaries on individuals. Reviewers can be divided into two groups depending on whether their contributions are intrinsically motivated, which is associated with voluntarily provided or user-generated content, versus extrinsically motivated, which typically applies when people write reviews as part of their job (professionally generated content), e.g., expert film reviews. This paper is focused on the former type. As a data source, we chose to use Amazon, because their product reviews seem to attract a large population of content providers. Details about data collection, annotation, and the labeling process can be found in the following sections.

Data Collection

Based on our prior collaboration with a foundation, we chose eight documentary films related to different social justice issues: "Fed Up," "This Changes Everything," "Pray the Devil Back to Hell," "Through a Lens Darkly," "Pandora's Promise," "Solar Mamas," "The House I Live in," and "Pay to Play." After obtaining permission from Amazon for our work, we collected 2,290 reviews. The films that relate to health and healthcare (Fed Up) and environmental issues (This Changes Everything) received the highest number of reviews (1,263, 664), which may suggest that individuals connect more with problems related to their everyday life compared to other social problems, e.g., criminal justice [14]. We randomly selected 1,000 reviews for labeling to keep manual annotation manageable. Very short and very long texts were excluded. The remainder, about 870 reviews, were annotated based on our codebook, which we introduce next.

Data Annotation

Codebook Development and Annotation Schema

What types of impact can an information product have on individuals? We use a data-driven and a theoretically-grounded approach to develop a practical solution to this question.

We randomly selected a small sample of our review corpus for close reading. With the help of a linguistics student, we qualitatively and collaboratively explored types of influence reflected in reviews.

To verify and expand the set of the identified categories, we reviewed prior work from media studies and psychology [1, 15, 17, 22, 25, 26, 31, 36, 41, 43, 44]. Media can have substantial short-term and long-term influence [41]. In a study conducted on children and adolescents, it was concluded that different kinds of media, such as movies, games, advertisements, and music, have significant influence on the behavior and attitude of viewers in different age groups [41]. Media products, such as films and social media, can influence the way of thinking, social relationships, brain activity, and human identity [44]. Besides raising awareness, documentary films can have an impact on individuals, society, and policies [1]. The impact of documentary films can be direct, indirect, or cumulative. Direct impact includes changes in individuals, and cumulative impact consists of changes in groups, systems, and conditions [17]. The level of impact on individuals varies, but based on different studies, media can change the behavior, cognition, belief, attitude, and emotion of a person [1, 15, 31, 41].

We conducted a three-step procedure for developing a codebook. First, we defined six impact types: "change in cognition," "change in attitude," "change in emotion," "change in behavior," "personal opinion," and "impersonal report." We wrote a codebook with precise definitions and examples, and trained two human annotators to label 50 reviews. Once completed, we closely studied the annotations and discussed the weaknesses and shortcomings of the codebook with the annotators. Based on their feedback, we found sentences related to "change in attitude" closely related to cognition and behavioral change. We also found that, in some cases, individuals talk about their future plans to change their behavior. To address these findings, we excluded "change in attitude" from the codebook, added a new class called "intention to change" to reflect the future plans, refined the codebook accordingly, and labeled a new set of reviews. We iterated through these steps (4 times) until we were sufficiently certain that the labels were comprehensive enough to cover different types of impact. Based on this process, we found that, in some cases, people also indicate previous influences from other sources, and reaffirm prior changes or current states. We accounted for these situations in the codebook.

The final category schema has nine types of impact: change in cognition, change in behavior, intention to change, change in emotion, reaffirm cognitive state, reaffirm behavioral state, reaffirm emotional state, personal opinion, and impersonal report (summary). We further grouped these nine types into four ranks that indicate the decreasing significance of impact. The codebook contains specific definitions and example sentences (short overview in Table 1). Examples are quoted from selected Amazon reviews.

Data Labeling

A review can entail none, one, or multiple types of impact. For example, a reviewer might start with a short summary of a film, then talk about their personal opinion, and later on mention the influence of the film on their personal life. To capture all these types of impact, we decided to label the reviews on the sentence level.

	Impact Types	Definition and Examples
Rank 1	Change in Behavior	A person indicates that they have changed their lifestyle or actions after viewing a documentary; person is influenced by the movie, e.g.: “Changed my lifestyle”; “I am doing more reading nowadays”; “buying healthier alternatives”
	Change in Cognition	A person changes their beliefs or way of thinking; a person clearly indicates that they have learned something new from the documentary and/or perceive something differently as a result, e.g.: “makes a person look at a problem from a new perspective”; “I knew so little!”
	Intention to Change	A person shows interest in changing their lifestyle in the near future; person is convinced by the movie enough to want to change something, e.g.: “I plan to use…”; “within a few years, I hope to do…”
	Change in Emotion	A person indicates that they experienced an affective change because of the documentary; person reacts emotionally to the general theme of the film or topics discussed in the film, e.g.: “The issue of… made me feel…”
Rank 2	Reaffirm Behavioral State	A person indicates that their behavior after viewing a documentary remains the same; person may have been influenced by a movie or a pre-existing experience, e.g.: “That is too bad that we will never be able to do anything about it...”
	Reaffirm Cognitive State	A person indicates that their cognition/knowledge after viewing a documentary remains the same; person may have been influenced by a movie or a pre-existing experience, e.g.: “I have had my experiences, and I opted to sober up of my own volition…”
	Reaffirm Emotional State	A person indicates that their emotion(s) after viewing a documentary remain the same; person may have been influenced by a movie or pre-existing experience, e.g.: “I am sick and tired of seeing my money go to waste”; “I felt like it would be such a downer. There is no doubt that lots of this is depressing”.
Rank 3	Personal Opinion	A person expresses the general idea or opinion about a film without confirming any changes to them, person mentions other movies/ books that they find relevant, or suggests a documentary to others. The opinion can be positive or negative, e.g.: “This is an important issue and an important book”; “a must read”; “it does a good job of…”
Rank 4	Impersonal Report	Person summarizes the documentary and does not share any personal thoughts or opinions; information that the reviewer provides is from the film or addresses artistic or technical features of the film, e.g.: “the author … suggests that only national…”; “tells story of how…”; “the authors wrote in the introduction…”; “the film is executive produced by …”

Table 1: Excerpt from impact codebook

To label the sentences, we first explained the task to two annotators individually, and asked them to annotate 10 reviews based on the codebook. After getting their results, we went through each sentence, discussed the chosen categories, resolved emerging issues, and gave each annotator more data to label.

Following the example of prior work, we had 10% of the data labeled by both coders [30]. In addition to cross-annotation, we also designed three check points during the process to get feedback from the annotators, resolve any issues, and check if they still have a good understanding of the task and codebook.

Since the annotators came from different educational and cultural backgrounds, they had different interpretations of some labels. For example, one misunderstanding between the annotators was about “change in emotion.” While one annotator marked sentences with emotional words such as “love” and “like” as “change in emotion,” the other one labeled them as “general opinion.” They also found the distinction between the classes of “general opinion” and “impersonal report” somewhat confusing without having a basic knowledge about the films. We resolved these issues through a detailed discussion, and asked the annotators to revise their previous work based on their new understanding, and to then update their labeling.

To calculate the agreement between the coders, we used weighted Cohen’s Kappa because that metric is mainly designed to be used for categorical data. In the primary stage, the average inter-coder reliability was around 45%. The lowest agreement was related to reaffirmations, and the highest was related to “change in behavior”. We understand that annotating the sentences with 9 levels of impact can be a cognitively demanding task for the coders, especially in the beginning. According to the codebook (Table 1), this task requires a high level of pragmatic knowledge of tags and sentences. After discussing the misunderstandings and resolving the confusions, the average agreement increased to 97%, with the lowest being related to “personal opinion” and “change in emotion”. To achieve 100% agreement, either the codebook developer picked the final tag for the 3% disagreements, or they were excluded from the dataset. It is necessary to mention that after resolving the final issues and misunderstandings, the annotators were asked to review their tags and revise the annotated sentences accordingly. At the end, 300 reviews were checked by the codebook developer to assess the correctness of the assigned labels. Overall, the process of labeling sentences and making revisions took about 90 days.

Naturally, some sentences in the reviews do not indicate any type of impact, for example statements about experiences with delivery time or the quality of a DVD box. We labeled these sentences as “Not Applicable” (NA). They were later excluded from the data set because they have no impact weight. In some studies, NA sentences can be used as negative examples for learning, especially when building binary classifiers. We did not choose this option for our work.

Overall, we labeled 3,972 sentences. Table 2 shows the number of instances for each type of impact. Only 6% of the sentences do not feature any of our defined types of impact (NA). The majority of sentences, around 51%, are related to general opinions. We could not find any instances of “reaffirming emotional state” in the studied dataset.

Importance	Impact Types	#Sentences
Rank 1	Change in Behavior	46
	Change in Cognitive	470
	Intention to Change	77
	Change in Emotion	170
Rank 2	Reaffirm Behavioral State	22
	Reaffirm Cognitive State	48
	Reaffirm Emotional State	0
Rank 3	Personal Opinion	2,060
Rank 4	Impersonal Report	831
--	NA	248

Table 2: Number of sentences of each type of impact

METHOD

Feature Selection

As mentioned in the background section, we build our models based upon previous work. Therefore, we decided to use a combination of features suggested in the literature, namely lexical features, linguistic features, and psychological features.

Lexical features can help us to find words that are both highly salient and highly informative in a text or text set. This process also entails the removal of a) dominant (with respect to the cumulative power law distribution of word frequencies in texts) yet not content bearing words, and b) highly rare words in a collection. Linguistic features entail the consideration of relation between words and their role in a sentence, subjectively connoted adjectives and other modifiers, punctuations as determiners of sentence type (such as declarative or exclamatory), and the ratio of different parts of speech in a sentence. In Natural Language Processing, these characteristics are known to be standard features for learning. In addition to these two features, we found some specific words to be uniquely indicative of (certain types of) impact in our data, e.g., authentic words. We refer to these features as psychological features and leverage prior work to capture them. In the following section, we provide more details regarding calculating each of these features.

Lexical Features

We considered salient unigrams, bigrams, and trigrams. After preprocessing the data, removing stop words, and the words with less than five occurrences, we selected the top 450 unigrams, top 300 bigrams, and top 100 trigrams based on their tf-idf values (Eq.3).

$$TF(t) = tf(t,d) \quad (1)$$

$$Idf(t) = \log\left(\frac{|D|}{1 + |\{d: t \in d\}|}\right) \quad (2)$$

$$Tf - Idf(t) = TF * Idf \quad (3)$$

In these formulas, t is a term, d is the document in which t occurs, and D is the document space (collection of documents). Equation 1 shows the term frequency of word t, equation 2 the inverse document frequency, and equation 3 the tf-idf score calculation for term t.

Linguistic Features

We considered a) grammatical features, i.e., presence of different parts of speech, b) sentence-level information, such as number of different punctuations, and length of sentences, c) sentiment of the sentence (computed as the ratio of positive and negative words to find the polarity of a string), d) ratio of dictionary words, i.e., words that can be found in a dictionary, and function words, i.e., words with less of a lexical meaning, but importance for sentence formation, and e) time orientation of sentences, conceptualized as past, present, and future, calculated by using different verb tenses and related adverbs.

We used a combination of the Apache OpenNLP library and the “Linguistic Inquiry and Word Count” tool (LIWC 2015) [32] to extract the linguistic features. LIWC is a validated and broadly used tool, which classifies words into categories based on proprietary, embedded dictionaries. To be consistent with the outcome of LIWC, we normalized our ratios by sentence length. We ended up with 45 attributes for linguistic features.

Psychological Features

Given the nature of this study, which is focused on personal effects of information products, we used LIWC’s set of psychological features, which are compound metrics (descriptions adapted from LIWC): a) “Cognition Processes”, which are words related to causation, discrepancy, tentative, differentiation, and certainty, b) “Informal Language Markers”, such as assents, fillers, and swears words, c) “Core Drives and Needs”, such as words that are related to personal drives like affiliation, power, achievement, reward, and risk, d) "Biological Processes", which are words related to health, body, and ingestion, e) "Perceptual Processes", such as words that refer to multiple sensory and perceptual dimensions associated with the five senses, f) "Social Words", which are words related to family and friends, g) "Clout", i.e., words related to the social status, confidence, or leadership of individuals presented in the text, h) "Tone", i.e., words related to the emotional tone of the writer, which are a combination of both positive and negative sentiment terms, i) "Authentic", which are words related to the real personality of the writer, and j) "Analytical Thinking", which comes from the words reflecting the experiences and logic of the writer. Overall, we considered 45 attributes for the psychological features set provided in LIWC.

Dealing with Imbalanced Class Distributions

As shown in Table 2, the high ranking impact classes have fewer instances than ranks 3 and 4. This imbalance can bias the classifier such that ranks 3 and 4 get predicted with higher accuracy. To mitigate this problem, different approaches have been proposed. In addition to cost-sensitive learning, methods such as over sampling, under sampling, and combinations of the two have been used [5, 24]. Based on prior work, oversampling and using a combination of different techniques can result in a better outcome compared to cost-sensitive learning [7].

To balance our dataset, we used a combination of two methods: oversampling for classes with small numbers of instances, and under sampling for large classes. For the first case, we used a method called Synthetic Minority Over-Sampling Technique (SMOTE). In this method, new instances are synthetically created using the k nearest neighbors. This method has a better performance compared to oversampling with replacement [6]. According to the number of instances of each class, a range between 100 to 500% was chosen using k=5 nearest neighbors to minimize the risk of over-fitting the classifiers. After oversampling and randomizing the data, we used random undersampling with the ratio of 9:1 to reduce the size of the large classes. These algorithms were implemented using WEKA. Table 3 shows the new number of instances after balancing the dataset. As shown in the table, the difference between the instances is minimized.

Importance	Impact Types	After Balancing
Rank 1	Change in Behavior	276
	Change in Cognitive	940
	Intention to Change	462
	Change in Emotion	850
Rank 2	Reaffirm Behavioral State	110
	Reaffirm Cognitive State	288
	Reaffirm Emotional State	0
Rank 3	Personal Opinion	990
Rank 4	Impersonal Report	831

Table 3: Number of sentences of each type of impact after balancing

Classification

To classify the sentences, we decided to use three different learning algorithms: Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB). We implemented the classifiers using WEKA [18] and conducted 10-fold cross validations. We compare performance in terms of accuracy.

To find the best combination of features, we 1) built a baseline model using the unigrams, 2) added bigrams, and 3) added trigram to complete the linguistic features. We then 4) added psychological features, and 5) linguistic features separately to the linguistic features. Finally, we 6) combined all three feature types.

Before classifying the sentences, we chose and ranked the best attributes using Information Gain (Eq.4) [34]. This algorithm was also implemented using WEKA.

$$InfoGain(Class, Attribute) = P(Class) - P(Class|Attribute) \quad (4)$$

For assessing prediction accuracy, we used the standard metrics of precision, recall, and F-score (with $\beta = 1$). The results for each feature and classifier are listed in Table 4.

Features		SVM			Random Forest			Naïve Bayes		
		Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Lexical	Unigram (Baseline)	53.3	46.4	47.3	63.8	61.0	61.3	50.9	49.2	49.3
	Unigram+Bigram	57.4	51.2	52.5	67.4	64.7	65.0	55.2	53.1	53.1
	Unigram+Bigram+ Trigram	57.3	51.5	52.7	67.7	65.2	65.3	56.1	54.4	54.3
Lexical + Psychological		71.0	70.6	70.6	80.2	79.2	79.5	55.2	52.8	52.5
Lexical + Linguistic		72.7	72.5	72.5	**81.4**	**80.8**	**81.1**	**64.4**	**64.1**	**63.0**
Lexical + Psychological + Linguistic		**73.0**	**73.1**	**73.0**	80.5	79.9	80.2	58.6	56.9	56.4

Table 4: Result of three classifiers using 10-fold cross validation (highest value per column in bold)

	Change Behavior	Change Cognition	Change Emotion	Intention Change	Reaffirm Behavior	Reaffirm Cognitive	Personal Opinion	Impersonal Report
Solar Mamas	**16.0**	0	**16.0**	0	0	0	48.0	20.0
Fed Up	2.4	**19.78**	4.81	1.58	**1.37**	1.92	49.31	18.82
The House I Live In	0.61	10.0	5.91	2.73	0	1.36	58.18	21.06
Pray the Devil Back to Hell	0	9.0	3.32	1.9	0	0	45.97	39.81
Pandora's Promise	0	8.64	0.82	1.23	0	**2.06**	63.79	23.46
This Changes Everything	0	7.01	4.24	2.37	0.2	0.59	**63.97**	21.42
Pay 2 Play	0	6.35	4.76	**4.76**	0	0	38.1	46.03
Through a Lens Darkly	0	1.89	3.77	3.77	0	0	41.51	**49.06**

Table 5: Different types of impact across each film (values are percent, the highest value of each column is highlighted)

RESULT

Class Distribution

Based on the labeled data, we found that 51% of the sentences contain general opinions, 20% provide summaries, and 6% do not contain any impact types (Table 2). Overall, approximately 20% of the sentences in our corpus feature emotional, cognitive, and behavioral impact (change or reaffirmation). This finding supports our effort to build a classifier that enables the detection of more fine-grained levels of micro-level impact of information products.

As the number of instances for each class shows (Table 2), the ratio of "intention to change" is higher than "change in behavior," which suggests that people are more prone to plan to change their course of action or way of thinking than actually implementing these changes. Both "change in cognition" and "change in emotion" have the highest number of instances. Also, "reaffirming emotional state" has no instances – in contrast to the other two types of impact in rank 2, which may indicate that individuals may seldom feel (or be motivated to express) a confirmation of emotional states compared to their cognitive and behavioral states.

In addition to the comparative ratio of each impact type, we also analyzed the amount of different types of impact across each film to find out to what level a film moved and motivated individuals (Table 5). Interestingly, we found that "Solar Mamas", a film about women, education, and mitigating poverty, changed the behavior rather than the cognition of reviewers. After reading the labeled sentences, we found that people stated that they had donated money to charitable organizations, which indicates a positive influence of the film. Similarly, we found that "Solar Mamas" and another film; "The House I Live In", which is related to minimum mandatory sentencing, affected individuals' emotions more than other films did. "Fed Up", a film related to health, sugar, and taxes, changed viewers' cognition and behavior. People also indicated more reaffirmation of behavioral and cognition states in the aforementioned film compared to the other considered

Lexical	**Unigram (baseline)**	change, food, sugar, people, movie, documentary, film, hope, eat, climate, years, war, healthy, book, life, watch, sad, real, nuclear, industry
Lexical	**Unigram+Bigram**	change, food, people, movie, hope, film, documentary, sugar, this movie, kids, eat, years, book, climate, war, life, i hope, sad, problem
Lexical	**Unigram+Bigram+Trigram**	change, food, movie, kids, people, this movie, hope, film, sugar, years, documentary, eat, problem, war, perspective, book, climate, i think, life, i hope
Lexical + Psychological		tone, clout, analytical thinking, biological process, discrepancy, ingestion, social words, authentic, relativity words, causation, tentative, differentiation, people, insight words, drives words, perceptual processes
Lexical + Linguistic		1st person singular , negative words, personal pronouns, overall sentimental words, focus on past, articles, all pronouns, length of sentence, verb, dictionary words, focus on future, positive words, change, people, function words, adjectives, food, adverbs
Lexical + Psychological + Linguistic		1st person singular, tone, clout, personal pronouns, negative words, analytical thinking, total sentimental words, length of sentence, focus on past, all pronoun, discrepancy, articles, verbs, biological process, social words, function words, anxiety words, focus on future

Table 6: Most informative attributes of each feature set (top 20 or less)

a	b	c	d	e	f	g	h	
58.7	26.0	0.7	8.5	0.6	4.2	1.2	0.0	**a =Impersonal Report**
19.3	**57.2**	1.7	11.3	1.1	5.3	3.6	0.5	**b =Personal Opinion**
0.7	5.4	**92.0**	1.8	0.0	0.0	0.0	0.0	**c = Change in Behavior**
7.7	17.1	1.9	**66.6**	1.5	2.4	2.3	0.4	**d = Change in Cognition**
1.0	8.0	0.0	2.8	**87.5**	0.7	0.0	0.0	**e = Reaffirm Cognitive state**
1.8	6.0	0.0	2.6	0.4	**89.1**	0.2	0.0	**f = Change in Emotion**
0.6	5.8	0.0	2.2	0.0	0.2	**91.1**	0.0	**g = Intention to Change**
0.0	2.7	0.0	2.7	0.0	0.0	0.0	**94.5**	**h = Reaffirm Behavioral State**

Table 7: Confusion matrix of SVM classifier (values are percent)

movies. Overall, when compared to affecting change in cognition, fewer films could change the behavior of reviewers. This finding indicates that a) not every film is capable and/or aims to change the behavior on the micro-level, and b) that it is difficult to change peoples' behavior. For instance, for "This Changes Everything," a film related to capitalism and environment, "change in cognition" is more desired than "change in behavior." In contrast to that, in "Fed Up," one would like to see both. These findings are shown in Table 5.

In summary, we found that information products can change individuals' perception of social justice problems, raise awareness in society, and move people to act. These findings are aligned with the results obtained by others, such as the Norman Lear Center [3], where researchers interviewed people and used quantitative analysis to identify micro-level impact. This shows that our codebook and classification algorithms can capture some dimensions of the impact of information products.

Classification

As shown in Table 4, we first created a baseline model by using the top salient unigrams. This baseline is needed to enable the assessment of the influence of added features on the models. The best performance with the baseline was achieved with the RF classifier. Adding in bigrams and trigrams increased the performance of all three classifiers by around 5% (for all three accuracy metrics).

Combining lexical (salient top unigrams, bigrams, and trigrams) and linguistic features further boosted the performance of all three classifiers. As shown in Table 4, accuracy increased by approximately 10-15%.

Adding psychologically connoted terms to the set of lexical features also resulted in a considerable jump in the performance of SVM and RF. All metrics for these classifiers increased by nearly 15%. However, for NB, adding the psychological features led to a drop in performance by roughly 2%.

Finally, the combination of all three set of features improved the performance of SVM. However, the performance of NB and RF slightly decreased compared to the performance with the combination of lexical and linguistic features.

Overall, RF outperformed SVM and NB when using lexical plus linguistic features with an overall F1-score of 81%. However, SVM benefitted the most from combining all three feature sets with a final F1-score of 73%.

In the following section, we analyzed the performance of the classifiers deeper by 1) examining the top attributes for each feature type, and 2) conducting an error analysis.

Feature Analysis

To identify the most contributing attributes of each feature, we calculated information gain to rank the attributes (Eq.4). The up to 20 most informative attributes per feature class are listed in Table 6.

As shown in Table 6, the best attributes of the lexical features come from the unigrams. Bigrams are rare in that set, and trigrams do not feature there.

The combination of lexical and psychological features mostly benefitted from attributes of the latter one. Clout, tone, and analytical thinking are the top attributes, while the presence of lexical features is limited to one word, namely "people." However, this set is joined by "change" and "food" in the combination of lexical and linguistic features. With respect to psychological features, 1st person singular pronouns ("I"), sentiment words, pronouns, and time orientation of the sentences had a significant role in both, the lexical and linguistic set, and the lexical plus linguistic plus psychological set. Finally, the consideration of all features benefitted from the combination of top psychological and linguistic features, where attributes of the latter set are more highlighted than the former one.

Based on these findings, we conclude that using linguistic and psychological features was beneficial for this task. As the analysis of the top informative attributes has shown, the structure of the sentences, grammatical indices, subjective words, and the tone of sentences are useful for predicting the impact.

Error Analysis

In addition to analyzing the contribution within and among features classes, we also studied the confusion matrix of the classifiers to find patterns in misclassifications. We chose the confusion matrix of the SVM because of its comparatively higher accuracy scores when using all sets of features. Table 7 shows the classified instances per impact category. As this matrix shows, "impersonal report," "personal opinion", and "change in cognition" are the most misclassified categories, where the first two classes have the lowest accuracy rate and the highest number of wrongly predicted instances, i.e., they are least orthogonal to other classes and/ or least predictable with the features we used. In fact, the highest error for these two classes comes from predicting the two other class. This finding is consistent with the feedback from our human annotators, who found it hard to distinguish "personal opinion" from "impersonal report" without prior knowledge about a given film. After studying sentences in these two classes, we found them to be very similar to each other in sentence structure and lexicon use. The overlap occurs in cases where people tend to agree with a concept in the film or want to add their own ideas to the concept.

Furthermore, "change in cognition" has been misclassified as "personal opinion" more often when compared to the other high-ranked impact categories. To further analyze this problem, for each of these classes, we randomly selected 30 sentences from different reviews, took them out of their original contexts, removed the labels, and asked the human annotators to label them again. Table 8 shows the underlying ground-truth result of the misclassified sentences labeled by the human annotators. The first column shows the original class types, column two provides the new labels chosen by the human annotators, and column three indicated the ratio of the given labels. From this case study, we see that human coders make similar mistakes like the classifier. This finding, which is also consistent with the confusion matrix, shows that some sentences are, in nature, hard to categorize, and more pragmatic analysis might be needed to solve this problem. Based on our discussion with the human annotators, we found that being able to see preceding sentences in a review and familiarity with the content of the film would lower these errors.

DISCUSSION

In this study, we developed a theoretically grounded and data driven classification schema, related codebook, corpus annotation, and prediction model for detecting multiple types of impact of documentaries (as a specific instance of information product) on individuals based on user-generated content (reviews).

Our analysis of a set of reviews showed that information products can change peoples' conception of an issue, and can be associated with changes in attitudes toward societal problems. This finding is a meaningful outcome for sponsoring organizations, such as foundations, and filmmakers, as it demonstrates the potential impact of documentary films, and highlights the importance of assessing impact beyond frequency metrics.

The data annotation and analysis procedures also showed that user-authored reviews contain or represent different types of impact, which justifies the development of a classification schema of micro-level impact types as well as the suitability of using reviews as a data source for studying impact.

Initial Tags	Secondary Tags	Ratio
Personal Opinion	Personal Opinion	68%
	Change in Cognition	16%
	Impersonal Report	16%
Change in Cognition	Change in Cognition	36%
	Personal Opinion	54%
	Impersonal Report	10%
Impersonal Report	Impersonal Report	53%
	Personal Opinion	46%

Table 8: Error analysis: example for misclassified instances and human annotation

To identify and define impact types, and generate a codebook, we used a combination of reviewing prior work from media studies and psychology on the effects of print and social media on individuals, and qualitative exploration through close reading techniques by an interdisciplinary team that included a linguist. Our resulting categorization schema is composed of four levels: (1) (intent to) change and (2) reaffirmation in cognition, behavior, and emotions, as well as (3) personal opinions, and (4) impersonal reports (Table 1). Around 20% of the sentences in our corpus indicate high impact (type 1 and 2), 6% do not contain any impact type considered herein, and 74% show lower levels of user engagement (types 3 and 4) (Table 2). Sentences of types 3 and 4 are often the focus of review mining studies that aim to predict ratings and sentiment. Our work builds upon and expands this line of research by separating impact into practically relevant and theoretically supported types.

To build classifiers, we worked with three sets of features: lexical, linguistic, and psychological ones. We trained three commonly used types of classifiers, i.e., Support Vector Machines, Random Forest, and Naïve Bayes. We first built a baseline model using top unigrams, gradually added the other feature types, and measured the incremental contribution of each type. The classification results (Table 4) showed that the combination of all three sets of features was most beneficial for SVM, where it improved the performance from 51% (baseline) to 73% final model (F1 score). The Random Forest classifier outperformed the other two training algorithms, and achieved the best overall performance, but did so by using only a combination of lexical and linguistic features (from 63% for baseline to 81%). Naïve Bayes also performed best with a combination of lexical and linguistic features only, however, its score for F1, recall, and precision was lower than those for the Random Forest. The comparison of the top attributes of each set revealed that using informative attributes from the linguistic and psychological feature sets were helpful in building impact prediction model (Table 6). We also conducted an error analysis of misclassified instances, finding that sentences related to "personal opinion" and "impersonal report" are very similar to each other in structure and lexicon use, which made labeling challenging for the classifier (Table 7). Distinguishing these two types of impact was also challenging for humans, especially when respective sentences were presented out of context, and these difficulties carry through to the labeling and learning steps (Table 8).

In contrast to similar research in the field of review mining, where it is a common goal to identify user opinions about products, we categorized and based on that predicted different types of impact that an information product can have on individuals with relatively high accuracy. The findings from this work can advance review mining research by introducing a classification schema for micro-level impact assessment.

Our outcomes may also be informative for sponsors, makers, and producers of documentaries as we provide a detailed yet comprehensive understanding of citizen engagement with issue-focused films. This might offer support in developing strategies for improving user engagement, and raising awareness for social justice issues. As shown in Table 5, the proposed impact codebook is helpful for formalizing and exemplifying a documentary film's various types of influence. As mentioned, some films can influence people to change their behavior and take action, e.g., by donating money or supporting a movement, while other films aim to raise awareness and change the cognition of (re)viewers. Our findings might also help social movements to better understand the kind of impact that outreach work on certain topics can have.

The potential future contributions of our codebook and classifiers are not limited to finding the impact of information products. These tools can also broaden our understanding of an individual's interactions with online communities, and the impact of the information products on individuals' everyday lives. Respectively, researchers and practitioners from different application domains of impact assessment can leverage our codebook to find the influence of policies or projects on the micro-level in their contexts. Our codebook can be domain-adapted and expanded to be applicable in other sectors. In addition, in the era of Big Data, gaining better knowledge of online reviews can be useful to both academia and the corporate sector.

CONCLUSION, LIMITATIONS, AND FUTURE WORK

The outcomes of this work confirm that documentary films can have different types of impact on individuals, and that these types can be identified from reviews. The developed codebook can advance research in review mining such that these types of impact can also be considered or be used as features. Our work might also improve research and methodology on impact assessment in different fields, from environmental studies to economics, in two ways. First, by advancing our knowledge about micro-level impact. Second, by increasing our understanding of the different types and magnitude of influence that various products or themes can have on individuals.

To study impact on the micro-level, we used online reviews instead of surveys and interviews. One advantage with this data source is a possible reduction in bias in comparison to results based on analyzing (text) data obtained through questionnaires and surveys. However, we do not know whether a review was only based on the impression that a person got from watching a film (if they watched it at all), and / or also by other information sources. In the future, it would be insightful to compare the types of impact that can be identified from interviews and surveys (offline sources) to those found in reviews (online sources, subject of this paper).

In addition, we have limited our study to exclusively finding the impact of documentary films, which often aim to raise awareness and affect the behavior, knowledge, or opinion of viewers. However, in addition to succeeding at the box office, some motion pictures might have similar goals. In our future work, we plan to identify and compare the types of impact of documentaries versus motion picture films on the same topics on the micro-level.

Our proposed method is not complete and has some shortcomings, which we plan to address in the future. Both humans and the predictor had difficulties with distinguishing two of the classes when labeling was done on a sentence level out of the review context. This problem will be further explored in our future work by using deeper linguistic techniques, such as pragmatic and deep syntactic analysis.

Another challenge that we faced with this project was understanding and implementing regulations, (local) norms, and (cultural) expectations for accessing, collecting, and using review data in a lawful and ethical manner. The fact that some of these data are publicly available does not necessarily mean that one has permission to collect and analyze them. We obtained permission from Amazon for this process, but cannot share our corpus due to regulatory reasons and terms of service. However, the categorization schema can be used and further tested by others. We also plan to release the codebook.

In our future work, we will analyze the impact of documentary films on the meso and macro level to find the impact of information products on social movements and legislation.

ACKNOWLEDGMENT

This work was supported by the FORD Foundation, grant 0155- 0370, and by a faculty fellowship from the National Center of Supercomputing Applications (NCSA) at UIUC. We are grateful to Amazon for giving us permission to collect reviews from their website. We thank Professor Corina Roxana Girju from the Linguistics department at UIUC for her helpful insights and advise for developing the codebook. We also thank Ming Jiang, Sandra Franco, Harathi Korrapati, and Julian Chin from the iSchool at Illinois for their help with this paper.

REFERENCES

1. Barrett, D. and Leddy, S., 2008. *Assessing creative media's social impact.* The Fledgling Fund.
2. Becker, H., A., 2001. Social impact assessment. *European Journal of Operational Research*, *128*, 2, 311-321.
3. Blakley, J., Huang, G., Nahm, S., and Shin, H., *Changing appetites & changing minds: Measuring the impact of "Food, Inc.".* The USC Annenberg Norman Lear Center.
4. Chattoo, C.B. and Das, A., 2014. *Assessing the social impact of issues-focused documentaries: research methods & future considerations.* Center for Media and Social Impact, School of Communication at American University.
5. Chawla, N.V., 2005. Data mining for imbalanced datasets: An overview. In *Data Mining and Knowledge Discovery Handbook* Springer US, 853-867.
6. Chawla, N.V., Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 321-357.
7. Chawla, N.V., Japkowicz, N., and Kotcz, A., 2004. Editorial: Special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, *6*, 1, 1-6.
8. Chinn, D., 2011. *Measuring the online impact of your information proje ct.* John S. and James L. Knight Foundation ,FSG.
9. Clark, J. and Abrash, B., 2011. *Social justice documentary: Designing for impact.* School of Communication at American University.
10. Cui, H., Mittal, V., and Datar, M., 2006. Comparative experiments on sentiment classification for online product reviews. In *Proceedings of the 21st International Conference on Artificial Intelligence, (AAAI'06).* (Boston, Massachusetts), 1265-1270.
11. De Albornoz, J.C., Plaza, L., Gervás, P., and Díaz, A., 2011. A Joint model of feature mining and sentiment analysis for product review rating. In *Advances in Information Retrieval, 33rd European Conference on IR Research* Springer, Dublin, Ireland, 55-66.
12. Devitt, A. and Ahmad, K., 2007. Sentiment polarity identification in financial news: A cohesion-based approach. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, (ACL'07).* (Prague, Czech Republic), ACL, 25-27.
13. Diesner, J. and Rezapour, R., 2015. Social computing for impact assessment of social change projects. In *Social Computing, Behavioral-Cultural Modeling, and Prediction, 8th International Conference , SBP,* Springer International Publishing, Washington, DC, 34-43.
14. Diesner, J., Rezapour, R., and Jiang, M., 2016. Assessing public awareness of social justice documentary films based on news coverage versus social media. In *Proceedings of the iConference* (Philadelphia, PA, USA).
15. Gentile, D., A. and Sesma, A., 2003. Developmental approaches to understanding media effects on individuals. In *Media Violence and Children*, 19-37.
16. Ghose, A. and Ipeirotis, P.G., 2011. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*, 10, 1498-1512.
17. Green, D. and Patel, M., 2013. *Deepening engagement for lasting impact: A framework for masuring media performance and results.* John S. and James L. Knight Foundation and Bill & Melinda Gates Foundation.
18. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H., 2009. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, *11*, 1, 10-18.
19. Hu, M. and Liu, B., 2004. Mining and summarizing customer reviews. In *Proceedings of the Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (KDD '04)*, 168-177.
20. John, S., James, L., Knight Foundation, and FSG, 2011. *Impact: A practical guide for evaluating community information projects.*
21. Kania, J. and Kramer, M., 2011. Collective impact Stanford Social Innovation Review, 36-41.
22. Karlin, B. and Johnson, J., 2011. Measuring impact: The importance of evaluation for documentary film campaigns. *M/C Journal*, *14*, 6.
23. Kim, S.M., Pantel, P., Chklovski, T., and Pennacchiotti, M., 2006. Automatically assessing review helpfulness. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP*, ACL, 423-430.
24. Kotsiantis, S., Kanellopoulos, D., and Pintelas, P., 2006. Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, *30*, 1, 25-36.
25. Latané, B., 1981. The psychology of social impact. *American psychologist*, *36*, 4, 343-356.
26. Leiserowitz, A., A., 2004. Day after tomorrow: study of climate change risk perception. *Environment: Science and Policy for Sustainable Development*, *46*, 9, 22-39.
27. Liu, B. and Zhang, L., 2012. A survey of opinion mining and sentiment analysis. In *Mining Text Data* Springer, 415-463.
28. McAuley, J. and Leskovec, J., 2013. Hidden factors and hidden topics: Understanding rating dimensions with review text. In *Proceedings of the 7th ACM Conference on Recommender Systems, (RecSys'13).* ACM, 165-172.
29. Napoli, P., 2014. *Measuring media impact: An overview of the field.*

30. Neuendorf, K.A., 2002. *The Content Analysis Guidebook*. Sage.
31. Nisbet, M.C. and Aufderheide, P., 2009. Documentary film: Towards a research agenda on forms, functions, and impacts. *Mass Communication and Society*, *12*, 4, 450-456.
32. Pennebaker, J.W., Boyd, R.L., Jordan, K., and Blackburn, K., 2015. The Development and Psychometric Properties of LIWC2015, UT Faculty/Researcher Works.
33. Priem, J., Taraborelli, D., Groth, P., and Neylon, C., 2010. Altmetrics: A manifesto.
34. Roobaert, D., Karakoulas, G., and Chawla, N.V., 2006. Information gain, correlation and support vector machines. In *Feature Extraction* Springer, Berlin Heidelberg, 463-470.
35. Rose, F., 2012. *The art of immersion: How the digital generation is remaking Hollywood, Madison Avenue, and the way we tell stories*. WW Norton & Company, New York, NY.
36. Scheufele, D.A. and Tewksbury, D., 2007. Framing, agenda setting, and priming: the evolution of three media effects models. *Journal of Communication*, *57*, 1, 9-20.
37. Schiffrin, A. and Zuckerman, E., 2015. *Can We Measure Media Impact? Surveying the Field.* Stanford Social Innovation Review.
38. Thelwall, M., Haustein, S., Larivière, V., and Sugimoto, C.R., 2013. Do altmetrics work? Twitter and ten other social web services. *PloS One*, *8*, 5, e64841.
39. Turney, P.D., 2002. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, (ACL'02).* Association for Computational Linguistics, 417-417.
40. Vanclay, F., 1999. Social impact assessment. *Environmental Assessment in Developing and Transitional Countries*, 125-135.
41. Villani, S., 2001. Impact of media on children and adolescents: a 10-year review of the research. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*, 4, 392-401.
42. Wang, H., Lu, Y., and Zhai, C., 2010. Latent aspect rating analysis on review text data: A rating regression approach. In *Proceedings of the 16th ACM International Conference on Knowledge Discovery and Data Mining, (KDD'10).* ACM, 783-792.
43. Whiteman, D., 2004. Out of the theaters and into the streets: A coalition model of the political impact of documentary film and video. *Political Communication*, *21*, 1, 51-69.
44. Zeitel-Bank, N. and Tat, U., 2014. Social Media and Its Effects on Individuals and Social Systems. In *Proceedings of the Human Capital without Borders: Knowledge and Learning for Quality of Life; Proceedings of the Management, Knowledge and Learning International Conference 2014*, ToKnowPress, 1183-1190.
45. Zhang, Z. and Varadarajan, B., 2006. Utility scoring of product reviews. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management, (CIKM '06)*, ACM, 51-57.
46. Zhao, L. and Li, C., 2009. Ontology based opinion mining for movie reviews. In *Knowledge Science, Engineering and Management: Third International Conference, (KSEM '09).* Springer, Vienna, Austria, 204-214.
47. Zhuang, L., Jing, F., and Zhu, X.-Y., 2006. Movie review mining and summarization. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management, (CIKM '06).* ACM, 43-50.