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

School of Information Sciences
The iSchool at Illinois

Rezvaneh Rezapour, Jana Diesner

School of Information Sciences, University of Illinois at Urbana-Champaign

Problem Statement and Contribution

- Open questions in the field of impact assessment of information products:
  - What types of impact can an information product have on individuals?
  - Can this impact be reliably measured and predicted from user-generated text data?
- Our contributions – we developed:
  - A theoretically grounded classification schema for micro-level impact.
  - A codebook and annotation schema for labeling data.
  - Identifying text level features that are indicative of these types of impact.
  - A probabilistic model for predicting the identified types of impact based on user-generated reviews of documentaries.

Method

Overall workflow (see also Figure 1):

1. Obtain permission from data provider.
2. Collect corpus of user-generated reviews of issue-focused documentaries.
3. Develop categorization schema for micro-level impact based on a) systematic review of literature from different fields (psychology, media studies, journalism, communication) and b) close reading of samples from our data.
4. Develop codebook for annotating reviews (Table 1) .
5. Train two human coders to apply codebook to the data. Iterative annotation process until inter-coder reliability was sufficiently high.
6. Analyze the annotated data. Select features. Train a classifier that predicts the outlined impact categories. Assess accuracy. Conduct errors analysis.

Data

We collected (with permission from data provider) 2,290 user-generated reviews from eight issue-focused documentaries: “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.”

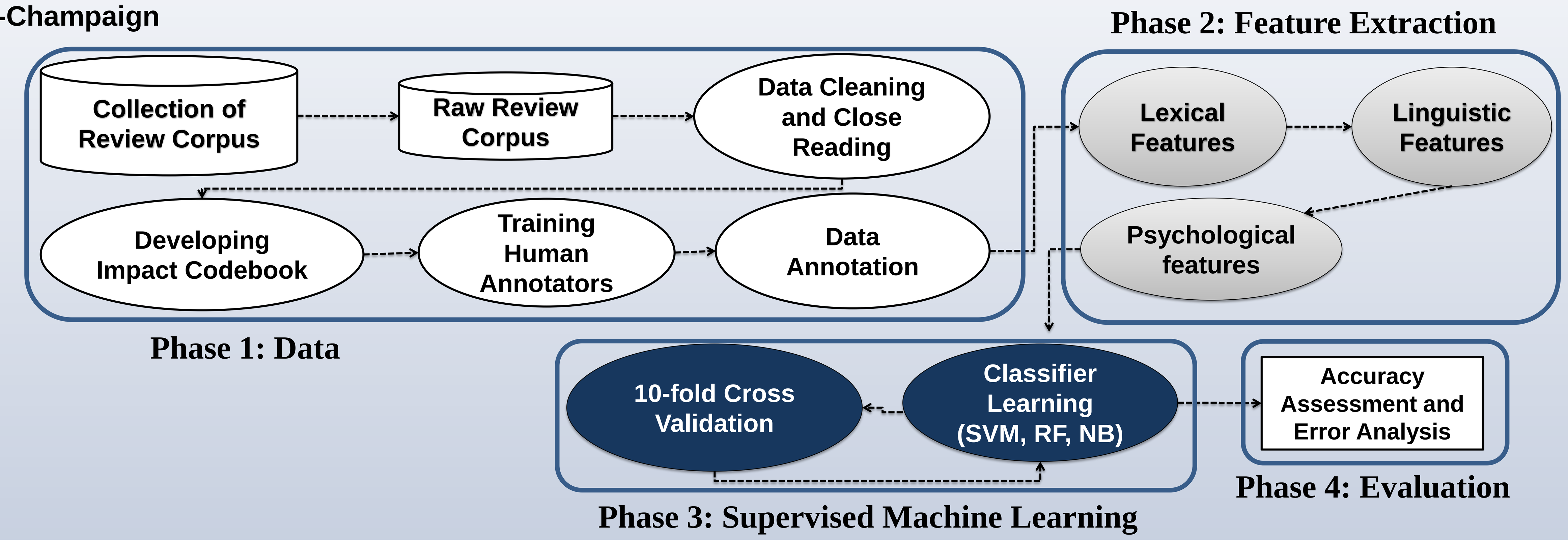

Figure 1: Experimental design and project workflow

Annotation Schema

Codebook Development and Annotation

- **Table 1**: **Types of Impact of Media on Individuals**

	Impact Types	Definition
Rank 1	Change in Behavior	A person indicates that they have changed their lifestyle after viewing a documentary; person is under the influence of the movie, e.g.: “Changed my lifestyle”
	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 it differently as a result, e.g.: “makes a person look at a problem from a new perspective”
	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, e.g.: “I plan to use…”
	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 under the influence of a movie or 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 under the influence of a movie or 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 remains the same; person may have been under the influence of a movie or pre-existing experience, e.g.: “I am sick and tired of seeing my money go to waste”
Rank 3	Personal Opinion	A person express the general idea about the film without confirming any changes in them, person mentions other movies that they find relevant; suggests documentary to others. The opinion can be positive or negative, e.g.: “This is an important issue and an important book”
Rank 4	Imperso-nal Report	Person summarizes the documentary and does not share any personal thoughts or opinions; information reviewer provides is from the film; artistic or technical features of the movie, e.g.: “the author … suggests that only national…”; “tells story of how…”; “the authors wrote in the introduction…”

Features and Learning

- ***Lexical Features***
  - Top 450 unigrams, top 300 bigrams, and top 100 trigrams based on their tf-idf values

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

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

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

- ***Linguistic Features***
  - Grammatical features, Sentence-level information, Ratio of punctuations, length of a sentence, Sentiment of a sentence, Ratio of dictionary words and function words, Time orientation of the sentences using different verb tenses and related adverbs
- ***Psychological Features***
  - Cognition Processes, Informal Language Markers, Core Drives and Needs, Biological Processes, Perceptual Processes, Social Words, Clout, Tone, Authentic, Analytical Thinking

Learning

- **Classifiers:**
  - SVM, Random Forest (RF), and Naive Bayes (NB). (implemented using WEKA)
- 10-fold cross validation

Attribute Selection

- **Information Gain**

$$InfoGain(CLass, Attribute) = P(Class) - P(Class|Attribute)$$

Dealing with Imbalanced Class Distribution

- “Synthetic Minority Over-Sampling Technique” (SMOTE) to increase the size of small classes.
- “Random Undersampling” with the ratio of 9:1 to reduce the size of large classes.

Table 2 : Number of Instances Before and After Balancing

Importance	Impact Types	Before Balancing	After Balancing
Rank 1	Change in Behavior	46	276
	Change in Cognitive	470	940
	Intention to Change	77	462
	Change in Emotion	170	850
Rank 2	Reaffirm Behavioral State	22	110
	Reaffirm Cognitive State	48	288
	Reaffirm Emotional State	0	0
Rank 3	Personal Opinion	2060	990
Rank 4	Impersonal Report	831	831

Results

Table 3: Classification Result for SVM, RF, and NB

Features		SVM			Random Forest			Naïve Bayes		
		P	R	F1	P	R	F1	P	R	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

$$Precision = \frac{TP}{TP+FP}, \quad Recall = \frac{TP}{TP+FN}, \quad F1 = \frac{2 \times Precision \times Recall}{Precision+Recall}$$

Table4: Ratio of Types of Impact per Film

	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**

Conclusions

- Information products, namely issue-focused documentaries, can impact peoples’ perception of an issue. This impact can be associated with changes in the understanding and attitudes toward societal problems.
- Micro-level impact on behavior, cognition, emotions and opinions can be predicted from user-generated text data.
- The developed codebook advances research in review mining, e.g., by enabling the extraction of different types of impact from reviews.
- This work advances the field of social impact assessment by providing a data driven, computational and probabilistic solution to identifying and classifying impact.

Acknowledgments

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

Reference –poster is based on the following paper:

Rezapour, R., & Diesner, J. (2017). Classification and Detection of Micro-Level Impact of Issue-Focused Films based on Reviews. In *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, ACM.