Verbal Behaviors and Persuasiveness in Online Multimedia Content

Moitreya Chatterjee, Sunghyun Park*, Han Suk Shim*,
Kenji Sagae and **Louis-Philippe Morency**
USC Institute for Creative Technologies
Los Angeles, CA 90094
`metro.smiles@gmail.com,`
`{ park, hshim, sagae, morency }@ict.usc.edu`

Abstract

Persuasive communication is an essential component of our daily lives, whether it is negotiating, reviewing a product, or campaigning for the acceptance of a point of view. With the rapid expansion of social media websites such as YouTube, Vimeo and ExpoTV, it is becoming ever more important and useful to understand persuasiveness in social multimedia content. In this paper we present a novel analysis of verbal behavior, based on lexical usage and paraverbal markers of hesitation, in the context of predicting persuasiveness in online multimedia content. Toward the end goal of predicting perceived persuasion, this work also explores the potential differences in verbal behavior of people expressing a positive opinion (e.g., a positive movie review) versus a negative one. The analysis is performed on a multimedia corpus of 1,000 movie review videos annotated for persuasiveness. Our results show that verbal behavior can be a significant predictor of persuasiveness in such online multimedia content.

1   Introduction

A message that is "intended to shape, reinforce or change the responses of another or others" is categorized as *persuasive communication* (Miller, 1980), and it is particularly important for the role it plays in creating social influence and altering other people's opinions (Reardon, 1991; Zimbardo and Leippe, 1991). For instance, a persuasive advertisement could be a potential profit churner.

   The growth of social networking sites on the Internet has resulted in an explosion of online content with the purpose of delivering persuasive messages. Websites such as YouTube, Vimeo and ExpoTV are examples of online media in which these messages propagate mainly in the form of videos. ExpoTV, in particular, is a repository of a large number of videos dedicated for product reviews in which people try to convince others in favor of or against the use of various products. This raises an interesting research problem as to what it is that makes certain speakers have a substantial impact on others' opinions while other speakers are ignored.

   In this paper, we present a novel analysis of spoken persuasion in online multimedia content. Our work is motivated by prior research findings in psychology indicating that verbal behavior is a promising indicator for persuasive communication (Chaiken and Eagly, 1979; Werner, 1982). Such prior findings allow us to hypothesize that two primary types of verbal features will be predictive of persuasion: lexical features and paraverbal markers of hesitation. Additionally we explore the relationship of the sentiment of the content and perceived persuasion, by hypothesizing that speakers' exhibit different verbal behavior when expressing a positive opinion versus a negative one and taking into account these differences will improve prediction performance. We conduct several experiments in order to validate these hypotheses using a multimedia corpus of 1,000 movie review videos obtained from ExpoTV.com, which is a great source of online reviews. Our experiments followed by a detailed analysis also reveal a set of predictive features which characterize persuasive online presentations.

---

* Both authors contributed equally to this work.

In the following section, we present an overview of related work. Section 3 elaborates on our research hypotheses. In Section 4, we present our multimedia corpus. The details of the experiments with computational descriptors and methodology are described in Section 5. We discuss the results and findings in Section 6, and finally we conclude our paper and present some future directions of research in Section 7.

2   Related Work

Content in the form of written text are omnipresent in our society. Starting from books, magazines and newspapers to the now prevalent emails and blog posts, text-based content are an invaluable component for effective communication. Prior research reports possibly greater persuasiveness in written messages compared to visual or acoustic modalities in certain situations (Chaiken and Eagly, 1979; Werner, 1982). Past research has also revealed that for sophisticated messages, such as those used in a martial setting, written messages are more persuasive (Chaiken and Eagly, 1979).

Although the importance of studying verbal behavior for determining persuasiveness has been underscored in prior work in the field of communication sciences (O'Keefe, 2002) and this line of research gives us useful pointers to the factors that contribute to persuasiveness in text or verbal communication, they present no computational aspect, which is where we put our emphasis in the paper.

In the field of natural language processing, text classification based on bag-of-words has been a long standing approach (Lewis and Gale, 1994; Mitchell, 1997; Dave et al., 2003). In fact, Young et al. (2011) have explored lexical features in the specific context of predicting persuasion, but they focus their attention on studying persuasion in dialogue. Our work draws inspiration from such approaches but explores it in the specific context of predicting persuasiveness in online multimedia content using lexical and paraverbal features.

3   Research Hypotheses

Motivated by prior works and theoretical background, we designed our experiments to validate three hypotheses.

Since multiple prior works point to the usefulness of the text modality in persuasive communication and also to the power of text classification with lexical features in various tasks, we explored the feasibility of capturing the difference in verbal behavior between persuasive and unpersuasive expressions of opinions in online social multimedia content (specifically, movie reviews). The following is the hypothesis that we specifically tested with our experiments:

Hypothesis 1: Verbal behavior, as captured by lexical usage, is indicative of persuasiveness in online social multimedia content, irrespective of whether the opinion expressed is positive or negative.

Paraverbal behaviors indicative of hesitation can constitute important information for predicting persuasiveness. For instance, a speaker's stuttering or breaking his/her speech with filled pauses (such as *um* and *uh*) has influence on how other people perceive his/her persuasiveness. Although previous work (DeVault et al, 2013) suggests paraverbal behavior may be indicative of depression, another work on emotion prediction however, (Devillers et al., 2006) raised questions about its predictive power when compared to using standard cues derived from lexical usage. This leads us to our second hypothesis on paraverbal behaviors in the context of predicting persuasiveness:

Hypothesis 2: Paraverbal behaviors related to hesitation are indicative of persuasiveness in online social multimedia content.

Past research highlights the importance of the knowledge of the affective state of a document towards its perceived persuasiveness (Murphy, 2001). We therefore hypothesize the following:

Hypothesis 3: Knowledge of the sentiment polarity of a movie review improves classification performance of the speaker's perceived persuasiveness.

4   Dataset

ExpoTV.com is a popular website housing videos of product reviews. Each product review has a video of a speaker talking about a particular product as well as the speaker's direct rating of the product on an integral scale from 1 star (for most negative review) to 5 stars (for most positive review). This direct rating is useful for the purpose of our study because this allows us to study perceived persua-

sion under different directions of persuasion (in favor of or against). For instance, the speaker in a 5-star movie review video would most likely try to persuade his/her audience in favor of watching the movie while the speaker in a 1-star movie review video would argue against watching the movie. We therefore collected a total of 1,000 movie review videos that were either highly positive or negative. The dataset consists of the following:

- **Positive Reviews**: 500 movie review videos with 5-star rating (315 males and 185 females).

- **Negative Reviews**: 500 movie review videos with 1 or 2-star rating, consisting of 216 1-star videos (151 males and 65 females) and 284 2-star videos (212 males and 72 females). We included 2-star videos due to the lack of enough 1-star videos on the website.

Each video in the corpus has a frontal view of one person talking about a particular movie, and the average length of the videos is about 94 seconds. The corpus contains 372 unique speakers and 600 unique movie titles and is available to the community for purposes of academic research[1].

4.1 Evaluation of Persuasiveness

Amazon Mechanical Turk (AMT), which is a popular online crowdsourcing platform, was used to obtain subjective evaluation of each speaker's perceived persuasiveness, following a similar annotation scheme as (Mohammadi et al., 2013). For each video in the corpus, we obtained 3 repeated evaluations on the level of persuasiveness of the speaker by asking the workers to give direct rating on each speaker's persuasiveness on a Likert scale from 1 (very unpersuasive) to 7 (very persuasive). A total of 50 native English-speaking workers based in the United States participated in the evaluation process online, and the task was evenly distributed among the 50 workers. To minimize gender influence, the task was distributed such that the workers only evaluated speakers of the same gender. The correlation between the mean score of every movie and the individual ratings was found to be 0.7 on the average (Pearson's Correlation Coefficient).

Once the evaluation was complete, we used the mean persuasiveness score for each video as the ground-truth measure of the speaker's perceived persuasiveness. In this initial effort, we focused on videos that were extremely persuasive or not persuasive at all. Hence, videos with a mean score of equal to or greater than 5.5 were taken as persuasive while those with a mean score of equal to or less than 2.5 were taken as unpersuasive. After this, we ended up with a total of 300 videos, specifically 157 videos of positive reviews (75 persuasive and 82 unpersuasive) and 143 videos of negative reviews (62 persuasive and 81 unpersuasive).

4.2 Transcriptions

Using AMT and 18 participants from the same worker pool for persuasiveness evaluation, we obtained verbatim transcriptions of these filtered 300 videos, including transcriptions for filled pauses and stutters. Each transcription was reviewed and edited by multiple in-house experienced transcribers for accuracy. We do not use automatic speech recognition techniques in order to avoid noisy transcriptions.

5 Experiments

In this section, we give details on the design of our computational descriptors followed by the experimental methodology.

5.1 Computational Descriptors

In our experiments, our main focus was on devising computational descriptors for verbal behaviors in terms of lexical usage and also in terms of paraverbal markers of hesitation that can capture indications of persuasiveness of the speaker.

---

[1] Dataset available online: http://multicomp.ict.usc.edu/

Verbal (Lexical) Descriptors: As in many text classification tasks, we designed our verbal descriptors based on the bag-of-words representation using term frequency of both unigrams and bigrams. Using the 300 filtered videos (see Section 4.1) and without feature selection, the numbers of unigrams reach around 4,500 and bigrams around 24,000. We did not proceed further with higher order n-grams because empirical evidence has shown that trigrams and other higher order n-grams do not always show improvement because they introduce problems related to the sparsity of features (Dave et al., 2003).

Paraverbal Descriptors of Hesitation: From the verbatim transcriptions of our corpus, we observed a set of frequent paraverbal cues that could potentially be associated with the level of persuasiveness. The set of descriptors is inspired from the findings of DeVault et al. (2013), who explored a similar set of generic paraverbal features in an interactive dialogue setting. However, we are interested specifically in the ones that capture signs of hesitation. The following were the descriptors that were used:

- **Pause-Fillers:** The verbal behaviors of reviewers are often characterized with various pause-fillers, such as *um* or *uh*. In order to account for the varying length of each review, we normalized the count of all instances of filled pauses by the number of words spoken in the video.

- **Disfluency Markers:** A prominent marker of disfluency in human speech is stuttering. To capture this disfluency, we counted all instances of stuttering in each video and normalized them by the number of words spoken in the video.

- **Articulation Rate:** Articulation rate is defined as the rate of speaking in which all pauses are excluded from calculation (Dankovicova, 1999). This descriptor was computed by taking the ratio of the number of spoken words in each video to the actual time spent speaking.

- **Mean Span of Silence:** Human speech is often interspersed with pauses. We therefore computed this descriptor, by measuring the total duration of silence during speech, normalized by the total length of the video.

5.2 Methodology

We processed all the videos in our dataset and automatically extracted the indicated lexical and paraverbal features. The extracted features were then used for several classification experiments under three different settings to test our hypotheses: only positive reviews, only negative reviews (called the *sentiment-dependent classifiers*) and a combined set of positive and negative reviews (called the *sentiment-independent classifiers*). For each such setting, we divided the set of samples (transcription of movie reviews) into 5 balanced folds that were both speaker-independent and movie-independent. In other words, in all our experiments, no 2 folds contained samples from the same speaker or movie title. This was done to remove any form of bias in the classifier based on either the speaker or the movie.

We then performed classification experiments using 5-fold cross-validation using the lexical features (unigrams and bigrams) on this combined set of reviews (positive and negative reviews together), each time leaving 1 fold for hold-out testing. Here, we note that for constructing the dictionary, only data from the training set was used. On average across 5-fold cross-validation, the number of unigrams was around 4,560 and bigrams around 23,701 for the combined set of movie reviews.

However, since such a feature design typically suffers from problems arising out of the sparsity of the entries of the dictionary in the dataset, we employed a feature selection step. For feature selection and analysis, we used Information Gain (IG), which is a measure of the number of bits of information obtained for category prediction by knowing the presence of a term in a document. Prior evaluation of feature-selection methods for text classification has revealed the superiority of IG as a metric over other ones such as Mutual Information, Term Strength or a simple Document Frequency thresholding for document classification tasks (Yang and Pedersen, 1997). This serves as an inspiring basis for using IG as a metric for feature selection.

The gain score $G(t)$ obtained from IG is a non-zero positive value for features that are strongly indicative of the extent of persuasiveness of the document, while ones that are not so informative have a value of 0. We therefore select only those lexical features (unigrams and bigrams) which have an IG >

0 based on the distribution obtained from the training set. This allows us to trim the dictionary significantly and use only meaningful features for classification.

Feature Group	Sentiment Dependent Classifier			Sentiment Independent Classifier
	Mean	**Positive Reviews**	**Negative Reviews**	
Lexical Features (Unigrams and Bigrams)	**83.92%**	81.74%	86.09%	76.73%
• Unigrams Only	**77.70%**	74.78%	80.62%	73.77%
• Bigrams Only	**84.05%**	81.64%	86.46%	75.81%
Para-Linguistic Features	**64.23%**	65.22%	63.23%	63.04%
Early Fusion	**84.54%**	82.61%	86.46%	78.56%
Majority Baseline	52.14%	50.43%	53.85%	51.09%

Table 1: Accuracies for our experiments using a Naïve Bayes classifier. The scores in bold indicate the dominance of the sentiment-dependent classifier under all circumstances.

This was then followed up by a 5-fold cross-validation using only the paraverbal features (no feature selection was used here since they were too few in number). The accuracy of classification based on paraverbal features was then compared with that obtained by classification using only the lexical descriptors and by a majority baseline classifier.

Furthermore, we also tried an early-fusion approach, where we simply use both lexical and paraverbal features together. Such an approach to fusion seemed more promising here than a decision-level fusion approach because of the few categories of features used (just lexical and paralinguistic, as motivated by the findings of (Gunes and Piccardi, 2005)).

5.3 Classification Model

For performing classification experiments we used the Naïve Bayes classifier. However a well-known issue with using the Naïve Bayes classifier is its incapability of handling new features, which is taken care of by performing a conditional uniform smoothing (Puurula, 2012).

6 Results and Discussion

Table 1 shows the results for our classification experiments, which confirm the predictive power of lexical features.

Hypothesis 1: The lexical features (unigrams and bigrams) are predictive of persuasiveness. This is manifested by the fact that they perform significantly better than a majority baseline, which is only 51.04% accurate on the combined set of positive and negative reviews, while the lexical features achieved an accuracy of around 77% (Figure 1). Considering the positive and the negative reviews individually, we note that the lexical features were accurate for nearly 82% of the test samples for the positive reviews and for 86% of the test samples for the negative reviews, again outperforming a simple majority baseline classifier (Table 1).

An analysis of the features (Table 2) reveals that certain lexical features contribute to the predictability of the persuasiveness of a speaker. The presence of unigrams such as *character* or *make* or bigrams such as *to make* or *this movie* for instance, contributes to the predictability of persuasiveness of the speaker, even though they are not emotionally salient terms. The high IG scores of such features irrespective of the setting we conduct our experiments in (positive reviews only, negative reviews only or a combined set of positive and negative reviews), highlights their importance. Moreover, a (+) sign for most of these unigrams or bigrams show that their presence contributes favorably to the speaker being perceived as persuasive. On the other hand a (-) sign for an informative bigram such as *it says* is indicative of lack of speaker's persuasiveness. This can be explained by the context of the usage of such features. For instance, the bigram *it says* in *it says that the movie duration is…* is a bigram that is uttered by the reviewers when they have to refer to the DVD cover of the movie to give
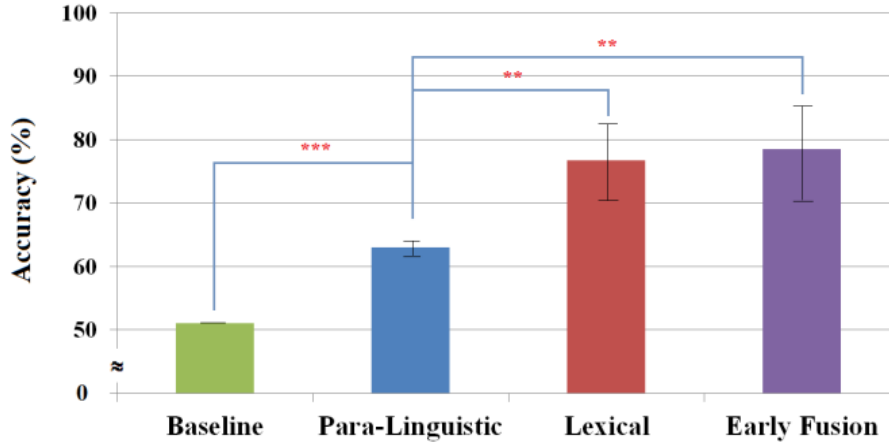
Figure 1: Bar graph visualization of the classification accuracies using different types of features on the combined set of reviews (i.e. sentiment-independent classifier). ** indicates 2-samples t-test results with p < 0.01 and *** indicates p < 0.001. The error bars show 1 SD.

Feature	Positive Reviews		Negative Reviews		Both Combined	
	Word	IG Score	Word	IG Score	Word	IG Score
Unigrams	The (+)	0.1183	Even (+)	0.11	Make (+)	0.1117
	Make (+)	0.0816	Make (-)	0.1082	Just (+)	0.0728
	Everything (+)	0.0806	Movie (+)	0.0969	Very (+)	0.0669
	Just (+)	0.0806	Real (+)	0.0873	Character (+)	0.0573
	Dollars (+)	0.0722	Not (+)	0.0867	Becomes (+)	0.0558
	Character (+)	0.0685	Big (+)	0.0858	Even (+)	0.0524
	Can (+)	0.0685	One (+)	0.0817	One (+)	0.051
	Product (+)	0.0685	Avoid (+)	0.079	Yourself (+)	0.05
	Famous (+)	0.0609	Feel (+)	0.079	You (+)	0.04571
	Enjoy (+)	0.0566	Character (+)	0.0773	Lot (+)	0.0456
Bigrams	There are (+)	0.1183	This movie (+)	0.1083	To make (+)	0.0905
	This movie (+)	0.0816	Do not (+)	0.1032	A lot (+)	0.0617
	I can't (+)	0.0806	I think (+)	0.1032	This movie (+)	0.0578
	To make (+)	0.0806	To make (+)	0.0989	Lot of (+)	0.0443
	Good movie (+)	0.0722	Not even (+)	0.091	It says (-)	0.0417
	Buy it (+)	0.0685	Don't even (+)	0.091	You will (-)	0.0417
	Really a (+)	0.0685	The story (+)	0.079	Twenty dollars (+)	0.0368
	Definitely one (+)	0.0685	The film (+)	0.0672	The character (+)	0.0386
	Best movies (+)	0.0609	At all (+)	0.0672	So many (+)	0.033
	It's awesome (+)	0.0566	It's so (+)	0.0672	See it (+)	0.033

Table 2: Important unigrams and bigrams when they are used individually as lexical features. (+) indicates that it increases persuasiveness while (-) indicates it contributes to the lack of it.

some more detailed information about it. This is identified as a sign of an unpersuasive reviewer. Such results confirm that the verbal behaviors, as captured by lexical usage, are extremely predictive of persuasiveness irrespective of whether the opinion expressed is positive or negative, which validates Hypothesis 1.

Hypothesis 2: Moreover, our experiments show that although the designed paraverbal features that are markers of hesitation can classify only about 63% of the speakers correctly (see Table 1), however they are statistically significant features, in terms of their p-values (Figure 2). While classification performance is lower than that obtained with purely lexical features, it is still far above a majority baseline, and thus confirms our second hypothesis.
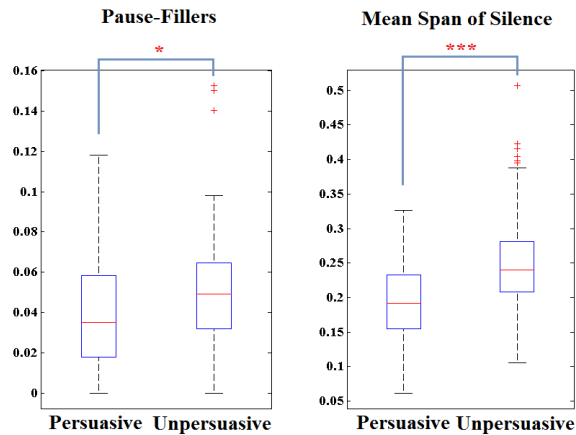
Figure 2: Boxplots for the paralinguistic hesitation markers for a classifier trained on the paralinguistic features only. * and *** indicate p <= 0.05 and 0.001, respectively.

Additionally, it is interesting to note from Table 1 that, although a feature-level fusion of the lexical features and paraverbal features gives us an improvement in classification performance, the difference between the results obtained with fusion and those with lexical features alone was minor and was not statistically significant.

Hypothesis 3: We also observe that a sentiment-dependent classifier trained individually on positive reviews or on negative reviews outperforms one that is trained on a combined set of reviews. This is supported by our empirical results in Table 1 which show that when classification is performed with any of the lexical features, the accuracies are significantly higher for the classifier trained only on the positive or only on the negative reviews (sentiment-dependent classifiers) than for the classifier trained on the combined set of reviews (sentiment-independent classifiers). For instance, when unigrams and bigrams were both used as our lexical features, we observed that for a sentiment-dependent classifier the classification accuracy jumps to over 84% on average. This is significantly better than the scenario where the classifier is not aware of the sentiment of the review. Figure 3 demonstrates this phenomenon.

We resort to feature analysis for an explanation of such an observation (Table 2). The analysis reveals that certain sentiment-based lexical features, i.e. emotionally salient terms, assume an important role in magnifying the discriminative power of language use in persuasiveness prediction, when prior knowledge about the speaker's opinion is known. For instance, in the case of a classifier trained only on the positive reviews, unigrams such as *enjoy* and *famous* and bigrams such as *good movie* or *it's awesome* become significant. In the context of persuading against watching the movie prominent sentiment-based unigrams are *not* and *avoid* while bigrams are *do not*, *don't even* and *at all*. This provides empirical support for our third hypothesis.

7   Conclusion and Future Work

This work presents several interesting findings about perceived persuasiveness prediction in online social multimedia content by analyzing the verbal behavior of the speaker, modeled using lexical features and paraverbal indicators of hesitation. We conducted experiments and showed that verbal behavior as captured by lexical descriptors is a strong indicator of persuasiveness, irrespective of whether we persuade in favor of or against something. Much of this is due to the presence of certain unigrams and bigrams that are either indicative of strong persuasiveness or of lack of persuasiveness. Our experiments further reveal the superiority of classifying with lexical features as compared to with paraverbal features alone. Moreover we empirically validate the hypothesis that a sentiment-aware classifier outperforms a sentiment-independent one. As future work, we intend to explore more paraverbal features for persuasiveness prediction and also try more sophisticated prediction models which explicitly model the temporal dynamic.
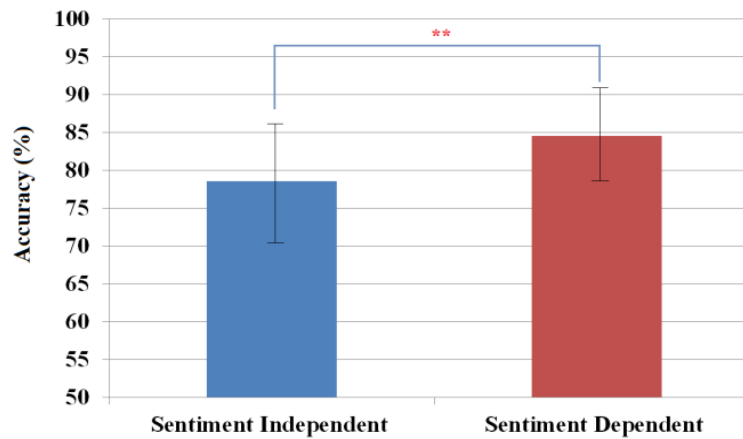
Figure 3: Bar graph visualization of the classification accuracies of lexical features using a sentiment-dependent classifier (mean) and a sentiment independent one. ** indicates 2-sample t-test results with $p < 0.01$ and the error bars show 1 SD.

References

Shelly Chaiken and Alice H. Eagly. 1979. Communication modality as a determinant of message persuasiveness and message comprehensibility. *Journal of Personality and Social Psychology,* 37:1387-1397.

Jana Dankovicova. 1999. Articulation rate variation within the intonation phrase in Czech and English. *14th Int. Congress of Phonetic Sciences,* San Francisco, Vol. 1, pp. 269-272.

Kushal Dave, Steve Lawrence, and David M. Pennck. Mining the Peanut Gallery: Opinion Extraction and semantic Classification of Product Reviews, 2003. *2003 Association for Computational Linguistics (ACL '03).*

David DeVault, Kallirroi Georgila, Ron Artstein, Fabrizio Morbini, David Traum, Stefan, Scherer, Albert (Skip) Rizzo, and Louis-Philippe Morency. 2013. Verbal indicators of psychological distress in interactive dialogue with a virtual human. *SIGDIAL 2013 Conf, 2013 Association for Computational Linguistics (ACL '13).*

Laurence Devillers and Laurence Vidrascu. 2006. Real-life emotions detection with lexical and paralinguistic cues on Human-Human call center dialogs. *Interspeech 2006.*

Hatice Gunes, and Massimo Piccardi. 2005. Affect Recognition from face and body: Early fusion vs. Late fusion. *IEEE Int'l Conf. on Systems, Man and Cybernnetic.*

Daniel J. O'Keefe. 2002. *Persuasion: Theory and research.* (2nd Edition). Sage Publications, Thousand Oaks, CA.

David D. Lewis and William A. Gale. 1994. A Sequential Algorithm for Training Text Classifiers. *Special Interest Group in Information Retieval (SIGIR'94).*

Gerald R. Miller (1980). *On being persuaded: Some basic distinctions*. In M. Roloff, & G. R. Miller (Eds.), Persuasion: New directions in theory and research, 11–28. Beverly Hills, CA: Sage.

Tom M. Mitchell. 1997. *Machine Learning*. McGraw-Hill.

Gelareh Mohammadi, Sunghyun Park, Kenji Sagae, Alessandro Vinciarelli, and Lois-Phillippe Morency. 2013. Who is persuasive? The role of perceived personality and Communication modality in social multimedia. *Int'l Conf. on Multimodal Interfaces (ICMI '13).*

P. Karen Murphy. 2001. What makes a text persuasive? Comparing students' and experts' conceptions of persuasiveness. *Int'l Journal of Education Research*, 35 (2001) 675-698.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. *Conf. on Empirical Methods in Natural Language Processing. (EMNLP '02).*

Antti Puurula. 2012. *Combining Modifications to Multinomial Naive Bayes for Text Classification.* Springer, LNCS.

Kathleen Kelley Reardon. 1991. *Persuasion in practice.* Sage Publication, Inc.

Carol Werner. 1982. Intrusiveness and persuasive impact of three communication media. *Journal of Applied Social Psychology,* 89:155-181.

Yiming Yang and Jan O. Pedersen. 1997. A comparative study on feature selection in text categorization. *Int'l Conf. on Machine Learning (ICML '97).*

Joel Young, Craig Martell, Pranav Anand, Pedro Ortiz and Henry T. Gilbert IV. 2011. A Microtext Corpus for Persuasion Detection in Dialog. *Analyzing Microtext: AAAI Workshop (AAAI-Workshop '11).*

Phillip G. Zimbardo and Michael R. Leippe. 1991. *The psychology of attitude change and social influence.* McGrew-Hill New York.