

Lie Detection using NLP Techniques

Swapnil Ghuge (10305907)

Jyotesh Choudhari (113059001)

This seminar is mainly based on the short paper “The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language” by Rada Mihalcea and Carlo Strapparava, published in ACL-2009

Outline

- Introduction
- Approach
- Corpus
- Experiments and Results
- Identifying Salient Features of Deceptive Text
- Conclusion

Introduction

- Discrimination between truth and falsehood in psychology, philosophy and sociology
- From psychological aspect, human behavior, gesture, repetition of words, stammering, eye contact are important cues to detect lies
- Can we apply computational approaches to recognize deceptive language in written text?

Introduction

- Objective is to answer following questions:
 - Are truthful and lying texts separable?
 - Does this property hold for different datasets?
 - Can computational approaches be used to separate truthful and lying texts?
 - What are the distinct features of deceptive text?

Approach

- Statistical approach
- Classes – Truth and Falsehood
- Classifier – Naïve Bayes and SVM
- Features
- Corpus

Corpus

- 3 Datasets
- True and lying opinions on
 - Abortion
 - Death Penalty
 - Best Friends
- 100 true and 100 false statements in every dataset.

Experiments and Results

- Used two classifiers
 - Naïve Bayes
 - SVM
- Minimal preprocessing – tokenization and stemming

Topic	Naïve Bayes	SVM
ABORTION	70%	67.5%
DEATH PENALTY	67.4%	65.9%
BEST FRIEND	75.0%	77.0%
AVERAGE	70.8%	70.1%

Table 1: Ten-fold cross validation classification using Naïve Bayes(NB) or Support Vector Machine(SVM)

Experiments and Results

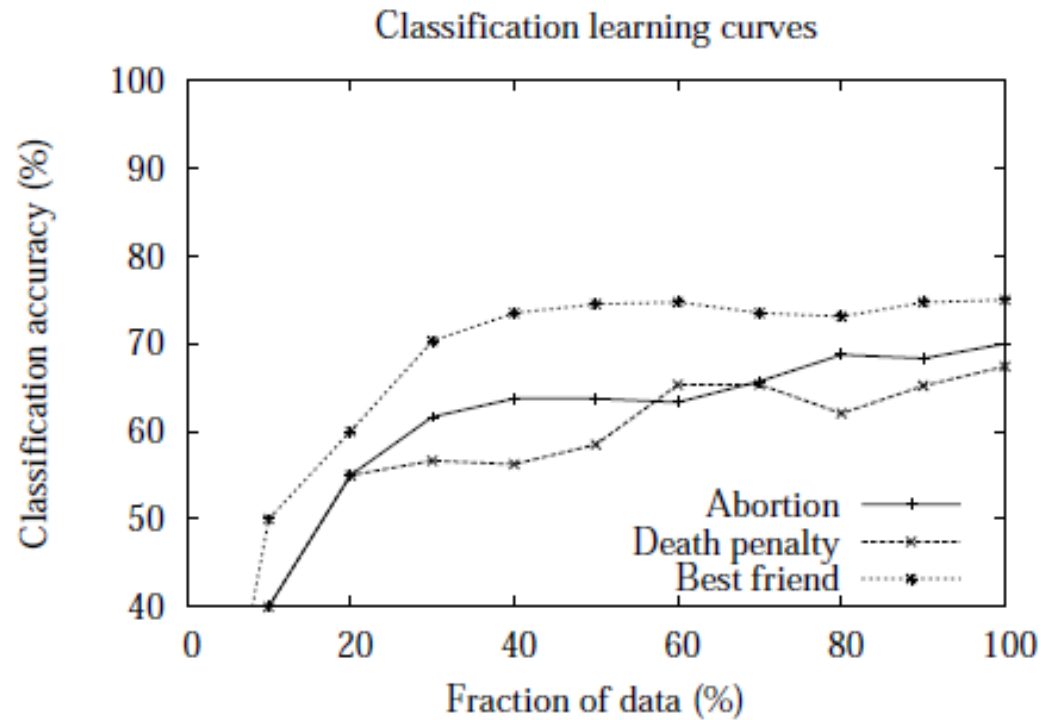


Figure 1: Classification Learning Curves

Experiments and Results

Training	Test	Naïve Bayes	SVM
DEATH PENALTY + BEST FRIENDS	ABORTION	62.0%	61.0%
ABORTION + BEST FRIENDS	DEATH PENALTY	58.7%	58.7%
DEATH PENALTY + ABORTION	BEST FRIENDS	58.7%	53.6%
AVERAGE		59.8%	57.8%

Table 2: Cross-topic classification results

Have we achieved it yet?

- Objective is to answer following questions:
 - Are truthful and lying text separable? - YES
 - Does this property hold for different datasets? - YES
 - Can computational approaches be used to separate truthful and lying texts? - YES
 - What are the distinct features of deceptive text?

Identifying Salient Features

- To understand the characteristics of the deceptive text, calculate a score for a given class of words

$$C = \{W_1, W_2, \dots, W_N\}$$

- Class coverage in the deceptive corpus D :

$$Coverage_D(C) = \frac{\sum_{w_i \in C} Frequency_D(w_i)}{size_D}$$

- Class coverage in the truthful corpus T :

$$Coverage_T(C) = \frac{\sum_{w_i \in C} Frequency_T(w_i)}{size_T}$$

Identifying Salient Features

- Dominance of word class C in deceptive corpus D :

$$\text{Dominance}_D(C) = \frac{\text{Coverage}_D(C)}{\text{Coverage}_T(C)}$$

- Dominance higher than 1 indicates word class C is dominant in deceptive texts.
- Dominance less than 1 states that word class C is more likely to appear in truthful texts.
- Dominance score close to 1 shows similar distribution of word class C in both the deceptive and truthful corpus.

Identifying Salient Features

- Word classes obtained from Linguistic Inquiry and Word Count (LIWC) by (Pennebaker and Francis, 1999).

Text	Class	Score	Sample Words
Deceptive	METAPH	1.71	God, mercy, sin, hell
	OTHER	1.47	She, him, themselves
	CERTAIN	1.24	Always, truly, completely
Truthful	OPTIM	0.57	Best, determined, accept
	I	0.59	I, myself, mine
	INSIGHT	0.65	Believe, understand, admit

Table 3: Dominant Word Classes along with sample words

Conclusion

- Explored computational approach to recognize deceptive language in written text.
- Truthful and false texts are separable and this property holds for different datasets.
- Salient features show interesting pattern of uses of words in deceptive texts including detachment from self and emphasizing certainty.

References

- Rada Mihalcea, Carlo Strapparava. The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. In *Proceedings of ACL-IJCNLP Conference Short Papers 2009*.
- J. Pennebaker, M. Francis. 1999. Linguistic Inquiry and Word Count: LIWC. Erlbaum Publishers.

Ask no questions and you will hear no
lies.

-James Joyce, *Ulysses*