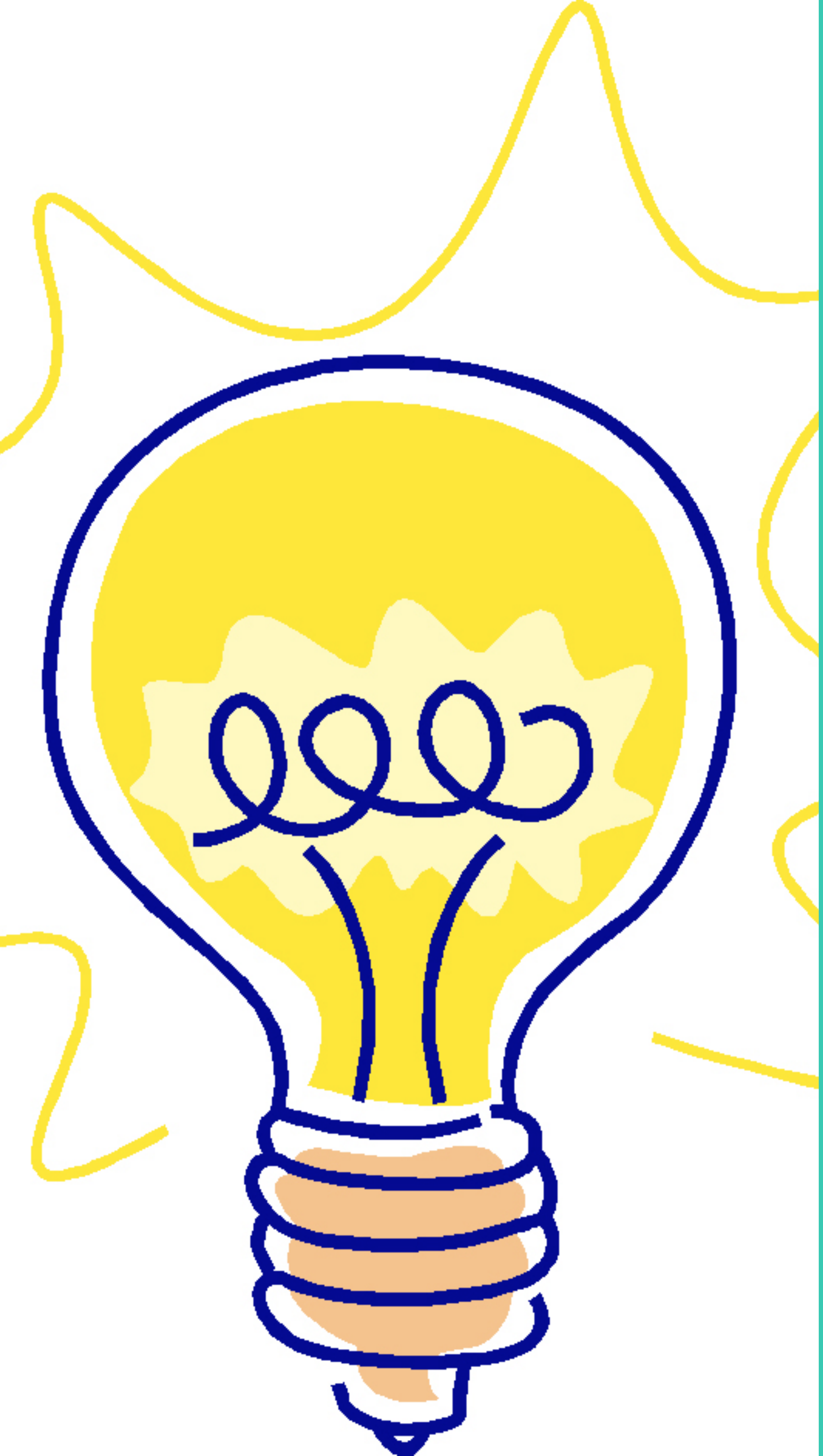
“DETERMINING THE SENTIMENT OF OPINIONS”

SOO-MIN KIM AND EDUARD HOVY
UNIVERSITY OF SOUTHERN CALIFORNIA

Aditya Bindra

Paul Cherian

Benjamin Haines



INTRODUCTION

A. Problem Statement
B. Definitions
C. Outline
D. Algorithm

PROBLEM STATEMENT

- Given a topic, and set of text related to that topic, find the opinions that people hold about the topic.
- Various models to classifying and combine sentiment at word and sentence level.

DEFINITIONS

- Define an opinion as a tuple `[Topic, Holder, Claim, Sentiment]`.
- Sentiment is positive, negative, or neutral regard toward the `Claim` about the `Topic` expressed by the `Holder`
  - I like ice-cream. (*explicit*) 😀
  - He thinks attacking Iraq would put US in a difficult position. (*implicit*) ☹️
  - I haven't made any decision on the matter 😐

OUTLINE

- Approached the problem in stages, first words, and then sentences.
- A unit sentiment carrier is a word.
- Classify each adjective, verb and noun by sentiment.
    - Ex: *California Supreme Court **agreed** that the state's new term-limit law was **constitutional***.
    - Ex: *California Supreme Court **disagreed** that the state's new term-limit law was **constitutional***.
- A sentence might express opinions about different people(Holders).
    - Determine for each holder, a relevant region within sentence.
- Various models to combine sentiments.

CLASSIFICATIONS

A. Holder Identification
B. Regions of Opinion
C. Word Sentiment Classifiers
D. Sentence Sentiment Classifiers

HOLDER IDENTIFICATION

- Used IdentiFinder named entity tagger.
- Only consider PERSON and ORGANIZATION.
- Choose `Holder` closest to the `Topic`.
    - Could have been improved with syntactic parsing to determine relations.
- `Topic` finding is done by direct match.

REGIONS OF OPINION

assumption: sentiments most reliably found close to the Holder

1. **Window1:** full sentence
2. **Window2:** words between Holder and Topic
3. **Window3:** *window2* ± 2 words
4. **Window4:** *window2* to the end of the sentence

WORD SENTIMENT CLASSIFICATION MODELS

Begin with hand selected seed sets for positive and negative words and repeatedly expand by adding WordNet synonyms and antonyms.

Problem: Words occur in both lists.

Solution: Create a polarity strength measure. This also allows classification of unknown words.

WORD SENTIMENT CLASSIFICATION MODELS

To compute $P(c|w) = P(c|\text{syn}_1, \ldots, \text{syn}_n)$ two models were developed

$$\textbf{Word Classifier1: } \underset{c}{\text{argmax}} P(c|w) =$$

$$\underset{c}{\text{argmax}} P(c) \frac{\sum_{i=1}^{n} \text{count}(\text{syn}_i, c)}{\text{count}(c)}$$

$$\textbf{Word Classifier2: } \underset{c}{\text{argmax}} P(c|w) =$$

$$\underset{c}{\text{argmax}} P(c) \prod_{k=1}^{m} P(f_k|c)^{\text{count}(f_k, \text{synset}(w))}$$

Example Outputs

abysmal:
NEGATIVE [+ : 0.3811][- : 0.6188]

adequate:
POSITIVE [+ : 0.9999][- : 0.0484e-11]

afraid:
NEGATIVE [+ : 0.0212e-04][- : 0.9999]

SENTENCE SENTIMENT CLASSIFICATION MODELS

Model 0: $\prod(\text{signs in region})$	Product of sentiment polarities in region.	"Negatives cancel out." Include "not", "never."
Model 1: $P(c\|s) = \frac{1}{n(c)} \sum_{i=1}^{n} p(c\|w_i)$, if $\underset{j}{\text{argmax}} p(c_j\|w_i) = c$	Harmonic mean of sentiment strengths in region.	Considers number and strength of words.
Model 2: $P(c\|s) = 10^{n(c)-1} \prod_{i=1}^{n} p(c\|w_i)$, if $\underset{j}{\text{argmax}} p(c_j\|w_i) = c$	Geometric mean of sentiment strengths in region.	

SENTENCE SENTIMENT CLASSIFICATION MODELS

example output

Public officials throughout California have condemned a U.S. Senate vote Thursday to exclude illegal aliens from the 1990 census, saying the action will shortchange California in Congress and possibly deprive the state of millions of dollars of federal aid for medical emergency services and other programs for poor people.

`TOPIC:` illegal alien

`HOLDER:` U.S. Senate

`OPINION REGION:` vote/NN Thursday/NNP to/TO exclude/VB illegal/JJ aliens/NNS from/IN the/DT 1990/CD census,/NN

`SENTIMENT_POLARITY:` negative

EXPERIMENTS

A. Word Sentiment Classifier Models
B. Sentence Sentiment Classifier Models

WORD SENTIMENT CLASSIFIER EXPERIMENT

human classification

- TOEFL English word list for foreign students
    - Intersected with adjective list of 19,748 English adjectives
    - Intersected with verb list of 8,011 English verbs
- Randomly selected 462 adjectives and 502 verbs for human classification
    - Humans classify words as positive, negative, or neutral

	Adjectives	Verbs
	Human1 vs Human2	Human1 vs Human3
Strict	76.19%	62.35%
Lenient	88.96%	85.06%

WORD SENTIMENT CLASSIFIER EXPERIMENT

human-machine classification results

- Baseline randomly assigns sentiment category (10 iterations)

$$\textbf{Word Classifier2:} \ \underset{c}{\operatorname{argmax}} P(c|w) = \underset{c}{\operatorname{argmax}} P(c) \prod_{k=1}^{m} P(f_k|c)^{\operatorname{count}(f_k, \operatorname{synset}(w))}$$

	Adjectives (test: 231)			Verbs (test: 251)		
	Lenient Agreement		Recall	Lenient Agreement		Recall
	Human1 vs Model	Human2 vs Model		Human1 vs Model	Human3 vs Model	
Random Selection	59.35%	57.81%	100%	59.02%	56.59%	100%
Basic Method	68.37%	68.60%	93.07%	75.84%	72.72%	83.27%

- System has lower agreement than human, higher than random

WORD SENTIMENT CLASSIFIER EXPERIMENT

human-machine classification results (cont.)

- Previous examination used few seed words (44 verbs, 34 adjectives)
- Added half of collected annotated data (251 verbs, 231 adjectives) to training set and kept other half for testing

	Adjectives (train: 231, test: 231)			Verbs (train: 251, test: 251)		
	Lenient Agreement		Recall	Lenient Agreement		Recall
	Human1 vs Model	Human2 vs Model		Human1 vs Model	Human3 vs Model	
Basic Method	75.66%	77.88%	97.84%	81.20%	79.06%	93.23%

- Agreement and recall for both adjectives and verbs improves

SENTENCE SENTIMENT CLASSIFIER EXPERIMENT

human classification

- 100 sentences from DUC 2001 corpus
- 2 humans annotated the sentences as positive, negative, or neutral
- Kappa coefficient = 0.91, which is reliable
    - Measures inter-rater agreement that takes agreement by chance into account
    - $\kappa = \frac{p_o - p_e}{1 - p_e}$ where p_o is the relative observed agreement between raters and p_e is the probability of agreement by chance

SENTENCE SENTIMENT CLASSIFIER EXPERIMENT

test on human annotated data

- experimented on 3 models of sentence sentiment classifiers:

Model 0: $\prod(\text{signs in region})$

$$\textbf{Model 1: } P(c|s) = \frac{1}{n(c)} \sum_{i=1}^{n} p(c|w_i), \quad \text{if } \underset{j}{\operatorname{argmax}} p(c_j|w_i) = c$$

$$\textbf{Model 2: } P(c|s) = 10^{n(c)-1} \prod_{i=1}^{n} p(c|w_i), \quad \text{if } \underset{j}{\operatorname{argmax}} p(c_j|w_i) = c$$

- using 4 window definitions:

Window1: full sentence
Window2: words between Holder and Topic
Window3: *window2* ± 2 words
Window4: *window2* to the end of the sentence

Model 0: 8 combinations (only considers polarities, word classifiers yield same results)

Models 1,2: 16 combinations

- and 4 variations of word classifiers (2 normalized):

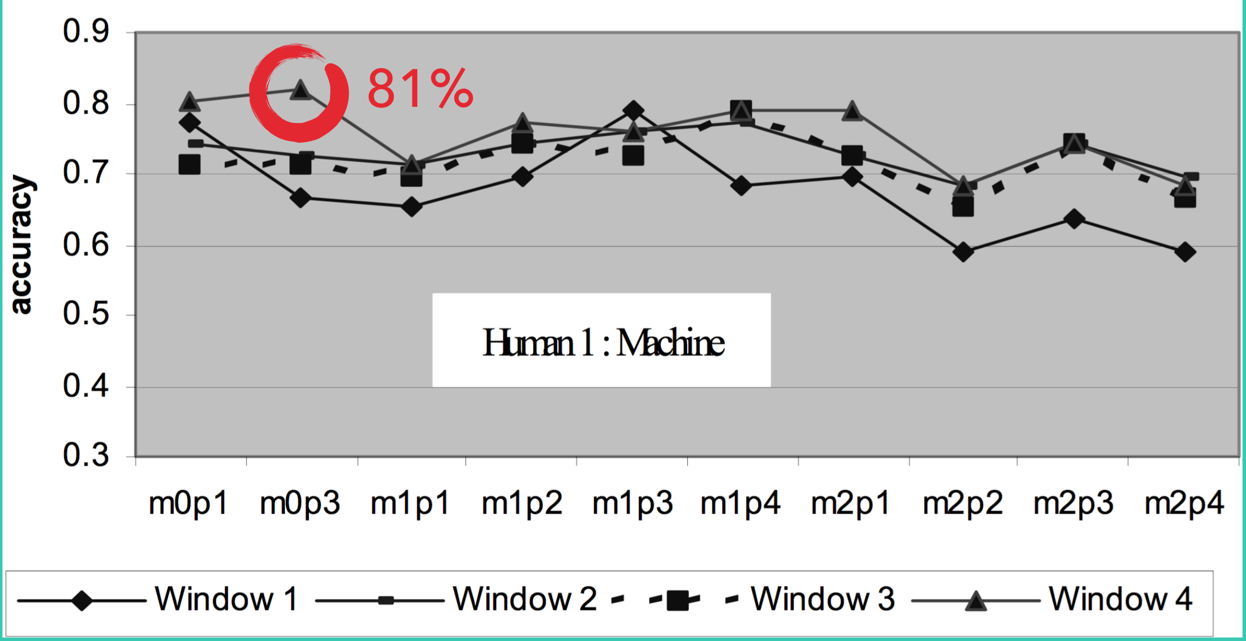
$$\textbf{Word Classifier1: } \underset{c}{\operatorname{argmax}} P(c|w) = \underset{c}{\operatorname{argmax}} P(c) \frac{\sum_{i=1}^{n} \text{count}(\text{syn}_i, c)}{\text{count}(c)}$$

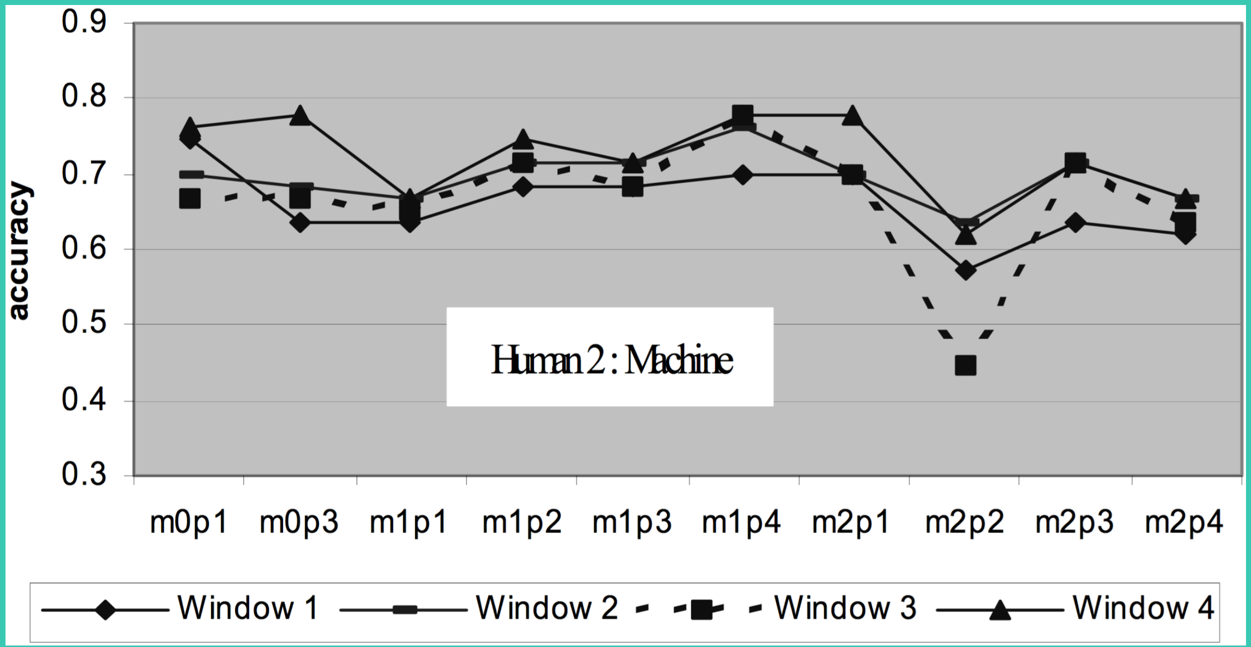
$$\textbf{Word Classifier2: } \underset{c}{\operatorname{argmax}} P(c|w) = \underset{c}{\operatorname{argmax}} P(c) \prod_{k=1}^{m} P(f_k|c)^{\text{count}(f_k, \text{synset}(w))}$$

SENTENCE SENTIMENT CLASSIFIER EXPERIMENT

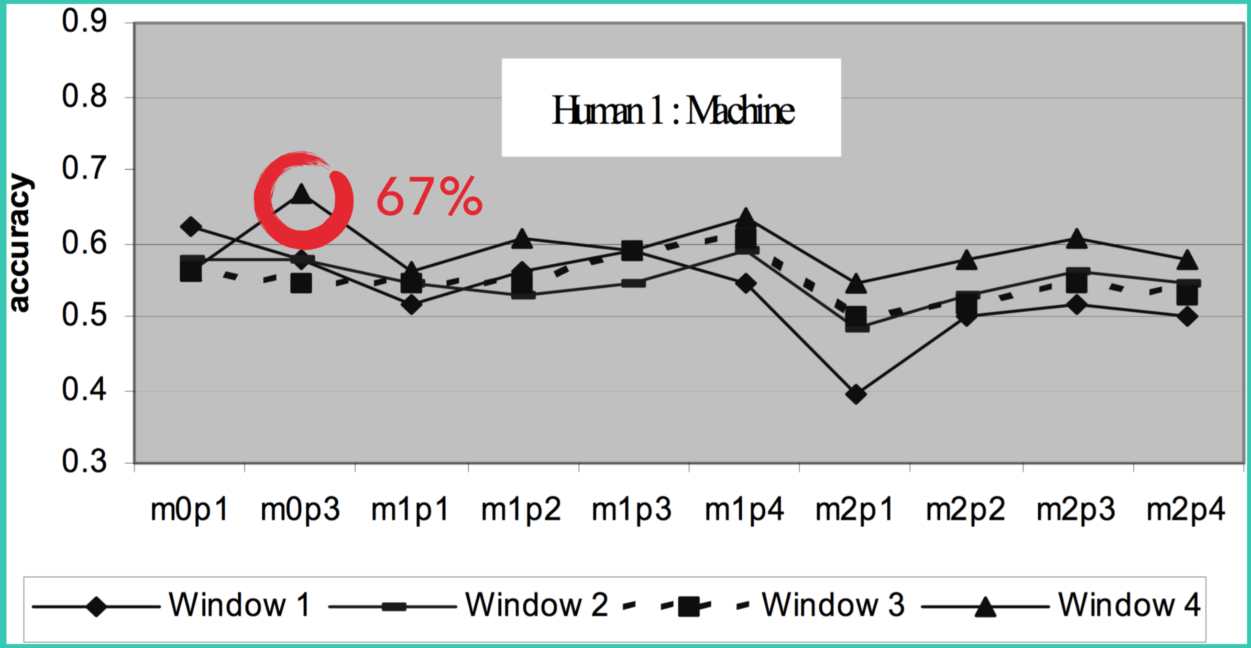
test on human annotated data (cont.)

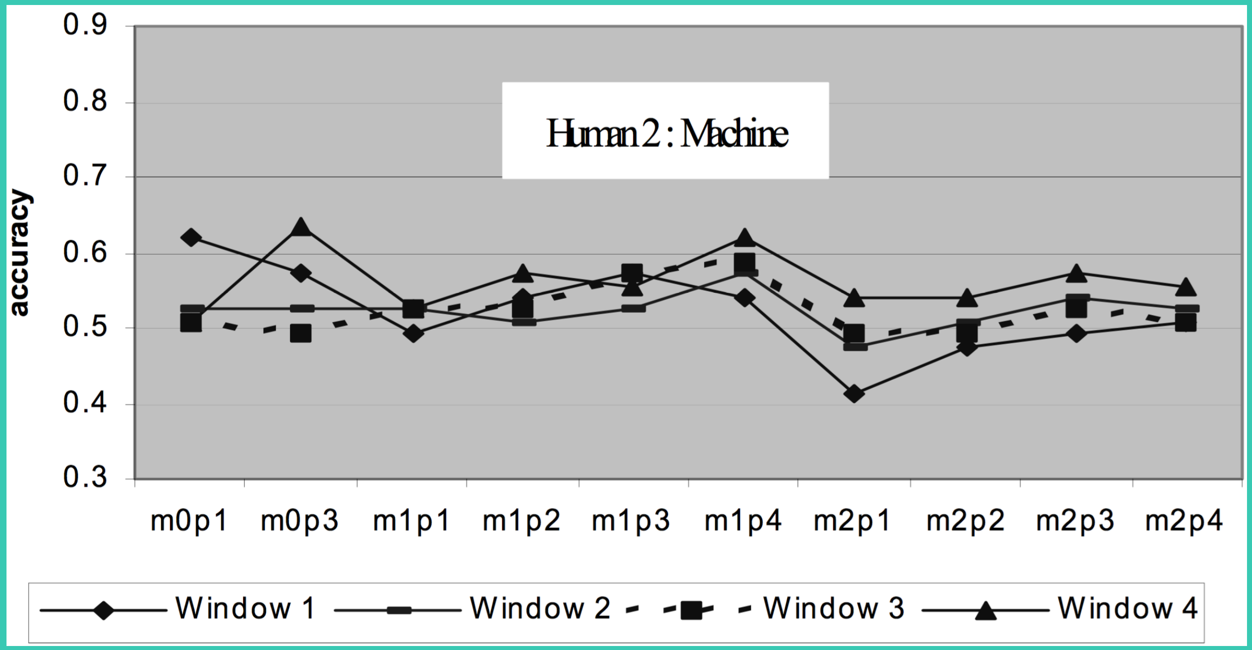
Manually Annotated Holder





Automatic Holder Detection





*m** = sentence classifier model; *p1/p2* and *p3/p4* word classifier model with/without normalization, respectively

RESULTS DISCUSSION

which combination of models is best?

- **Model 0:** $\prod(\text{signs in region})$ provides the best overall performance.
  - Presence of negative words is more important than the sentiment strength of words.

which is better, a sentence or region?

- With manually identified topic and holder, *window4* (Holder to sentence end) is the best performer

manual vs automatic holder identification

Average Difference between Manual and Automatic Holder Detection

	positive	negative	total
Human1	5.394	1.667	7.060
Human2	4.984	1.714	6.698

~7 sentences (11%) were misclassified

DRAWBACKS

word sentiment classification acknowledged drawbacks

- Some words have both strong negative and positive sentiment. It is difficult to pick one sentiment category without considering context.
- Unigram model is insufficient as common words without much sentiment can combine to produce reliable sentiment.
  - Ex: *'Term limits really **hit** at democracy,' says Prof. Fenno*
  - Even more difficult when such words appear outside of the sentiment region.

DRAWBACKS

sentence sentiment classification acknowledged drawbacks

- A Holder may express more than one opinion. This system only detects the closest one.
- System cannot differentiate sentiments from facts.
  - Ex: *"She thinks term limits will give women more opportunities in politics"* = positive opinion about term limits
    - The absence of adjective, verb, and noun sentiment-words prevents a classification.
- System sometimes identifies the incorrect Holder when several are present. A parser would help in this respect.

DRAWBACKS

general unacknowledged drawbacks

- Methodology for selecting initial seed lists were not defined.
- The 19,748 adjectives and 8,011 verbs used as the adjective and word lists, respectively, for the word classifiers were undefined.
- Word sentiment classification experiment never examined

$$\textbf{Word Classifier1:}\ \underset{c}{\mathrm{argmax}} P(c|w) = \underset{c}{\mathrm{argmax}} P(c) \frac{\sum_{i=1}^{n} \mathrm{count}(\mathrm{syn}_i, c)}{\mathrm{count}(c)}$$

- Normalization technique used on word sentiment classifiers is never defined
- Precision and F-measure for classifier analysis needed

CONCLUSION

future plans

- Extend work to more difficult cases
  - sentences with weak-opinion-bearing words
  - sentences with multiple opinions about a topic
- Use a parser to more accurately identify Holders
- Explore other learning techniques (decision lists, SVMs)

QUESTIONS?