



Informativeness and Objectivity of Texts on the Web



Alisa Zhila, Alexander Gelbukh

Natural Language Processing Laboratory, Center for Computing Research, Instituto Politécnico Nacional, Mexico

Characteristics of Texts

Humans distinguish various characteristics of texts:

- Understandability • Objectivity
- Credibility • Informativeness

They can be measured objectively via automatic methods.

However, they are related to evaluation as *good* or *bad*:

Understandable? = **Good** Not understandable? = **Bad**

Problem

- **Are human judgments of these characteristics biased?**
- **Do their automatic estimates coincide with human judgments?**

Automatic Measures

I. Informativeness

[Lex et al. 2012, Horn et al. 2013] showed that informativeness can be assessed via factual density.

Fact: (Abraham Lincoln; was the President of; the United States)

Features used: POS sequences NP – VP – NP

Tool: *ReVerb* [Fader et al. 2011]

- a fact extraction system for English
- Recall = 0.60, Precision = 0.76 (at a certain confidence level)

Document's score = $\frac{fc(d)}{size(d)}$

where $fc(d)$: fact count in the document d ,
size(d): size of the document in characters

II. Objectivity

[Riloff and Wiebe 2003, Wiebe and Riloff 2005] introduced a subjectivity classifier based on a list of > 1000 nouns.

Features used: counts of subjective words

Tool: *OpinionFinder 2.0* [Riloff and Wiebe 2003]

- Each sentence classified as subjective or objective
- Recall = 0.77, Precision = 0.81

Document's score = $\frac{\#obj\ sentences}{\#subj\ sentences}$

Human Judgment Dataset

Web texts: • 50 news articles from GigaWord corpus
• 50 hotel reviews from Trip Advisor corpus

Questions on subjective characteristics of texts

Is the text:

- understandable
- believable
- supported by evidence or examples
- informative

3-9 evaluators for each text

Answers on Likert scale from 0 to 4

Document's score = mean score (from 0 to 4)

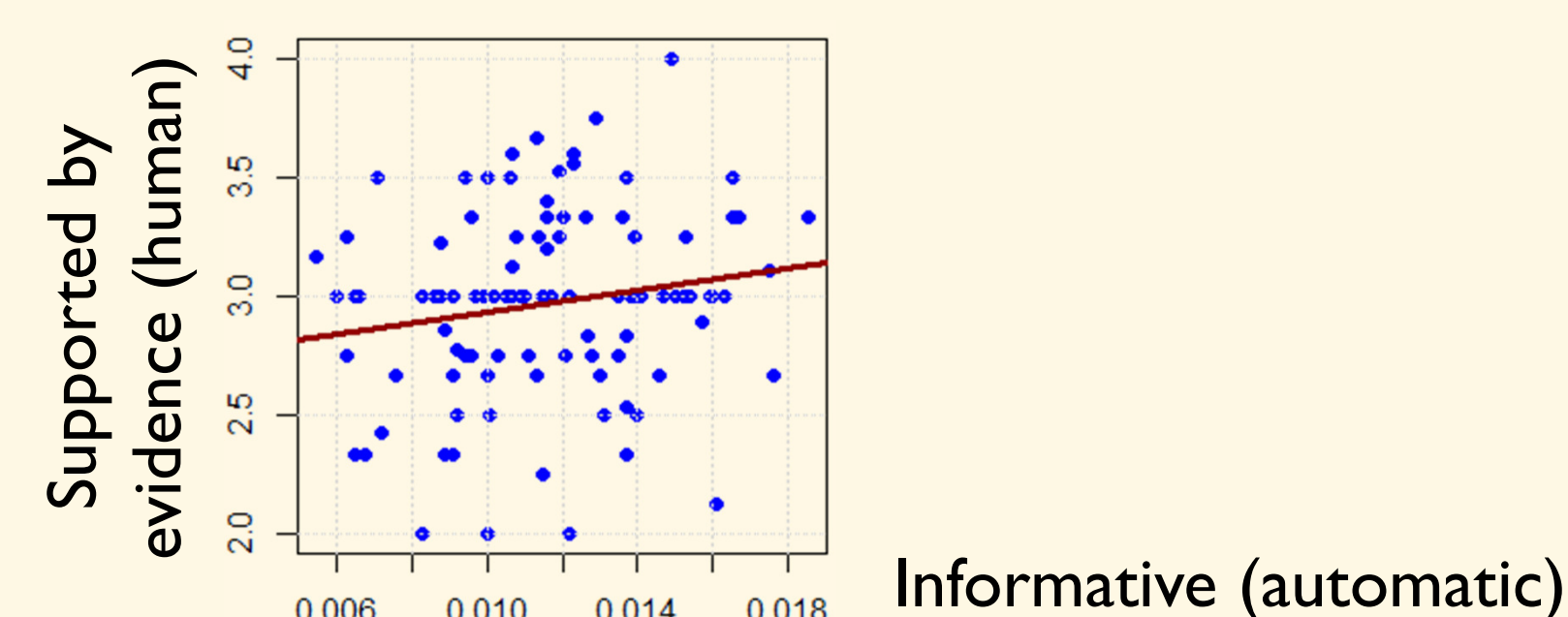
Experiments: Correlation

I. Informativeness vs. Human Judgments

NB: calculates factual density

- with 'supported by evidence and examples':

Low positive correlation: Spearman's $\rho \sim 0.2$ (91% significance)

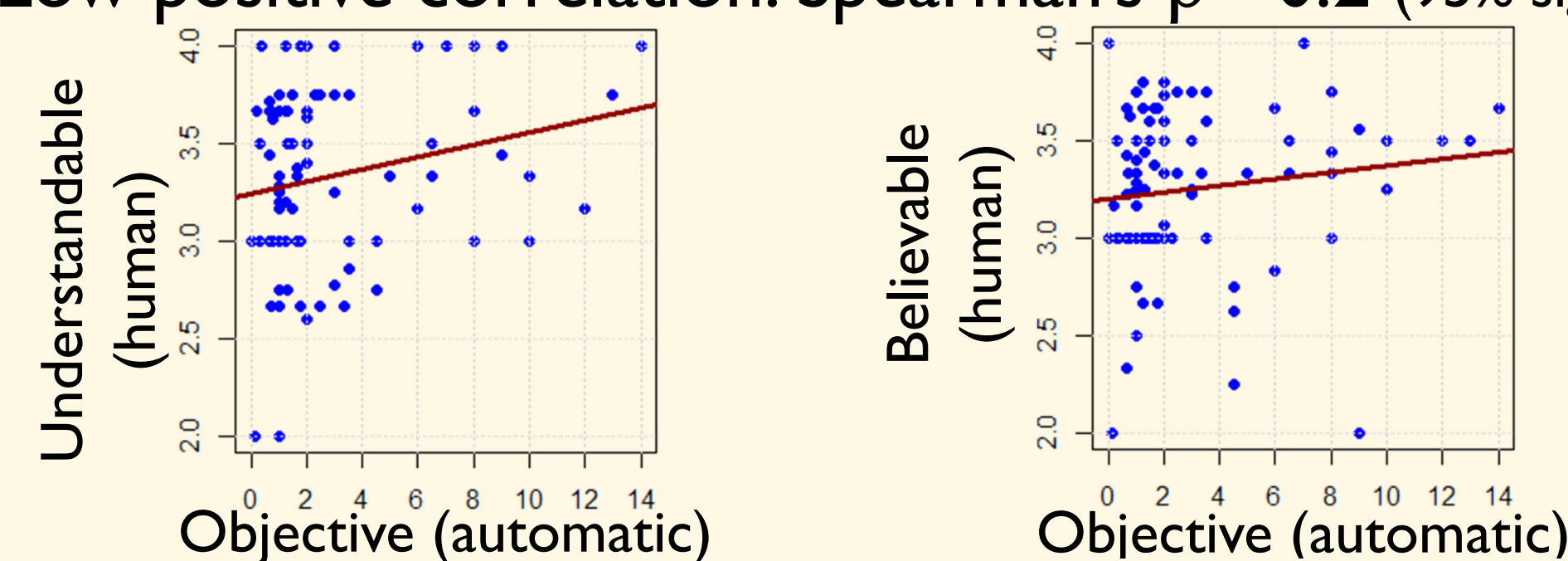


- with other characteristics: **no** significant correlation

II. Objectivity vs. Human Judgments

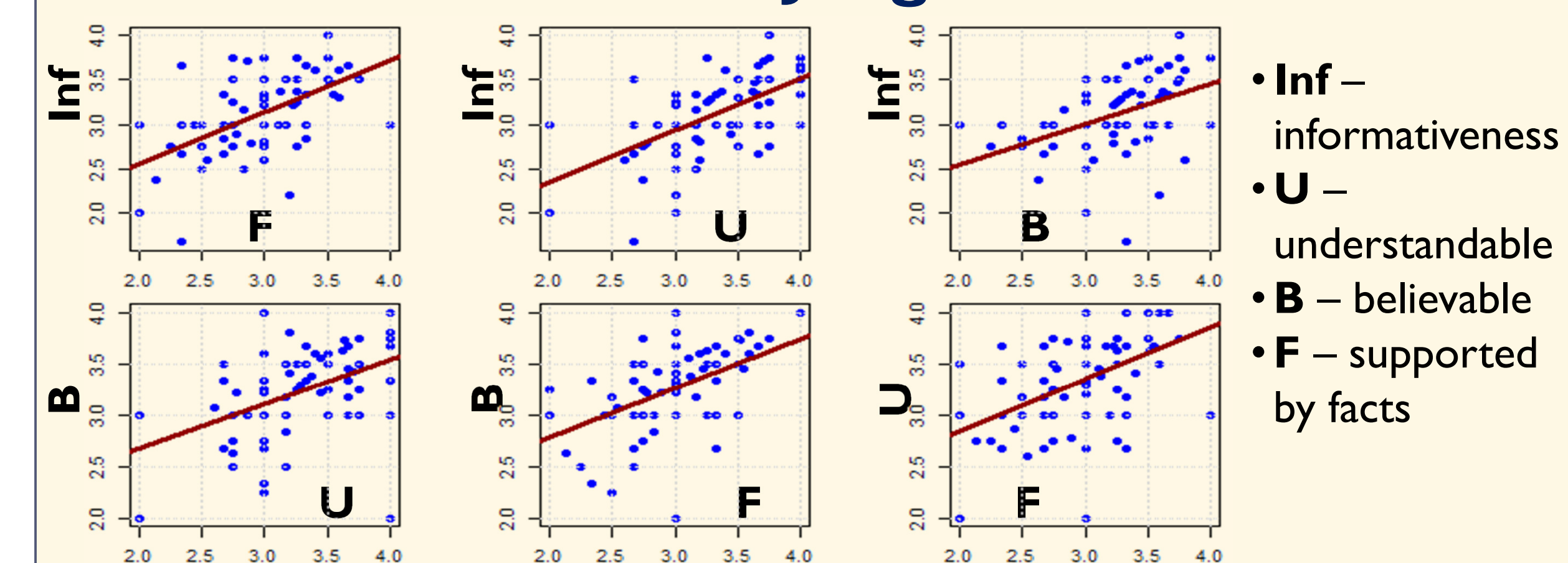
- with 'understandable' and 'believable':

Low positive correlation: Spearman's $\rho \sim 0.2$ (95% significance)



- with 'supported by evidence and examples' and 'informative':
no statistically significant correlation

III. Between Human Judgments



All pairs of characteristics show

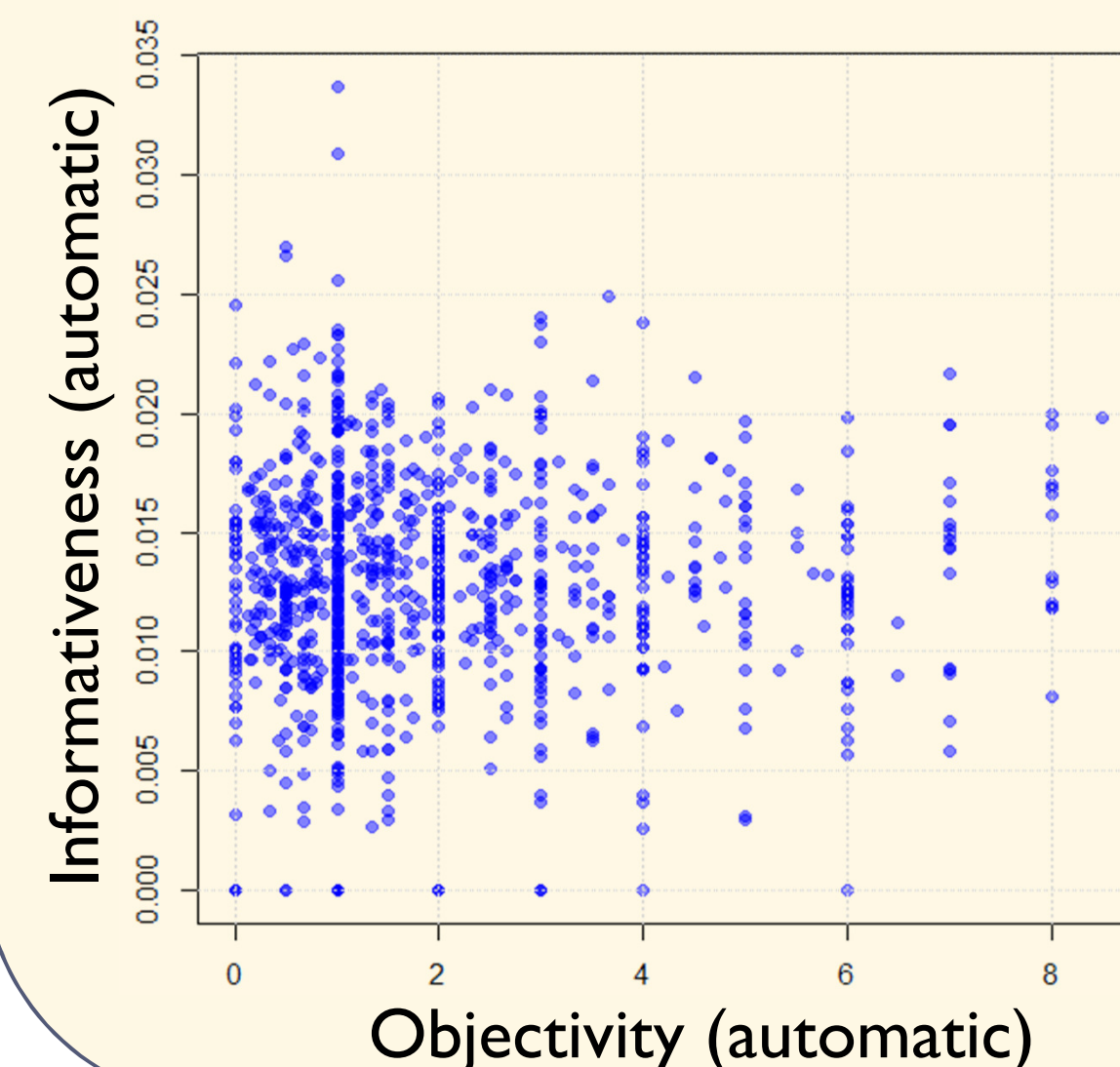
moderate positive correlation (Spearman's $\rho \sim 0.5$)
with high statistical significance (>99.99%)

Hypothesis: humans tend to judge a text uniformly:

If one likes a text, s/he tends to answer 'good' for all

IV. Automatically Measured Characteristics

Dataset: 1000 hotel reviews from Trip Advisor



No statistically significant correlation (p-value > 0.1)

Explanation:

Objective features (syntactic structures and use of subjective words) are not expected to correlate

Discussion

- Automatic methods score objectively. True features are independent.
- Human judgment is biased to "generally positive" or "generally negative" side: I (dis)like everything in this text.
- Questions intended to measure independent features but implying a scale "good/bad" lead to biased answers

Conclusions: better to formulate questions so that they are not perceived as on the scale of "good / bad", e.g.:
Is this a bright color? vs. *Is this a high intensity color?*