Sentiment Analysis

A Presentation By

2LT Zachary Szewczyk
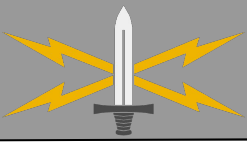
Presentation Outline

POC: 2LT Szewczyk - zachary.j.szewczyk.mil@mail.mil

An Introduction to Sentiment Analysis

A Definition

"The process of <u>computationally identifying</u> and categorizing <u>opinions</u> expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral."

- Oxford University Press

3

The Technology Behind Sentiment Analysis

Wordlists

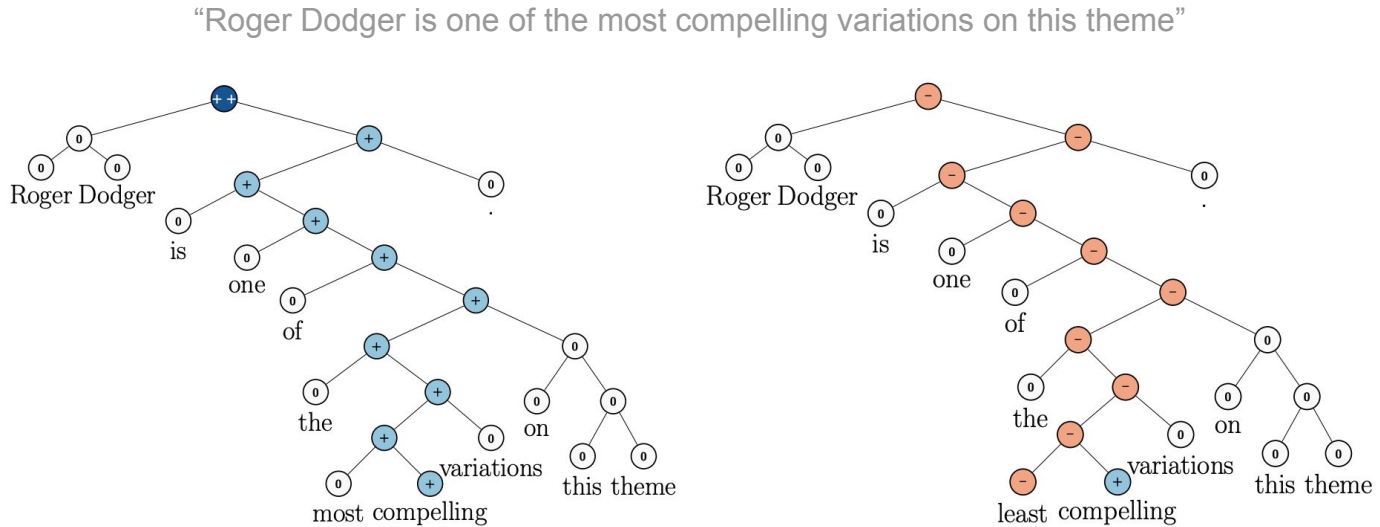| louder	-2
louts	-2
lovely	4

Figure 1: SentiStrength input scored -5 to 5

Inference

Is "Type" an actual term in the source documents …	0
With the transition to ORM all we should care of is …	-1
(btw. this field will occupy 7 bytes for our 64 bit builds …	0

Figure 2: SentiCR input scored -1, 0, or 1

Treebanks

"Roger Dodger is one of the most compelling variations on this theme"

Figure 3: Sentiment Treebank scoring in Stanford CoreNLP, from very negative (--) to very positive (++)
Ref: "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", by Richard Socher

04

Popular Tools

Stanford CoreNLP

- Sub feature of Natural Language Processing library
- <u>Resource intensive</u>
- Accurate: 85.4% correct estimation
- Under active development

- Uses Sentiment Treebank scoring on a dataset of 215,154 unique words and phrases, evaluated by 3 judges

SentiStrength

- Born of EU CyberEmotions project
- Lightweight
- Correct estimation of positive sentiment 60.6%, negative 72.8%
- <u>No longer under active development</u>

- Scores individual words

SentiCR

- <u>New tool created for Software Engineering</u>
- Lightweight, Python-based
- 83.03% accurate estimation
- Under active development

- Infers sentiment based on manual classification of 2,000 developer commit logs
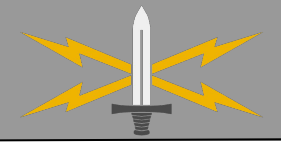
Sentiment Analysis in Industry

Recall that Sentiment Analysis is the process of:

"...<u>computationally identifying</u> and categorizing <u>opinions</u>…"

Industry uses Sentiment Analysis on:
- Social media text, in particular Twitter
- Product reviews, i.e. Amazon
- Software Engineering (SE) datasets, i.e. commit logs
- Markets

Sentiment Analysis in Cyber Operations

- Past and present:

  - No public records of Sentiment Analysis in ARCYBER

  - "Opinion mining" and "subjectivity analysis" are organic to the Army

- Looking to the future:

  - Real-time situational awareness

  - Automated Information Operations
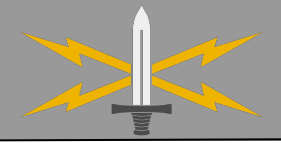
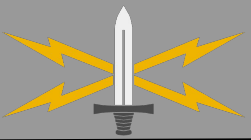# Challenges to and Shortfalls of Sentiment Analysis

- Inordinate emphasis on certain words and phrases
  - "Remove", "got rid of"
- Full sentence and individual preposition negation
  - "They did the best" vs "They did not do the best"
  - "They did the best at the best event" vs "They did the best at the worst event"
- Lack of context
- Domains
  - Social media vs software engineering
- Languages
  - Language implementation
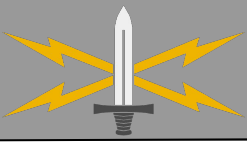
Conclusion

- Sentiment Analysis is a metric, but not the metric

- Within strict constraints, it performs well

- Outside those constraints, it performs poorly

- Nevertheless, Sentiment Analysis has a bright future

Questions

POC: 2LT Szewczyk - zachary.j.szewczyk.mil@mail.mil