

IDENTIFYING EXPRESSIONS OF EMOTION IN TEXT

Saima Aman¹ and Stan Szpakowicz^{1,2}

¹University of Ottawa, Ottawa, Canada; ²Polish Academy of Sciences, Warszawa, Poland

- **Goal** – Investigate the expression of emotion in text through a corpus annotation study
Explore computational techniques for emotion classification – we took a knowledge-based approach
- **Data** – Drawn from blog posts and manually annotated with emotion-related information
- **Applications** – Affective interfaces for HCI and CMC, Text-to-Speech systems, psychological analysis of text, personality modelling
- **Emotion categories** – happiness, sadness, anger, disgust, surprise, fear (Ekman, 1992)

Emotion annotation task

- Seed words selected for each emotion category
- Blog posts containing the seed words retrieved from the Web
- 173 posts collected (containing 5205 sentences)
- Four judges involved in annotation
- Each sentence subjected to two judgments

Types of annotations

- Emotion categories – happiness, sadness, anger, disgust, surprise, fear, mixed emotion*, no emotion
- Emotion intensity – high, medium, low
- Emotion indicators – spans of text

Example

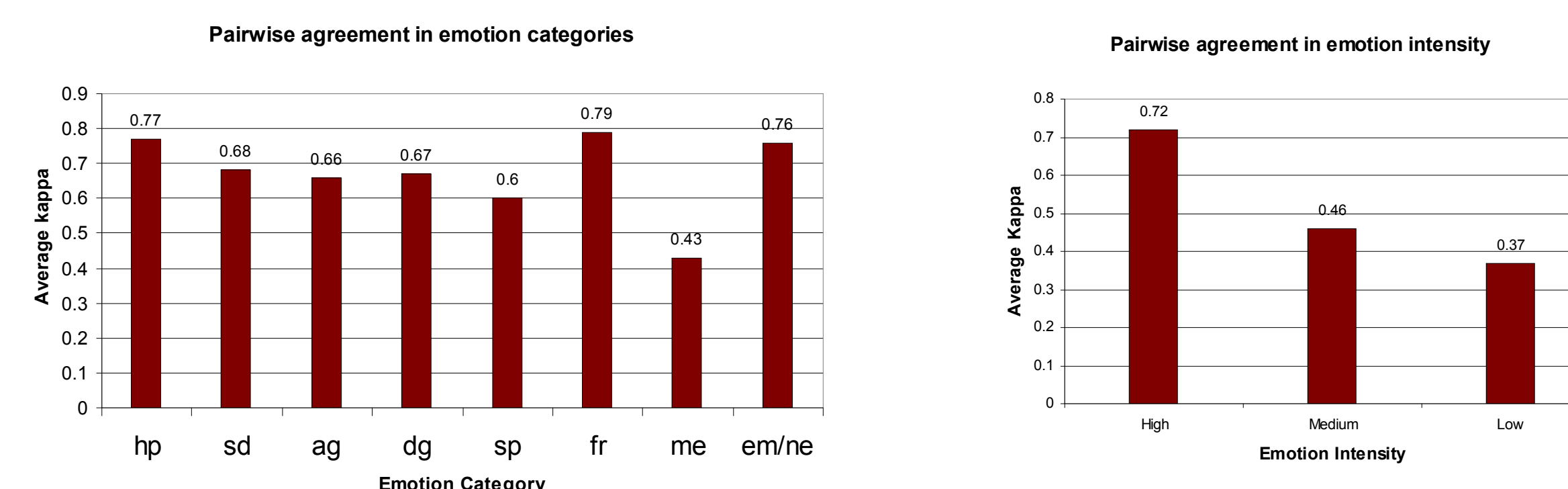
- *I have to look at life in her perspective, and it would break anyone's heart. (sadness, high)*

* Cases of more than one emotion type in a sentence, or an emotion not belonging to any category, were assigned the “mixed emotion” category.

Annotation agreement measurement

Emotion categories and intensity

- Cohen's kappa (1960) used to measure agreement



Emotion Indicators

(1) **MASI**—Measure of agreement on set-valued items (Passonneau, 2006)

A = set of emotion indicators labelled by Judge 1

B = set of emotion indicators labelled by Judge 2

$$\text{MASI} = J * M$$

$$J = |A \cap B| / |A \cup B|$$

$$M = \begin{cases} 1, & \text{if } A = B \\ 2/3, & \text{if } A \subset B \text{ or } B \subset A \\ 1/3, & \text{if } A \cap B \neq \phi, A - B \neq \phi, \text{ and } B - A \neq \phi \\ 0, & \text{if } A \cap B = \phi \end{cases}$$

(2) **I/O method**

- each word labeled **In** or **Out** of an emotion indicator
- example: “*I/O am/O very/I happy/I*”
- kappa can now be used for In and Out classes

[Avg. MASI = 0.61 ; Avg. kappa = 0.66]

Emotion classification

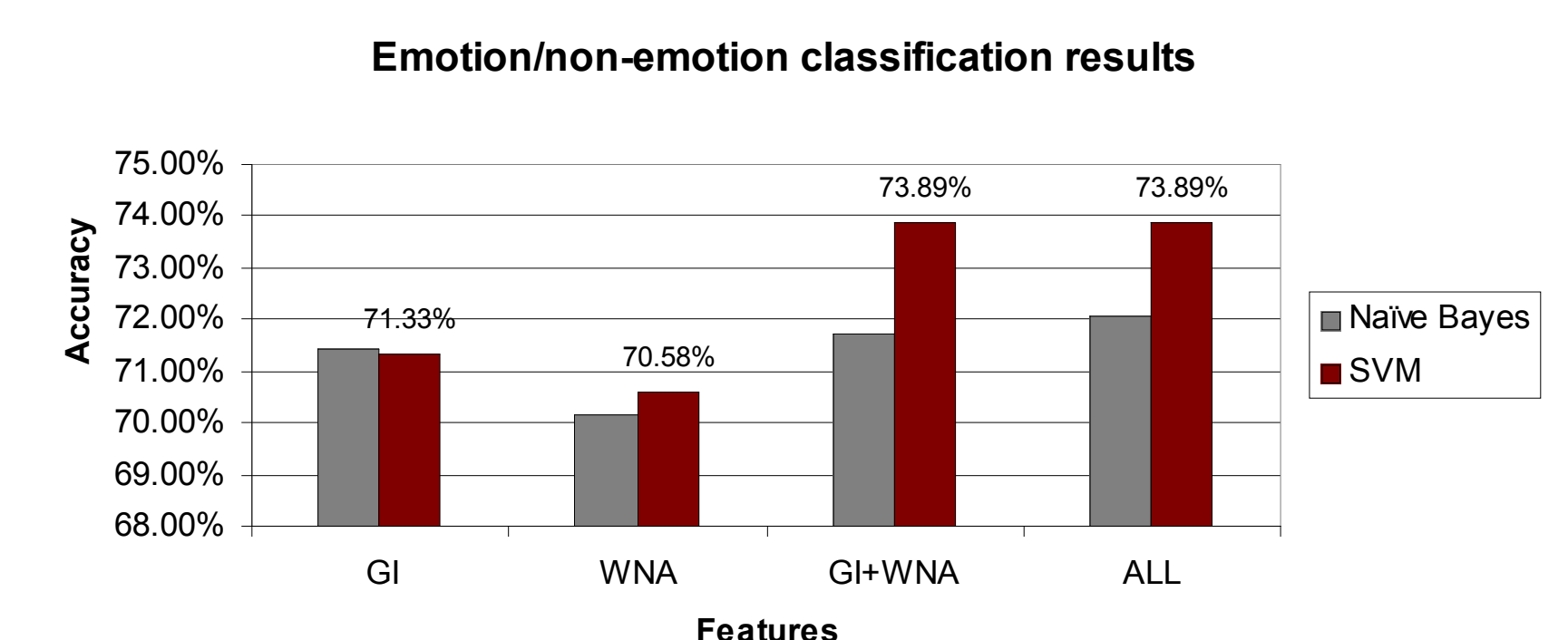
- Aim – To automatically recognize emotion sentences
- All emotion category sentences assigned to class “EM”
- # EM = 1466; # NE = 2800

Features

- General Inquirer (GI) – emotion, positive, negative, interjection, pleasure, and pain word lists
- Emotion words from WN-Affect (WNA) – happiness, sadness, anger, disgust, surprise, fear (used publicly available lists)
- Special symbols – emoticons, punctuation (“?” and “!”)

Experiments

- Baseline accuracy = 65.6% (using majority class labeling)
- Ten-fold cross validation



- **Results** – Average inter-annotator agreement on emotion categories ranged from 0.6 to 0.79
Agreement on intensity labelling higher for high-emotion sentences
Emotion classification accuracy = 73.89%, significantly higher than the 65.6% baseline
- **Future work** – Automatic fine-grained classification of emotions
Automatic recognition of emotion intensity