Recognizing Emotions in Text

Saima Aman

Master’s Thesis Presentation

Supervisor: Dr. S. Szpakowicz

University of Ottawa

2007

Agenda

Introduction
- Problem Definition
- Related Work

Data
- Emotion Annotation
- Annotation Agreement Measurement

Experiments
- Emotion/Non-emotion Classification
- Fine-grained Emotion Classification
- Emotion Intensity Recognition

Conclusions

Problem Definition

Objective
- Determine emotions expressed in text at the sentence level

Recognize Emotion Class
- happiness, sadness, anger, disgust, surprise, fear (Ekman, 1992)
- mixed emotion, no emotion

Determine Emotion Intensity
- high, medium, low, neutral

Data
- Drawn from blogs
- Manually annotated with emotion labels

Application Areas

Affective Interfaces

- make sense of emotional input
- provide emotional responses
- human-computer interaction (HCI)
- computer-mediated communication (CMC)
- e-learning systems

Text-to-Speech (TTS) Systems

- natural emotional rendering of text

Psychological Analysis of Text

- learn user preferences, inclinations, and biases
- personality modeling
- consumer review analysis

Related Work

Sentiment Analysis

- finding subjectivity, opinion, appraisal, orientation, affect, emotions
- finding polarity – positive/negative sentiment
- finding intensity – high, low, neutral

Genres

- news articles, editorials, opinion pieces (edited, professional)
- movie reviews, product reviews, blogs (unedited,informal)

Sentiment Analysis Methods

- Machine Learning methods
- Unsupervised methods

Related Work

Knowledge Sources

For identifying semantic orientation of words/phrases

- Specialized lexicons (e.g., GI, WN-Affect, SentiWordNet)
- Lexicons built using
  - domain-specific words/phrases (e.g., "great acting")
  - syntactic patterns (e.g., adverb-adj as in "very happy")
  - existing general-purpose lexicons (e.g., WordNet, Roget's)
- Corpus-driven approaches
  - PMI-IR (based on co-occurrence with similar words)
  - probabilistic sentiment scores (based on relative frequency in labeled documents)
- Contextual valence shifters
  - intensifiers, diminishers, negations

Data

Data Collection

- Used seed words for each emotion category
- 173 blog posts collected (5205 sentences)

Annotation Process

- four judges involved in the annotation process
- each sentence subjected to two decisions

Types of Annotations

- Emotion Category – {hp, sd, ag, dg, sp, fr, me, ne}
- Emotion Intensity – {h, m, l}
- Emotion Indicators (individual words / strings of words)

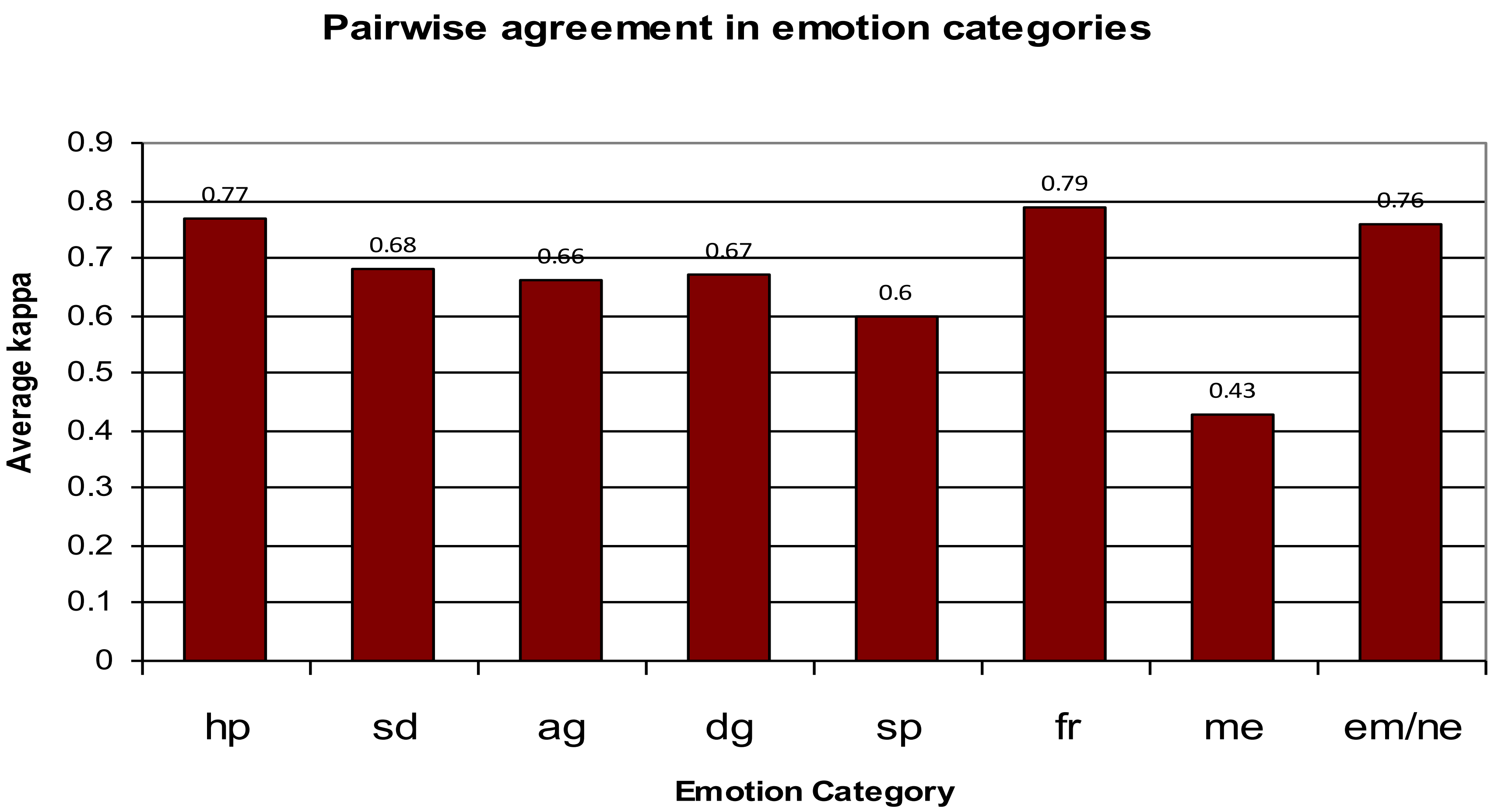
Example

But <u>all of a sudden</u> it's <u>hit</u> me that I have all this work due. (sp, h)

Annotation Agreement Measurement

Emotion Category

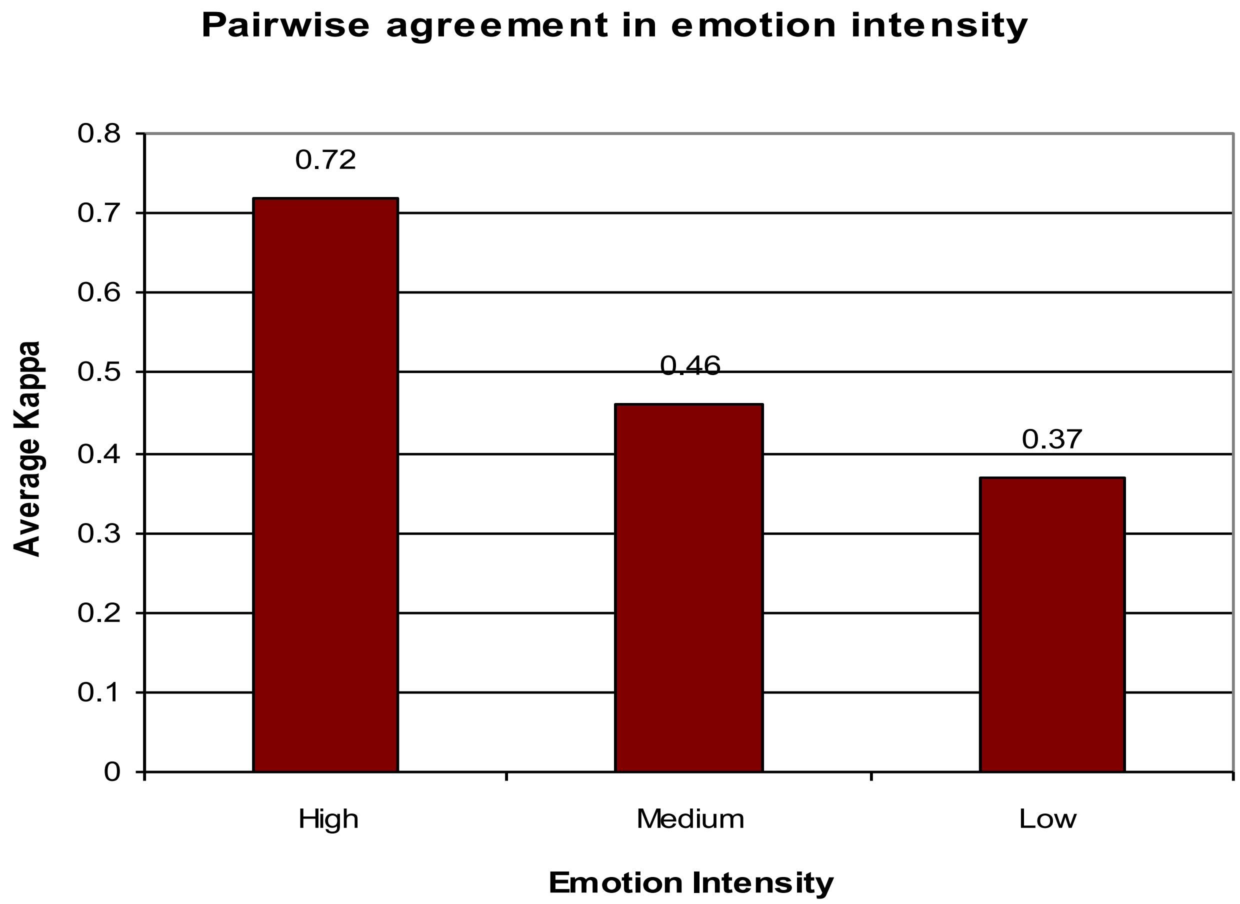
- Cohen's kappa used for agreement measurement (Cohen, 1960)



Annotation Agreement Measurement

Emotion Intensity

- Cohen's kappa used for agreement measurement (Cohen, 1960)



Annotation Agreement Measurement

Emotion Indicators

- MASI (Passonneau, 2006)

  A/B = set of emotion indicators identified by Judge1/Judge2

  MASI = J * M

  J = |A∩B| / |A∪B|

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

- I/O Method

  each word labeled (**I**n) or (**O**utside) an emotion indicator

  Example – "*I/O am/O very/I happy/I*" (kappa can be used)

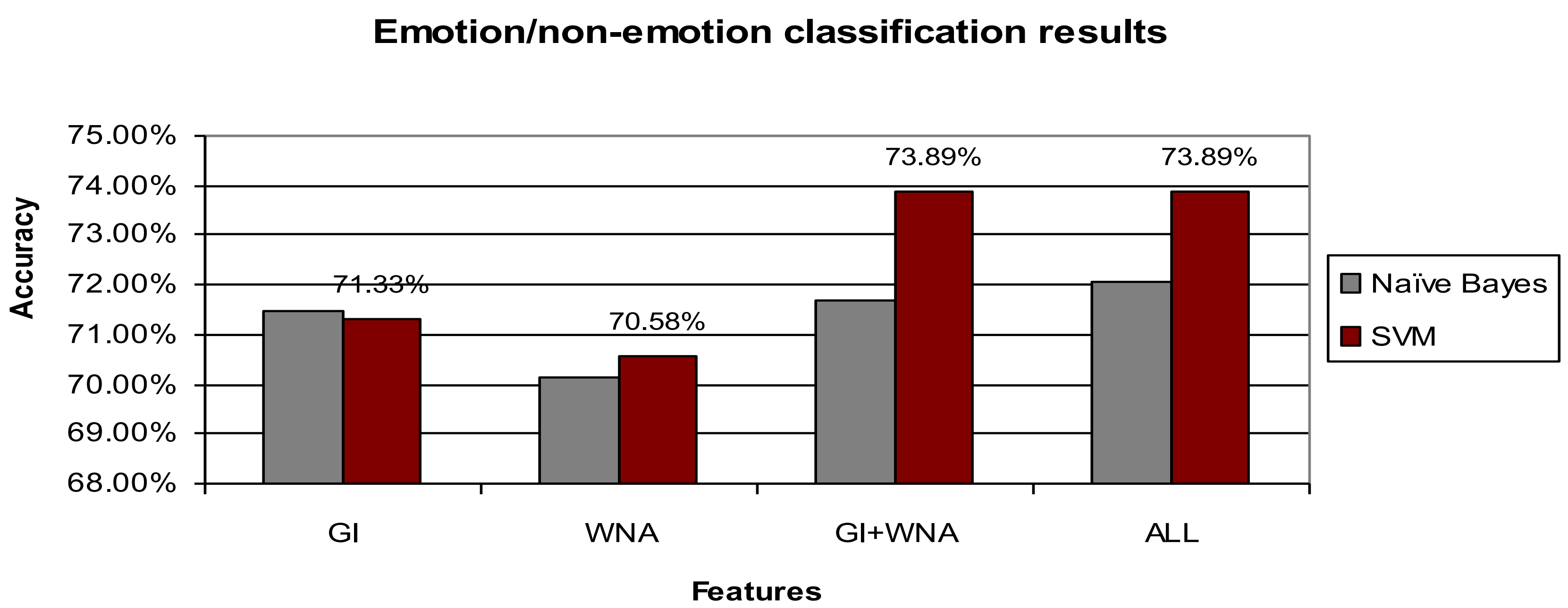
- Avg. MASI = 0.61 ; Avg. kappa = 0.66

Experiments – Emotion/Non Emotion Classification

Used ML methods – SVM and Naïve Bayes

Features

- GI – Emotion, Positive, Negative, Interjection Pleasure, Pain words
- WN-Affect – Happiness, Sadness, Anger, Disgust, Surprise, Fear words
- Special symbols – Emoticons, Punctuations ("?" and "!")



Experiments – Fine-grained Emotion Classification

Baseline

Term counting method using emotion words from WordNet-Affect

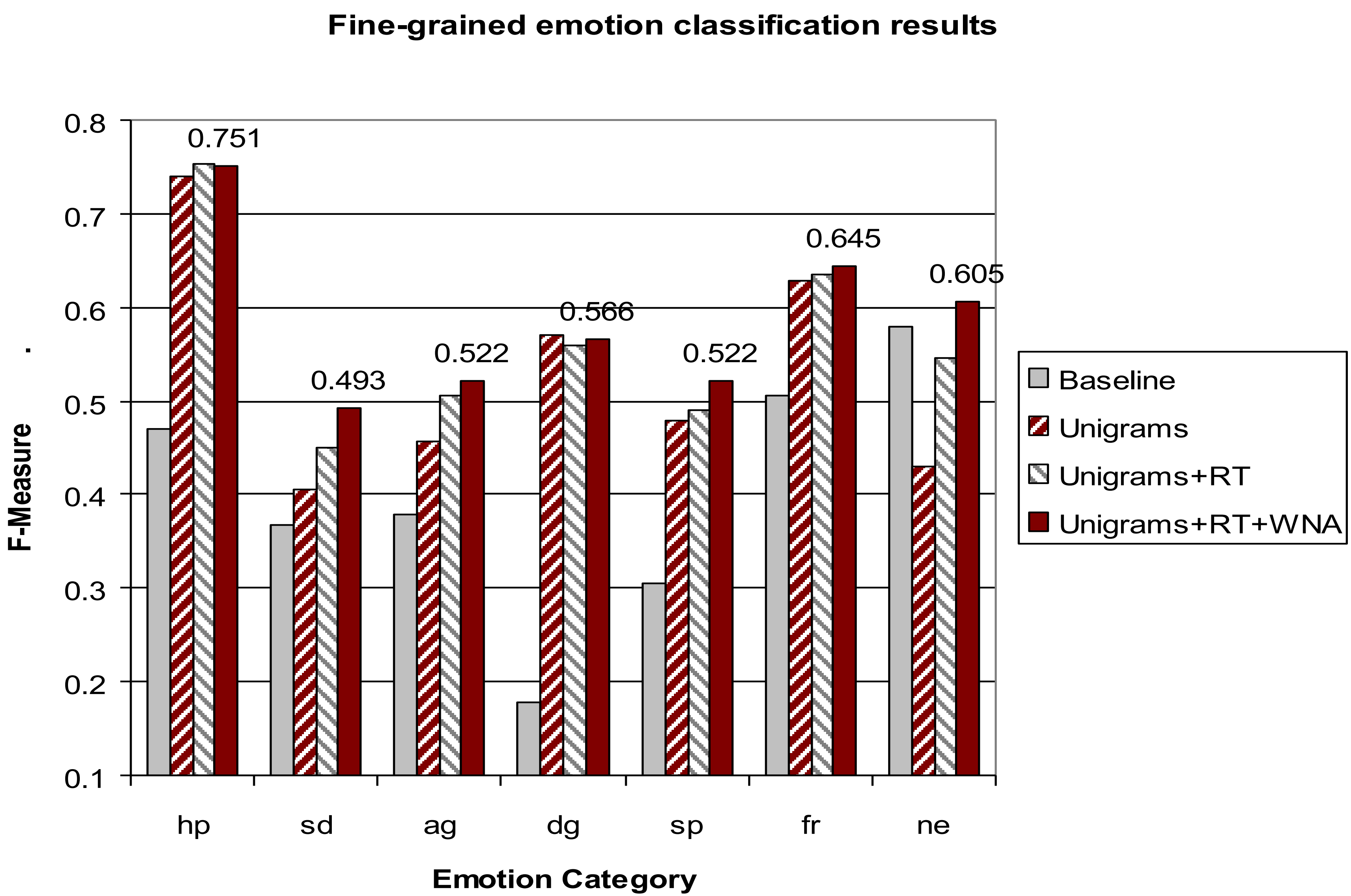
Features

- Corpus-based unigram features (excluding low-freq words and stopwords)
- Features from emotion lexicons -
  - WordNet-Affect (existing emotion lists)
  - emotion lexicon automatically built from *Roget's* Thesaurus

Lexicon from *Roget's* Thesaurus

- Words in *Rogets'* classification hierarchy considered as nodes in a network
- Related words likely to be located close to each other in the network
- They can be found using Semantic Similarity Measure (Jarmasz and Szpakowicz, 2004)
- Emotion words for each emotion category acquired by selecting words similar to {happy, sad, anger, disgust, surprise, fear}

Experiments – Fine-grained Emotion Classification



Experiments – Emotion Intensity Recognition

Emotion Intensity Modifications

- relatively weak and strong words (e.g., "*dislike*" and "*abhor*")
- intensifiers (e.g., "*very* happy", "*highly* grateful", "*much* disappointed")
- diminishers (e.g., "*little* embarrassed", *somewhat* apprehensive", "*not* pathetic")
- comparative and superlative forms of adjectives ("*happier*", "*greatest*")

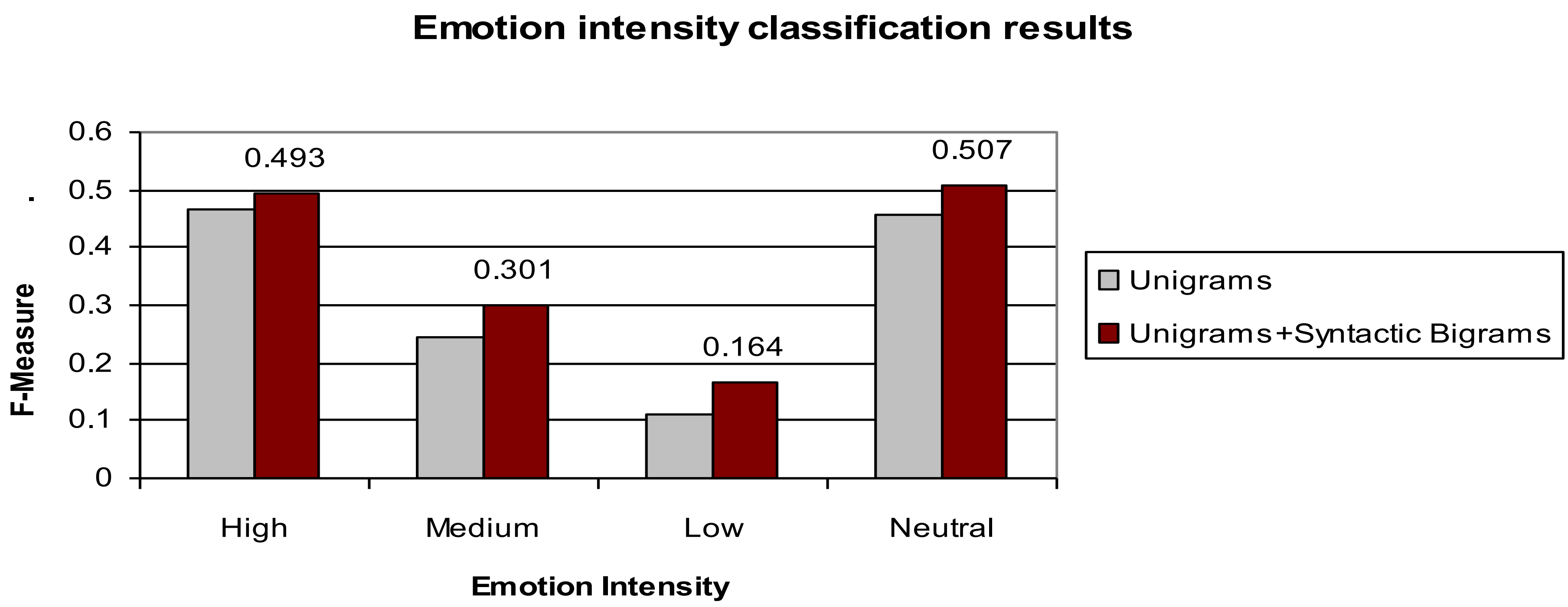
Syntactic Bigrams

- Represent English language constructs used to express and modify emotion
- Identified using the Link Parser
- Pairs of words connected by links output by the parser
- Link examples:
  - EA connects adverbs to adjectives (e.g., <more, happy>)
  - EE connects adverbes to other adverbs (e.g., <so, angrily>)
  - Other adjective and adverb related links (e.g., <awful, lot>, <much, more>)
  - Idiomatic expressions (e.g., <very, very>), etc.

Experiments – Emotion Intensity Recognition

Features

- Corpus-based unigram features (excluding low-freq words and stopwords)
- Syntactic bigrams



Conclusions

Summary

- Studied emotion expressions in text during manual annotation
- Investigated computational methods to identify the type and strength of the expressed emotion

Results

- Use of external knowledge resources helpful in determining emotion-related words
- Use of syntactic features along with the corpus-based unigram features helpful in recognizing emotion intensity

Contributions

- Prepared an emotion-labeled corpus
- Demonstrated the feasibility of applying computational methods for automatic emotion recognition
- Introduced a novel approach of automatically building Emotion Lexicon using *Roget's* thesaurus

References

[1] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20 (1): 37–46.

[2] Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion*, 6, 169-200.

[3] Jarmasz, M. and Szpakowicz, S. (2004). Roget's Thesaurus and Semantic Similarity. In N. Nicolov, K. Bontcheva, G. Angelova, R. Mitkov (eds.) *Recent Advances in Natural Language Processing III: Selected Papers from RANLP 2003*, John Benjamins, Amsterdam/Philadelphia, *Current Issues in Linguistic Theory*, 260, pages 111-120.

[4] Passonneau, R. (2006). Measuring agreement on set-valued items (MASI) for semantic and pragmatic annotation. In *Proceedings of LREC-2006*, Genoa, Italy.

Resources

[1] Jarmasz, M. and Szpakowicz, S. (2001). The Design and Implementation of an Electronic Lexical Knowledge Base. In *Proceeding of the 14th Biennial Conf. of the Canadian Society for Comp.Studies of Intelligence (AI-2001)*, Ottawa, Canada, 325-333.

[2] Stone, P.J., Dunphy, D.C., Smith, M.S., Ogilvie, D.M., and associates. (1966). *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.

[3] Strapparava, C. and Valitutti, A. (2004). WordNet-Affect: an affective extension of WordNet. In *Proceedings of LREC2004,* 1083 – 1086, Lisbon, Portugal.

Thank you!