

Using *Roget's* Thesaurus for Fine-grained Emotion Recognition

Saima Aman

SITE, University of Ottawa, Ottawa, Canada

Stan Szpakowicz

SITE, University of Ottawa, Ottawa, Canada

ICS, Polish Academy of Sciences, Warszawa, Poland



Overview

Objective

- Recognize emotive meaning of text
- Motivation – Growing interest in recognizing sentiment and emotions in text

Task

- Automatically identify emotion expressed in a sentence
- Categorize sentences into emotion classes – happiness, sadness, anger, disgust, surprise, fear (Ekman, 1992)

Data

- Drawn from blogs
- Manually annotated with emotion labels

Approach

- Machine learning experiments for emotion classification
- Corpus-based unigram features
- Features derived from Emotion lexicons



Application Areas of Automatic Emotion Recognition

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



Previous Work in Emotion Recognition

Emotion Recognition Tasks

- Classification of valence (positive/negative) and distinct emotion categories
- Classification at word-level and sentence level

Knowledge Sources

For identifying emotional affinity of words/phrases:

- Specialized lexicons (e.g., General Inquirer, WN-Affect)
- Lexicons built using
 - 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 emotion words)
 - probabilistic sentiment scores (based on relative frequency of words in emotion-labeled documents)



Emotion-labeled Data

Data Collection

- Data drawn from blogs – potentially rich in emotion
- 173 blog posts collected (5205 sentences)

Emotion Annotation Process

- four judges involved in the emotion annotation process
- each sentence subjected to two decisions
- Emotion labels – Ekman's six emotion classes, mixed emotion, no emotion

Example

This was the best summer I have ever experienced. (*happiness*)

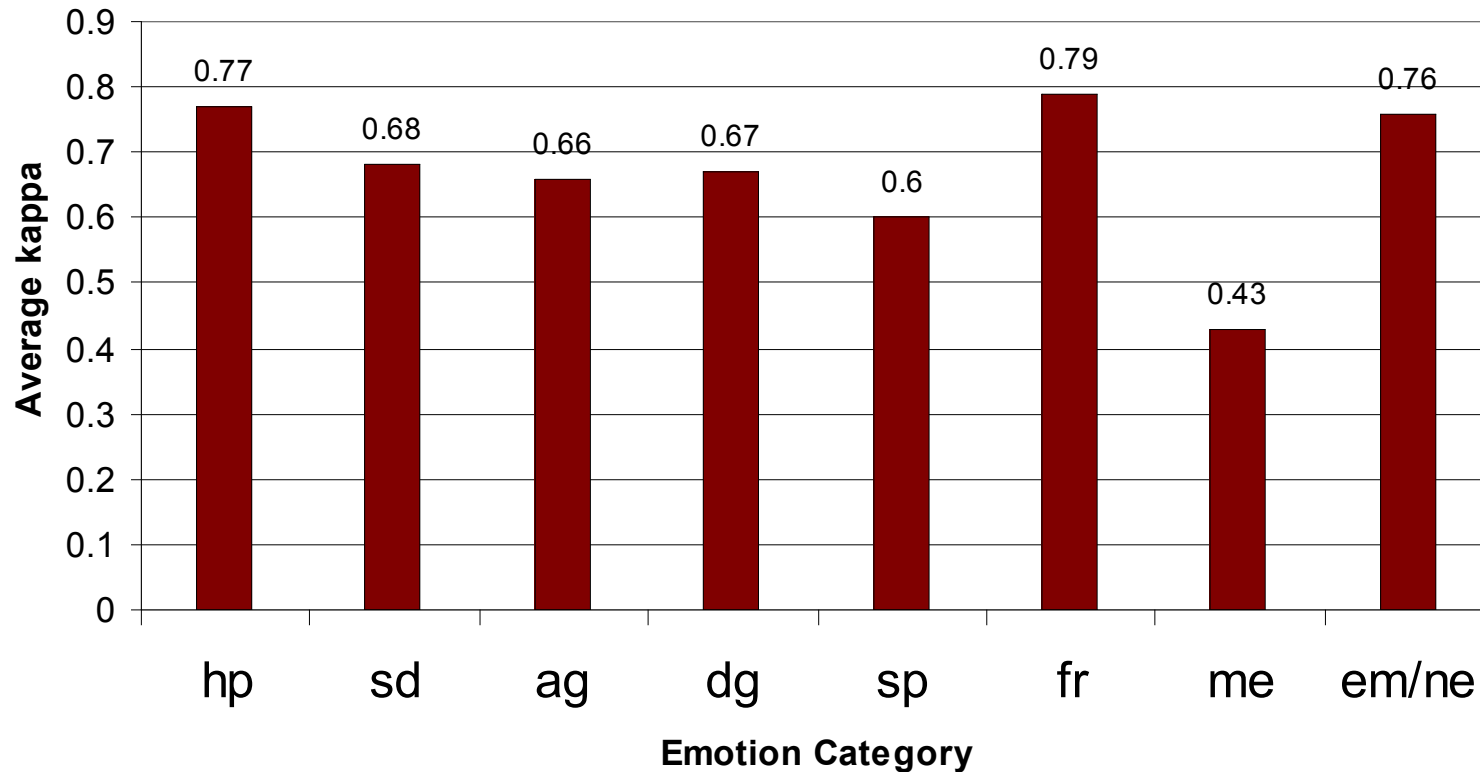


Annotation Agreement Measurement

Emotion Category

- Cohen's kappa used for agreement measurement. (Cohen, 1960)
- Average pair-wise agreement for emotion classes ranged from 0.6 to 0.79.

Pairwise agreement in emotion categories



Experiments – Emotion Classification

Baseline Approach

- Term counting method using emotion words from WordNet-Affect
- Count words of each emotion category in a sentence and assign it the category with the largest number of words

Machine Learning Approach

- 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



Building Emotion Lexicon from *Roget's*

- Goal – Build a lexicon of emotion words
- *Roget's* classification system used to infer emotion-relatedness of words
- Words in *Roget's* classification hierarchy considered as nodes in a network
- Related words likely to be located close to each other in the network
- Those words can be found using the Semantic Similarity Measure (introduced in Jarmasz and Szpakowicz, 2004) based on path lengths between nodes.
- Similarity scores vary from 0 (dissimilar) to 16 (very similar)
- Begin with a list of primary emotion words – one for each emotion category - {happy, sad, anger, disgust, surprise, fear}
- Cut-off score for similarity was chosen as 12 (based on previous studies)
- All words with score higher than 12 w.r.t. primary emotion words included in the lexicon
- A large variety of emotion-related words of different POS identified



Emotion Lexicon from *Roget's* - sample words

Similarity Score	Happiness	Sadness	Anger	Disgust	Surprise	Fear
16	family, home, friends, life, house, rest, loving, bed, partying, pleasure	crying, lost, wounds, bad, pills, falling, messed, spot, unhappy	pride, fits, stormed, abandoned, bothered, mental, anger	shock, disgust, dislike, loathing	plans, catch, expected, early, slid, slipped, earlier, caught, act	nervous, cry, terror, panic, feelings, run, fog, fire, turn, police, faith
14	love, like, feel, pretty, lovely, better, smiling, nice, beautiful, hope, cutest celebrations	ill, bored, feeling, ruin, blow, down, wrong, awful, evil, worry, death, bug	hate, burn, upset, dislike, wrong, blood, ill, flaws, bar, bitter	hate, pain, horrifying, ill, pills, sad, wear, blood, appalling, end, work, bad, regrets	left, swing, noticed, worry, times, amazing, break, interesting	falling, life, stunned, pay, broken, hate, blast, times, hanging, broken
12	gift, treats, adorable, fun, hug, kidding, bigger, great, lighting, won, stars, enjoy, favourite	defeat, nasty, boring, ugly, loser, end, victim, sick, hard, serious, aggravating	lose, throw, offended, hit, power, feel, flaring, pills, broken, life, forgot, ranting	feel, fun, lies, drawn, lose, missed, deprived, lack, sighs, defeat, down, hurt	realize, pick, wake, sense, jumped, new, late, magic, omen,	fearful, spy, night, upset, chased, hazardous, tomorrow, victim, grim, terrorists,



ML Experiments

- Used Support Vector Machines (SVM) for emotion classification experiments

Feature groups tested

- Unigrams - Corpus based unigram features
- RT - All words in the emotion lexicon acquired from *Roget's* Thesaurus
- Unigrams + RT
- Unigrams + RT + WordNet-Affect

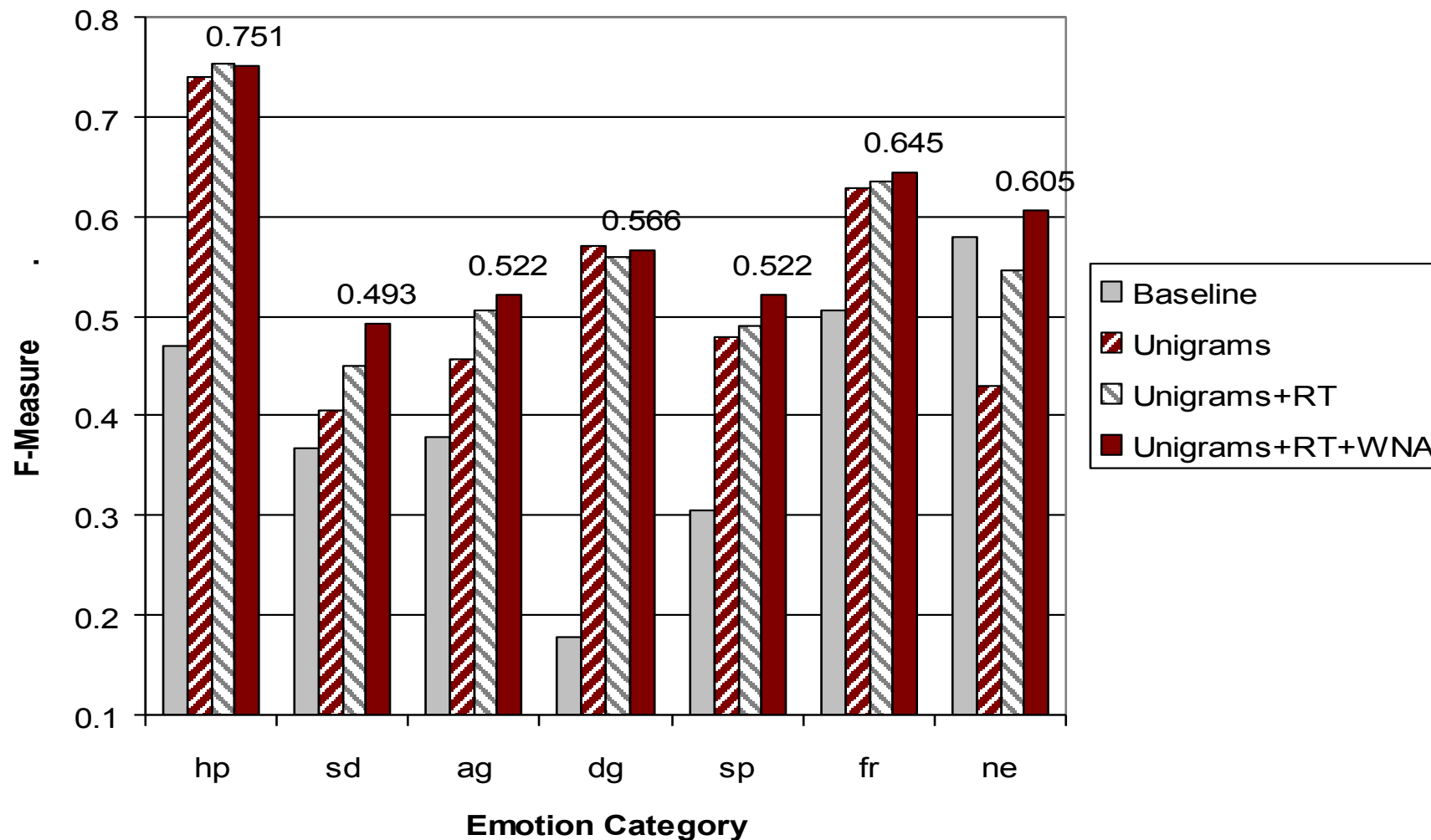
Results

- Highest recall values achieved when all features are combined
- The resulting F-measure values surpass baseline values for all emotion classes



ML Experiments - Results

Fine-grained emotion classification results



Conclusion

- Any automatic method of recognizing emotions should take into account a wide variety of words that are semantically related to emotions
- Some words are obviously affective, while many more are potentially affective depending on their conceptual notions in human psyche (e.g. home, family)
- Use of external knowledge resources (Roget's and WN-Affect) helpful in determining emotion-related words

Contributions

- Demonstrated that a combination of corpus based unigram features and features derived from emotion lexicons can help distinguish basic emotion classes in text
- 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.

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



Thank you!

