Textual Emotion & Affect Computing Literature Review

November 13, 2013

UCI Computation of Language Laboratory

by Kristine Lu

State of Current Research

- What kinds of questions are researchers in the field interested in?
- What techniques and methods are utilized?
- How does the article interact with other researchers in the field or related sub-areas (i.e., sentiment analysis)? Other disciplines?

Challenges of Interdisciplinarity

- Researchers from fields ranging from:
  - Computer Science
  - Information Sciences
  - Linguistics
  - Psychology & Cognitive Sciences (Neuroscience)

Challenges of Interdisciplinarity

- Researchers from fields ranging from:
  - Computer Science
  - Information Sciences
  - Linguistics
  - Psychology & Cognitive Sciences (Neuroscience)

- What challenges are run into when researchers come to the field from different disciplines?

General Considerations

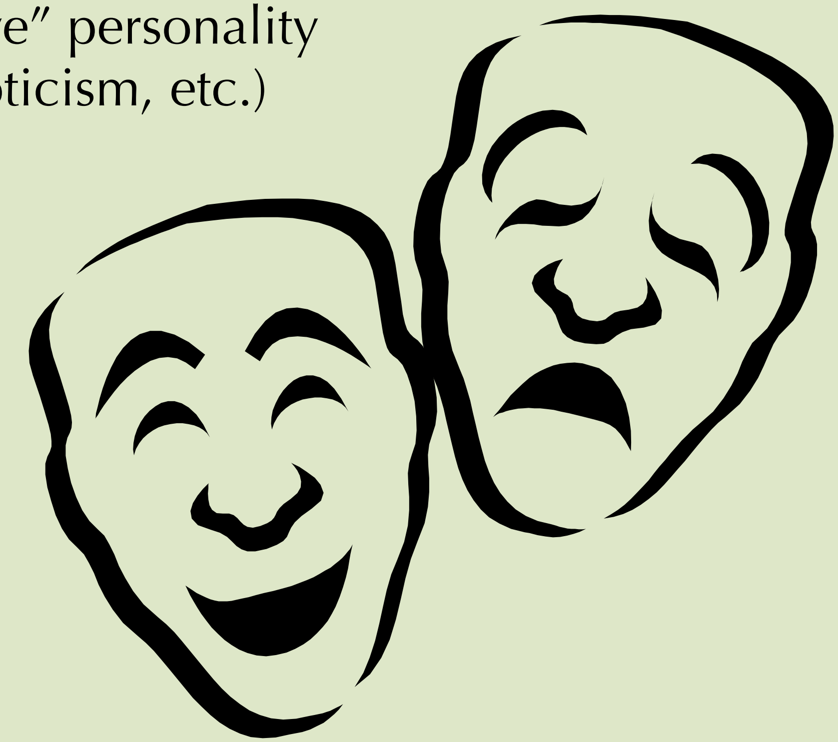
- Considerations that apply across the board:
  - What kind of Natural Language Processing and Statistical Methods are used?
    - General computational approach. Every study will use at least one, but some will focus on improving computational method

General Considerations

– What data set is used?
  - Chaffar 2011: utilized a heterogeneous emotion-annotated data set combining news headlines, fairy tales, blogs
  - Some studies use LiveJournal. Convenient because LiveJournal already includes writer-determined moods (Mishne, 2005; Keshtkar et al., 2009)
  - The OpenMind CommonSense Corpus: Large-scale real-world knowledge about people's common affective attitudes.

General Considerations

– Emotion/affect categories
  - Personality traits (Mairesse 2007)
    - » For example, the "Big Five" personality traits (extraversion, neuroticism, etc.)



General Considerations

- Emotion/affect categories
  - Personality traits (Mairesse 2007)
    - » For example, the "Big Five" personality traits (extraversion, neuroticism, etc.)
  - 6 basic emotions (Ekman)
    - Most frequently used. Includes: happy, sad, fearful, angry, disgusted, surprised

General Considerations

- Emotion/affect categories
  - Personality traits (Mairesse 2007)
    - For example, the “Big Five” personality traits (extraversion, neuroticism, etc.)
  - 6 basic emotions (Ekman)
    - Most frequently used. Includes: happy, sad, fearful, angry, disgusted, surprised
  - 9 affect categories (Izard 1971)
    - Anger, disgust, fear, guilt, interest, joy, sadness, shame, surprise

General Considerations

- Emotion/affect categories
  - Personality traits (Mairesse 2007)
    - » For example, the “Big Five” personality traits (extraversion, neuroticism, etc.)
  - 6 basic emotions (Ekman)
    - Most frequently used. Includes: happy, sad, fearful, angry, disgusted, surprised
  - 9 affect categories (Izard 1971)
    - Anger, disgust, fear, guilt, interest, joy, sadness, shame, surprise
  - LiveJournal moods: In LiveJournal, users can choose from 132 moods, hierarchically classified into 5 levels.
    - Can also look at a subset of 132 moods: Mishne (2005) only considered top 40 (amused, tired, happy, cheerful)
    - 5 Levels example:
      - » Level 1: sad, 2: uncomfortable, 3: exhausted, 4: tired, 5: sleepy

General Considerations

- What features are used?
  - "When designing a classification experiment, the most important decision – more important than the choice of the learning algorithm itself – is the selection of features to be used for training the learner" (Mishne 2005)
- Examples:
  - Frequency counts
  - Length-related (length of bytes, length of entry)
  - Semantic orientation (moods clearly negative or clearly positive)
  - More complex: Pointwise Mutual Information (PMI) – the measure of degree of association between two terms – for example, the association between words used in an entry and various moods.

Some Theoretical Considerations

– Differences amongst **emotion, mood, affect** (Davis 2010)
  - Important for those interested in computational cognitive sciences
    – i.e., Where do you draw the line between what machines can do and what is "felt" by humans only?

Some Theoretical Considerations

- Differences amongst **emotion, mood, affect**
  - Important for those interested in computational cognitive sciences
    - i.e., Where do you draw the line between what machines can do and what is "felt" by humans only?

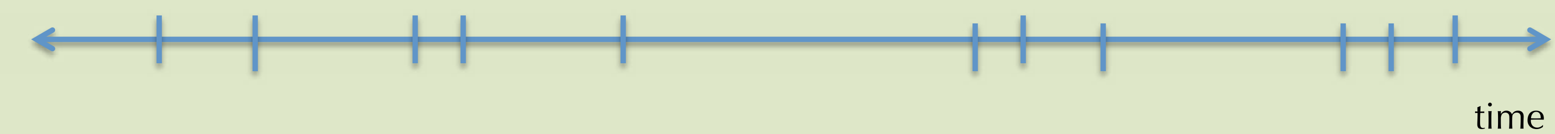
Some Theoretical Considerations

- Emotion
  - Emotions occur with a rapid onset, with little awareness and with involuntary changes in expression and physiology.
  - Short duration
  - Related to rewards and punishments (more later)
- Mood
  - refers to a longer-term affective state. A mood may arise when an emotion is repeatedly elicited
  - Long duration

Some Theoretical Considerations

Emotion vs. Mood

Mood =



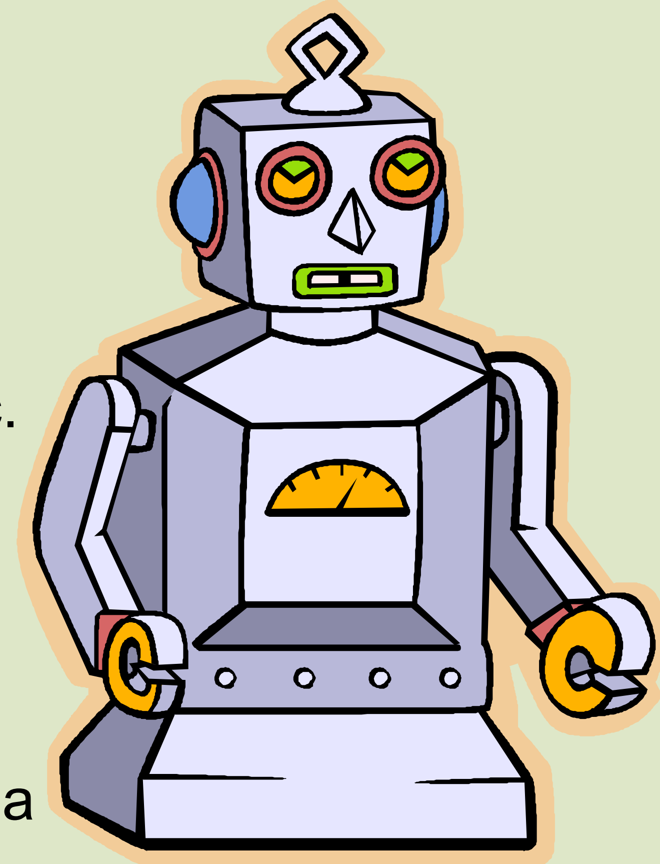
= 1 emotion

More Theoretical Considerations

– Emotion vs. Affect
  - Emotion in “goal-based” theories:
    – Described in terms of goals and roles
    – ‘‘a state usually caused by an event of importance to the subject’’ (Oatley 1992).
    – Involves mental states directed towards an external entity, physiological change, facial gestures and some form of expectation

More Theoretical Considerations

- Affect
  - Affect is a dimension of emotion that reflects valence - positive, negative or neutral valences.
  - Not necessarily accessible to conscious thought and is not necessarily describable using experiential linguistic concepts and labels such as hate, fear, surprise, etc.
- How Affect is Important:
  - Valence is likely to be an important feature for automatic detection of emotion
  - For example, valence: assigning a value between -1.0 to 1.0 indicating if a word has positive or negative connotations (Liu 2003)



Integrating Areas of Expertise

- Categories of current research
  - We propose categories by which to group studies to offer a clearer understanding of their main contributions to the field

Pearl, Lu, Barouni (PLB) Categories of Approaches

- Keyword Level (Linguistics/Social Psychology)
- Linguistic (Linguistics)
- Statistical NLP (Computer Science)
- Handcrafted Models (Mix, but will typically involve sophisticated computer science methods)

- Developed from categories of Liu et al. (2003)
- These are features that any study can have. May involve more than one.

Keyword Level

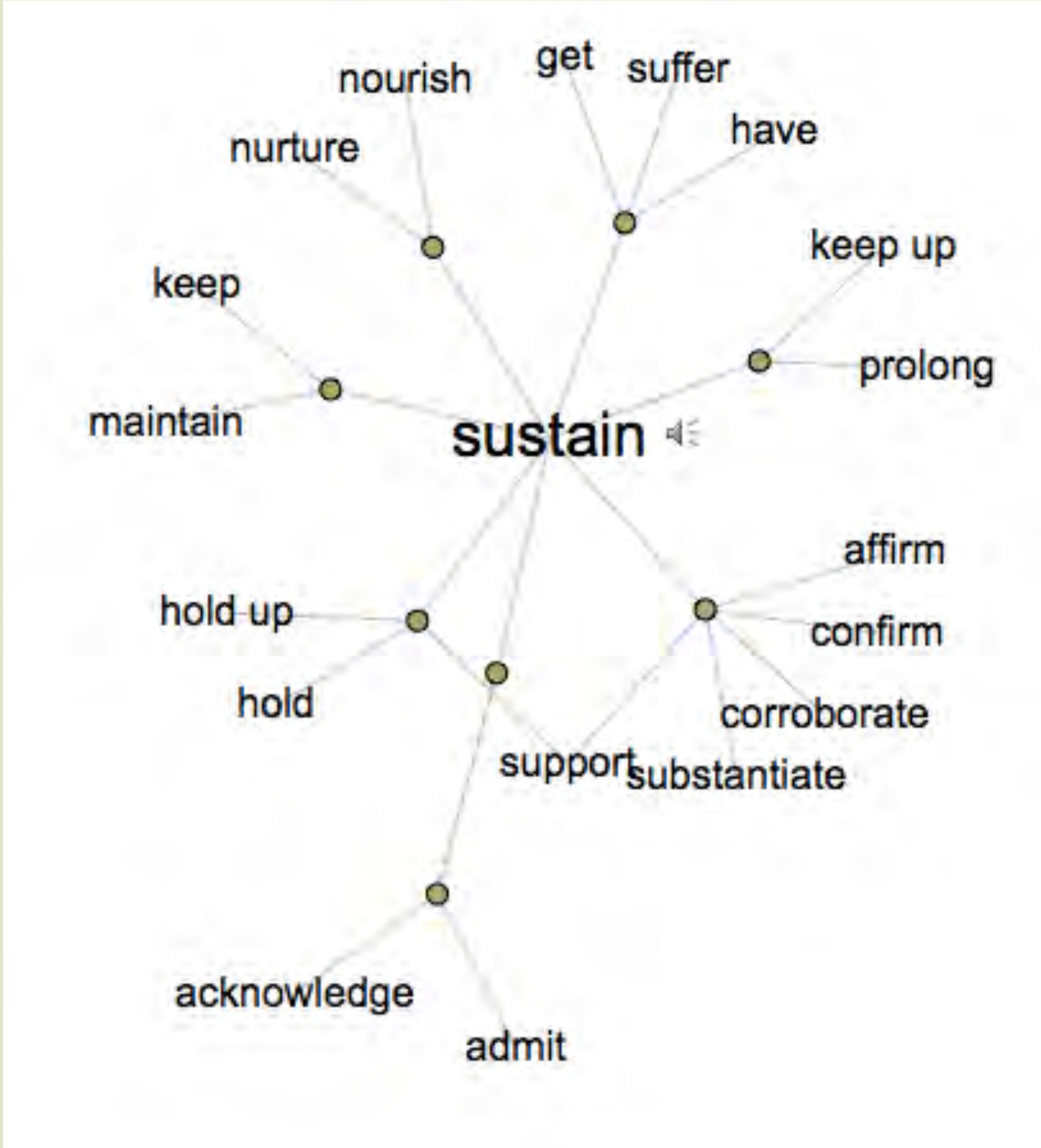
- Two main methods utilized at this level:

1) Keyword spotting: most naïve approach, but most popular
   - Text classified into affect categories based on unambiguous affect words (i.e., distressed, happy)

- Tools: Affective Lexicon (Ortony)
- Pros: Easy to use and implement
- Cons: poor recognition with negation ("not happy"), surface features

Keyword Level, continued

2) Lexical Affinity: slightly more sophisticated than keyword spotting
   - Assigns arbitrary words a probabilistic "affinity" for a particular emotion, trained linguistically (i.e., a visual thesaurus: "happy," and "glad")


nourish
get
suffer
have
nurture
keep up
keep
prolong
maintain
sustain
affirm
hold up
confirm
hold
corroborate
support
substantiate
acknowledge
admit

Keyword Level, continued

2) Lexical Affinity

Tools: Sentiment Analysis, WordNet Affect (Strappavara)

- Pros: Easy to implement (a little more work than keyword spotting, need to infer based on a list or lexicon)
- Cons: poor recognition of/can be easily tricked by negation or other word senses (“not happy”); biased toward particular genre, making it difficult to develop domain-independent model

Example: Mairesse et al., 2007

- Automatic recognition of personality traits (how personality affects linguistic production). Recognition of Big Five personality traits in conversation and text.
  - Big Five personality traits:
    - Extraversion vs. Introversion (sociable, assertive, playful vs. aloof, reserved, shy)
    - 2. Emotional Stability vs. Neuroticism (calm, unemotional vs. insecure, anxious)
    - 3. Agreeableness vs. Disagreeable (friendly, cooperative vs. antagonistic, faultfinding)
    - 4. Conscientiousness vs. Unconscientious (self-disciplined, organised vs. inefficient, careless)
    - 5. Open to experience vs. Closed to experience (intellectual, insightful vs. shallow, unimaginative)

Example: Mairesse et al., 2007

- Features used:
  - Semantic features = LIWC word categories (i.e., anger words, metaphysical issues, physical state/function, inclusive words, etc.)
    - LIWC: Linguistic Inquiry Word Count – mostly lexical affinity – a word count utility developed by Pennebaker et al., 2001 that groups words into syntactic and semantic (emotion) categories

Example: Mairesse et al., 2007

- Features used:
  - MRC (Medical Research Council) psycholinguistic features (associate each word with a numerical value): i.e., Concreteness: Low = patience, candor; high = ship.
    - MRC Psycholinguistic Database is a machine usable dictionary containing 150837 words with up to 26 linguistic and psycholinguistic attributes for each - psychological measures are recorded for only about 2500 words.

Example: Mairesse et al., 2007

- Success?
  - Openness to experience produces the best ranking model
    - Streams of consciousness, or more generally personal writings, are likely to exhibit cues relative to the author's openness to experience.
    - Extraversion is the easiest trait to model from spoken language, followed by emotional stability and conscientiousness
- Features very important – features useful for different traits, though LIWC useful for all traits
- They conclude: "It is not clear whether the accuracies are high enough to be useful" 😐

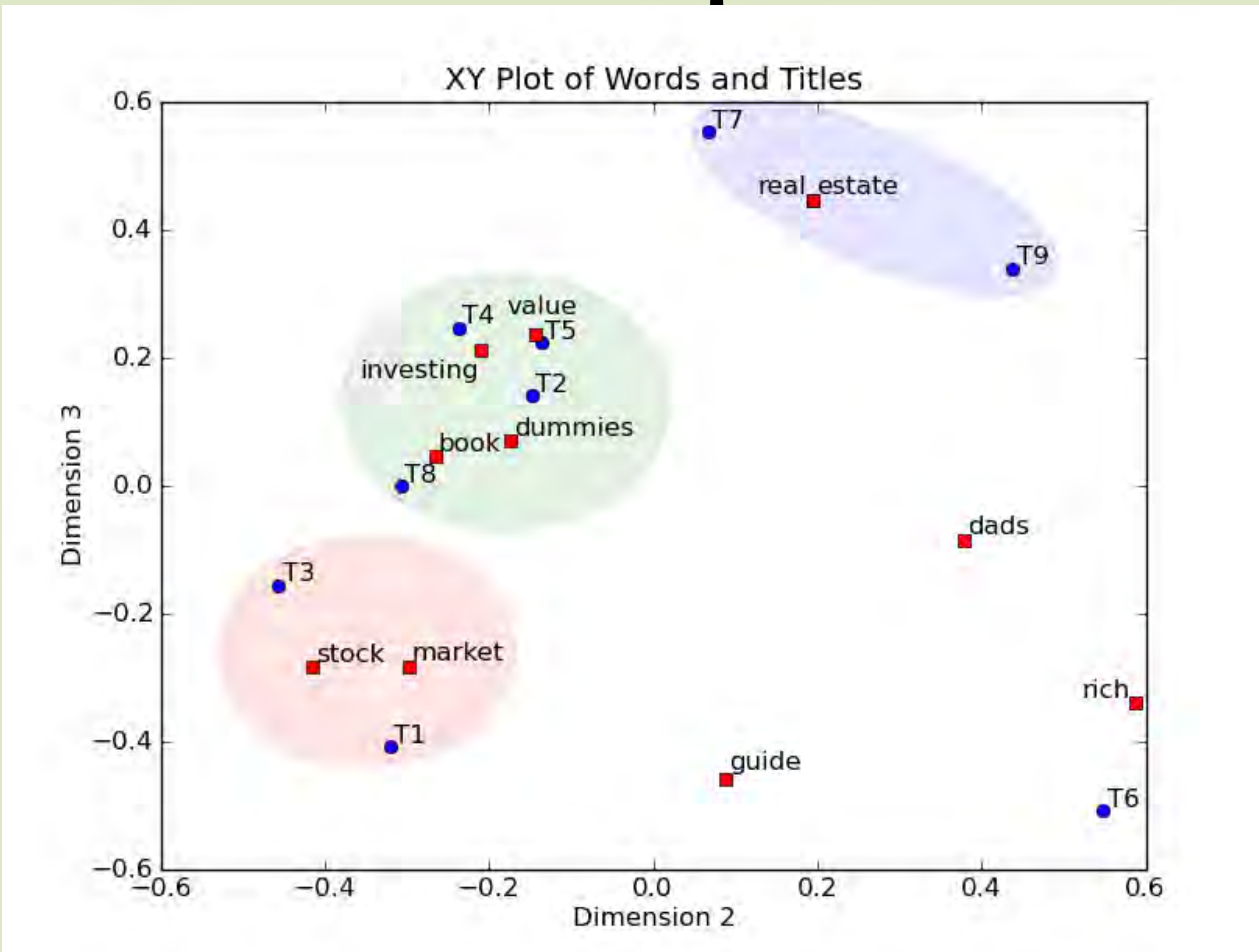
2. Statistical Natural Language Processing

- Feeding a machine learning algorithm a large training corpus of affectively annotated texts
- Nearly all of studies will use Statistical NLP in one form or another
- Pros: Can learn affective valence and also valence of other arbitrary keywords, punctuation, and word co-occurrence frequencies
- Cons: Usually semantically weak, so statistical text classifiers only work with large text input. Won't work well on smaller units like sentences

Examples of Statistical NLP Techniques

- Latent Semantic Analysis (LSA)
  - Decomposition so you can tell how far apart words are in semantic space. Giving you a way to plot words in high-dimensional space.
  - Axes with words of interest. LSA can give you a sense of how far away the different words are to each other. Can also tell you what words are important to track. Quantify how alike and not alike things are. A cool way of implement lexical affinity.
  - How far away it is = confidence level.

Examples of Statistical NLP Techniques



Examples of Statistical NLP Techniques

- N-grams
  - Unigrams and bigrams (Alm et al., 2005; Aman and Szpakowicz, 2007; Neviarouskaya et al., 2009; Chaffar and Inkpen, 2011)
- Bag of Words (BOWs)
  - Text represented as unordered collection of words, disregarding grammar and word order
- *But*…N-grams and Bag of Words tend to be experimentally weaker. Sometimes, adding features from Keyword Level approaches make NLP approaches more effective (Saif 2012)

More examples of Statistical NLP Techniques

- Classifiers
  - Support Vector Machines (SVM)
    - Supervised learning models with associated learning algorithms that analyze data and recognize patterns
      - Can train it on any kind of implementation. For example, SMO (Chaffar 2011) or other kinds of classifiers
    - SVMs are popular in text classification since they scale to the large amount of features often incurred
  - Naïve Bayes Classifier
    - Probablistic classifier based on applying Bayes's theorem with strong independence assumptions (all features considered independently)

Saif, Mohammed 2012

- Showed that even in supervised settings, an affect lexicon can provide significant gains.
- While n-gram features tend to be accurate, they are often unsuitable for use in new domains; On the other hand, affect lexicon features tend to generalize and produce better results than n-grams when applied to a new domain.
- Successful? Affect + n-gram features performed best (515/1290 guessed correctly), affect features only performed 2nd best (439/1290), and n-gram features only performed worst (375/1290). ☺

Mishne 2010

- Preliminary work on classifying blog text according to the mood reported by its author during the writing (LiveJournal)
- Obtain modest, but consistent improvements over a baseline
- Used many features: bag-of-words; Part-of-Speech tags; lemmas (POS and lemmas both from TreeTagger); unigrams (due to constraints for paper); length of blog post; semantic orientation features; Mood Pointwise Mutual Information Retrieval (PMI-IR); Emphasized words; Special symbols (Emoticons).
- Used SVMlight for experiments.
- Significant contributions: LiveJournal provided 40 moods for classification.
- Successful? Low accuracy both for human and machine annotation.

  Possibly to do with training data size, however. ☹
  - May also have to do with use of light version of SVM or shallow linguistic & domain features

Keshtkar et al., 2009

- A novel approach for using the hierarchy of possible moods to achieve better results than a standard machine learning approach. Introduced a hierarchical approach to mood classification (132 moods from LiveJournal – beyond Mishne's 40)
- Used 5 "levels" of 132 moods provided by LiveJournal service in attempt to improve on Mishne's near-baseline accuracy. Used SVM algorithm
  - Levels, an example:
    - Level 1: sad, 2: uncomfortable, 3: exhausted, 4: tired, 5: sleepy
- Successful? Hierarchical approach leads to substantial performance improvement over flat classification ☺

Linguistic Approaches

- Linguistic: Involving structural knowledge of some kind. Syntax/semantics relationship
- Pros: Sophistication and complexity
- Cons: Difficult to implement (due to specificity) and thus often leads to hand-crafted models

Liu, et al. 2003

- Textual affect sensing using large-scale real-world knowledge about the inherent affective nature of everyday situations to classify sentences into “basic” emotion categories.
- Generated 4 “Affect models”:
  1) Subject-Verb-Object-Object Model (with emotion values)
  2) Concept-Level Unigram Model (concepts = verbs, noun phrases, standalone adjective phrases)
  3) Concept-Level Valence Model
  4) Modifier Unigram Model

Liu, et al. 2003

- Linguistic models:
  - <u>Subject-Verb-Object-Object Model</u>
    - Example: "Getting into a car accident can be scary" = [<Subject>: ep_person_class*, <Verb>: get_into, <Object1>: car accident, <Object2>:] whose value is: [0 happy, 0 sad, 0 anger, 1.0 fear, 0 disgust, 0 surprise]
    - For declarative sentences in SVOO frame
    - Most specific model preserving accuracy of affective knowledge, but rather specific
  - <u>Modifier Unigram Model:</u>
    - » "Moldy bread is disgusting," "Fresh bread is delicious."
      - *Moldy* vs. *Fresh* bread
    - Sometimes modifiers are wholly responsible for carrying emotion of a verb or noun phrase

Liu, et al. 2003

- Successful? User testing suggests textual affect sensing engine works well enough to bring benefit to affective user interface application ☺

Handcrafted Models

- Thorough approaches that require deep understanding and analysis of text
- Example: Neviarouskaya et al.'s Affect Analysis Model
- Pros: Accuracy and sophistication
- Cons: generalizability limited because symbolic modeling of scripts, plans, goals, and plot units must be hand-crafted

Liu, et al. 2003 is NOT handcrafted

- Why not a “handcrafted model”?
  - Use of “Commonsense database” - the Open Mind CommonSense Corpus
  - More flexible than hand-crafted models because the knowledge source, OMCS, is a large, multi-purpose, collaboratively built resource. Compared to a hand-crafted model, OMCS is more unbiased, domain-independent, easier to produce and extend, and easier to use for multiple purposes.

Neviarouskaya et al.

- Highlights: SentiFul, Affect Analysis Model, and applications by the Compositionality Principle

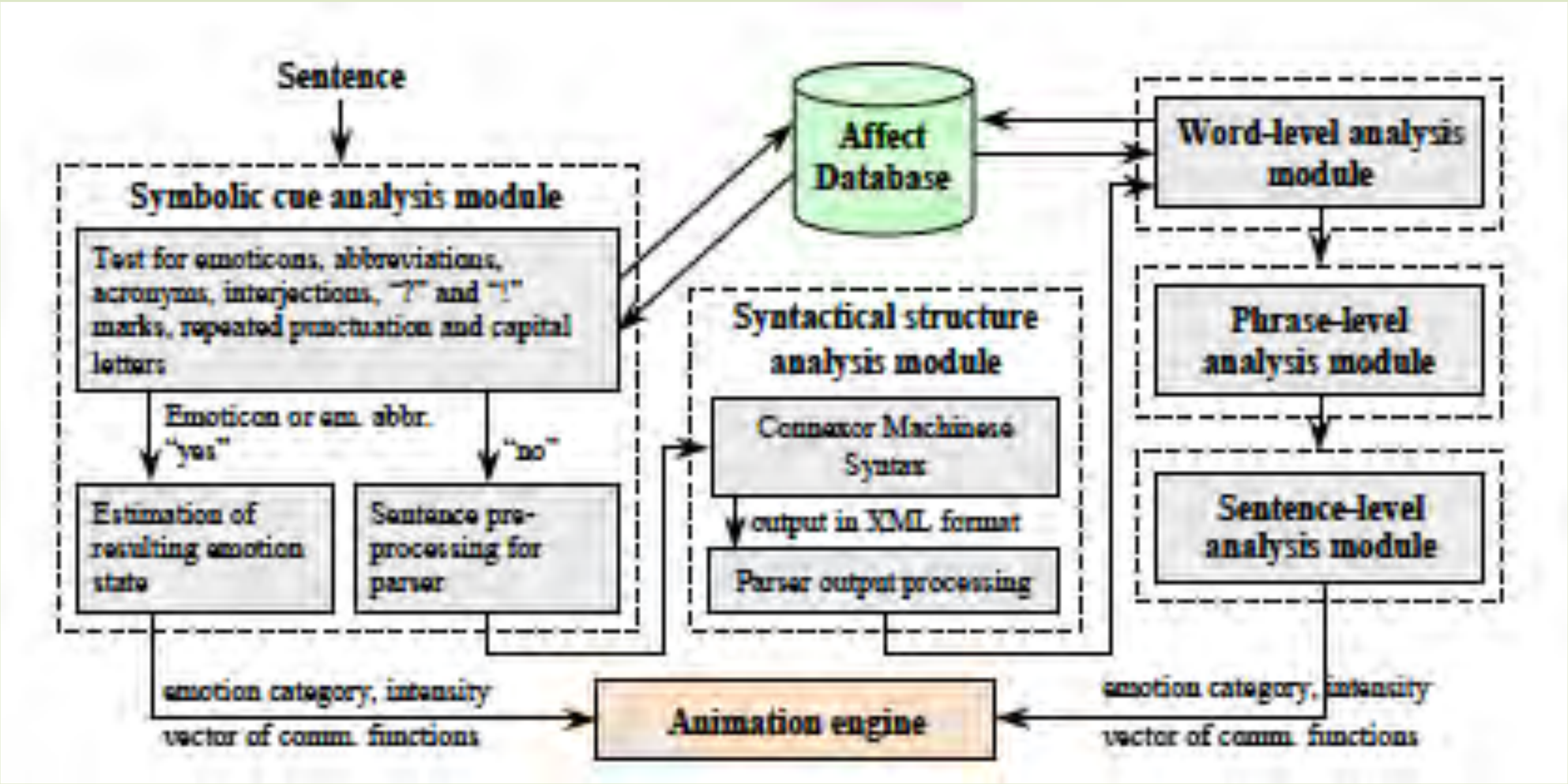
Neviarouskaya et al.

- SentiFul: Lexicon for attitude analysis includes: 1) attitude-conveying terms; modifiers; "functional" words; modal operators.
- Affective features of each emotion-related word are encoded using 9 emotion labels and corresponding emotion intensities that range from 0.0 to 1.0
  - expanded through direct synonymy and antonymy relations, hyponymy relations, derivation, and compounding with known lexical units
  - distinguish four types of affixes (used to derive new words)
  - elaborated the algorithm for automatic extraction of new sentiment-related compounds from WordNet using words from SentiFul as seeds for sentiment-carrying base components and applying the patterns of compound formations.

Neviarouskaya et al. Continued

- The Affect Analysis Model (AAM): a rule-based approach to affect sensing from text at a sentence level.
  - Algorithm for analysis of affect consists of 5 stages:
    - 1) symbolic cue analysis
    - 2) syntactical structure analysis using Connexor Machinese Syntax parser
    - 3) word-level analysis
    - 4) phrase-level analysis
    - 5) sentence-level analysis

Neviarouskaya et al. Continued



Neviarouskaya et al. Continued

- 2010: Compositionality Principle determines attitudinal meaning of a sentence by composing pieces that correspond to lexical units or other linguistic constituent types governed by rules of polarity reversal, aggregation (fusion) propagation, domination, neutralization, intensification at various grammatical levels

Neviarouskaya et al. Continued

$$\overbrace{\text{Paparazzi, }\overbrace{\text{who}}^{SF^{dep1}}\ \overbrace{\text{got}}^{VF^{dep1}}\ \overbrace{\text{best photo award}}^{OF^{dep1}}\ \overbrace{\text{last year,}}^{VF^{dep1}}}^{SF^{main}}\ \overbrace{\text{had attacked}}^{VF^{main}}\ \overbrace{\text{famous actress, }\overbrace{\text{who}}^{SF^{dep2}}\ \overbrace{\text{was enjoying}}^{VF^{dep2}}\ \overbrace{\text{her life}}^{OF^{dep2}}\ \overbrace{\text{despite troubles of upcoming divorce.}}^{VF^{dep2}}}^{OF^{main}}$$

Steps in affect recognition:

1) e^{dep1}(‘*who got best photo award last year*’) = coeff(tense:‘past’; FPP:‘no’) * e^{dep1}(SF^{0dep1} & VF^{0dep1} & OF^{+dep1}) = 0.4 * [0,0,0,0,0,0.42,0,0,0] = [0,0,0,0,0,0.17,0,0,0] = e^{+dep1};
SF^{main} = ‘*Paparazzi*’ & e^{+dep1} = [0,0,0,0,0,0.17,0,0,0] = SF^{+main};

2) e^{dep2}(‘*who was enjoying her life despite troubles of upcoming divorce*’) = coeff(tense:‘past’; FPP:‘no’) * e^{dep2}(SF^{0dep2} & VF^{+dep2}, where ‘*troubles of upcoming divorce*’ are neutralized due to ‘*despite*’, & OF^{0dep2}) = 0.4 * [0,0,0,0,0,0.6,0,0,0] = [0,0,0,0,0,0.24,0,0,0] = e^{+dep2};
OF^{main} = ‘*famous actress*’ & e^{+dep2} = [0,0,0,0,0,0.3,0,0,0.2] & [0,0,0,0,0,0.24,0,0,0] = [0,0,0,0,0,0.3,0,0,0.2] = OF^{+main};

3) e^{main}(‘*Paparazzi, who got best photo award last year, had attacked famous actress, who was enjoying her life despite troubles of upcoming divorce*’) = coeff(tense:‘past’; FPP:‘no’) * e^{main}(SF^{+main} & VF^{-main} & OF^{+main}) = 0.4 * ([0,0,0,0,0,0.17,0,0,0] & [0.4,0,0.9,0,0,0,0.8,0,0] & [0,0,0,0,0,0.3,0,0,0.2] yield [0.4,0,0.9,0,0,0,0.8,0,0]) = [0.16,0,0.36,0,0,0,0.32,0,0] = e^{-main}.

Figure 1. Example of affect sensing in a complex sentence with relative clauses

Neviarouskaya, et al. Continued

- Success?
  - Promising results, but main limitations are: strong dependency on source of lexicon, Affect database, and the commercially available syntactic parser; no disambiguation of word senses; and disregard of contextual information ☺

A Visual Overview

Author(s)	Date	Keyword-Level	Statistical NLP	Linguistic	Handcrafted	☺ or ☹
Danescue-Niculescu-Mizil et al.	2013	✔	✔	✔		☺
Keshtkar et al.	2009	✔	✔			☺
Liu et al.	2003	✔	✔	✔		☺
Mairesse et al.	2007	✔	✔			😐
Mishne, Gilad	2005		✔			☹
Neviarouskaya et al.	2007		✔	✔	✔	☺
Neviarouskaya et al.	2009		✔	✔	✔	☺
Neviarouskaya et al.	2010		✔	✔	✔	☺
Neviarouskaya et al.	2011	✔	✔	✔	✔	☺
Nigam et al.	2004	✔	✔			😐
Saif, Mohammed	2012	✔	✔			☺
Soumaya et al.	2011	✔	✔			☺
Strapparava et al.	2008	✔	✔			☺

Who's doing the best?

- We have seen that successful studies have utilized:
  - A form of affective lexicon
    - WordNet Affect to OCMS, use of affect-annotated lexicon
  - Of course, the more specific the better (Handcrafted yields high accuracy), but sacrifice generalizability
  - Hierarchical approach offers an alternative
    - Whether that is in moods (Keshtkar 2009) or in models (Liu 2003)
    - Breakdown into steps of generalizability
  - SVM is a popular and successful method, trained on different classifiers.

Citations

Chaffar, Soumaya, and Diana Inkpen. "Using a Heterogeneous Dataset for Emotion Analysis in Text." School of Information Technology and Engineering, University of Ottawa Ottawa, ON, Canada. University of Ottawa, 2011. Web. 7 Oct 2012.

Davis, D. N. (2010). Cognitive architectures for affect and motivation. *Cognitive Computation*, *2*(3), 199-216.

H. Liu, H. Lieberman, and H. Selker, "A model of textual affect sensing using real-world knowledge," IUI, Proceedings of the 8th international conference on Intelligent user interfaces, pp. 125–132, 2003.

Keshtkar, F., & Inkpen, D. (2009, September). Using sentiment orientation features for mood classification in blogs. In *Natural Language Processing and Knowledge Engineering, 2009. NLP-KE 2009. International Conference on* (pp. 1-6). IEEE.

Mairesse, Francois, et al. "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text." *Journal of Artificial Intelligence Research*, 2007.

Mishne, G. (2005, August). Experiments with mood classification in blog posts. In *Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access* (Vol. 19)

Citations

Mohammed, Saif. "Portable Features for Classifying Emotional Text." 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 587–591, Montreal, Canada, June 3-8, 2012.

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2007). Textual affect sensing for sociable and expressive online communication. In *Affective Computing and Intelligent Interaction* (pp. 218-229). Springer Berlin Heidelberg.

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2009, May). Compositionality Principle in Recognition of Fine-Grained Emotions from Text. In *ICWSM*.

Neviarouskaya, Alena, et al. "Recognition of Affect, Judgment, and Appreciation in Text." Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pages 806–814,
Beijing, August 2010

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2011). SentiFul: A lexicon for sentiment analysis. *Affective Computing, IEEE Transactions on*, *2*(1), 22-36.

Nigam, K., & Hurst, M. (2004, July). Towards a robust metric of opinion. In*AAAI spring symposium on exploring attitude and affect in text* (pp. 598-603).

Oatley K. Best laid schemes. Cambridge: Cambridge University Press; 1992.

Strapparava, C., & Mihalcea, R. (2008, March). Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing* (pp. 1556-1560). ACM.

Extra Studies

Nigam, et al. 2004

- Automatic opinion analysis and topic mining
  - An example with practical applications in market research
- Use general-purpose polar language module and a topic classifier (variant of Winnow classifier) trained with machine learning techniques hand-labeled with binary relation of topicality) to identify individual polar expressions about a topic of interest
  - Aggregate individual expressions into a single score
- Successful? Yes for identifying positive and negative topical language, but automated recall and polarity recognition significantly lower 😐

Chaffar et al. 2011

- Adopted a supervised machine learning approach to recognize six basic emotions (anger, disgust, fear, happiness, sadness and surprise) using a heterogeneous emotion-annotated dataset which combines news headlines, fairy tales and blogs.
- The Support Vector Machines classifier (SVM) trained on SMO implementation performed significantly better than other classifiers, and it generalized well on unseen examples. ☺