Toward a Computational Approach for Natural Language Description of Emotions

Abe Kazemzadeh

University of Southern California

Abstract. This is a précis of the author's dissertation proposal about natural language description of emotions. The proposal seeks to explain how humans describe emotions using natural language. The focus of the proposal is on words and phrases that refer to emotions, rather than the more general phenomena of emotional language. The main problem is that if descriptions of emotions refer to abstract concepts that are local to a particular human (or agent), then how do these concepts vary from person to person and how can shared meaning be established between people. The thesis of the proposal is that natural language emotion descriptions refer to theoretical objects, which provide a logical framework for dealing with this phenomenon in scientific experiments and engineering solutions. An experiment, *Emotion Twenty Questions* (EMO20Q), was devised to study the social natural language behavior of humans, who must use descriptions of emotions to play the familiar game of twenty questions when the unknown word is an emotion. The idea of a theory based on natural language propositions is developed and used to formalize the knowledge of a sign-using organism. Based on this pilot data, it was seen that approximately 25% of the emotion descriptions referred to emotions as objects with dimensional attributes. This motivated the author to use interval type-2 fuzzy sets as a computational model for the meaning of this dimensional subset of emotion descriptions. This model introduces a definition of a variable that ranges over emotions and allows for both inter- and intra- subject variability. A second set of experiments used interval surveys and translation tasks to assess this model. Finally, the use of spectral graph theory is proposed to represent emotional knowledge that has been acquired from the EMO20Q game.

1 Background and Motivation

The dissertation proposal begins by claiming that natural language communication and emotional empathy are two distinguishing characteristics of human beings and that these together make it possible for emotional information to be transmitted to and understood by people who were not necessarily first-hand observers of the original emotional information being communicated. It is claimed that modeling this phenomenon will be the basis of both scientific insights and engineering solutions. In particular, description of abstract phenomena like emotions can carry over to other types of behavior description, which currently rely on categorical labels or Likert scales as a description methodology. These labels and scales explicitly refer to the data they describe, but the author claims that

S. D´Mello et al. (Eds.): ACII 2011, Part II, LNCS 6975, pp. 216–223, 2011.

they implicitly refer to theoretical objects that are at a conceptual level once removed from observational data. The notion of a definite description and a theoretical object are attributed to Russell and Carnap, respectively [1,2]. Regarding natural language descriptions of emotions as definite descriptions and theoretical entities allows the author to propose the formulation of an algebraic framework for dealing with these. The conceptual meaning of emotion words are interpreted as a region of an emotional space, and natural language descriptions of emotions further refine this space.

The author gives two engineering applications that such a model could be applied to. First, it would enable computers to understand natural language descriptions of emotions and use these to describe emotional data to a users. Currently there is much research being done on emotional user interfaces, but these often aim to recognize the user's emotional state. However, there are times when a user may want to convey emotional information that is not his or her current state, such as in an after-the-fact emotional self-report, or for social purposes of telling a story or gossip.

Another application is new-media technologies that expand the notion of *presence*, which allows a user to broadcast their status to others. Current approaches to processing emotional language are derived from processing longer documents, while recent trends in electronic communication tend toward very short documents. Also, current approaches rely on models of emotion that are limited to a specific theory or set of emotion labels, while current technologies tend toward allowing user to use any emotional description they choose. The proposed approach addresses these issues by focusing on shorter word or utterance level descriptions and explicitly aiming to understand a potentially unlimited variety of emotion descriptions.

The author makes the distinction between what he calls scientific descriptions and natural language descriptions of emotions. Scientific descriptions seek to precisely define emotions for people who study them. Natural language descriptions are how ordinary people describe emotions. Scientific descriptions can be seen as a linguistic division of labor for those whose study emotions as an occupation or avocation, but this division of labor is problematic because its terminology often overlaps with the common, natural language terms. Mindfulness of this distinction is important when the object of study is natural language and naturalistic emotions that do not fit into precise categories.

2 Thesis and Approach

The author's proposed thesis is that while natural language descriptions of emotions *can* be referentially grounded in specific human behavior and situations in the world, *in general* there is an intermediate conceptual representation, in lieu of the physical reference, that is referred to in cases when emotional information is communicated among those who are not first-hand witnesses of the observed emotional behavior and situation. Because this conceptual representation is not directly tied to physical observations, and because it cannot be directly shared with other agents, this representation is vague and varies from person to person.

However, this vagueness and variability does not hinder the ability of people to communicate and reason about such data. It is hypothesized that such a conceptual representation contains information that helps humans reason about emotions abstractly by the use of logical relations between emotions, such as similarity and subsethood, and by association of certain sets of behaviors and situations with particular emotion concepts. Although all the attributes of these concepts are not shared among each agent, it is possible for agents to arrive at shared meaning through natural language communication.

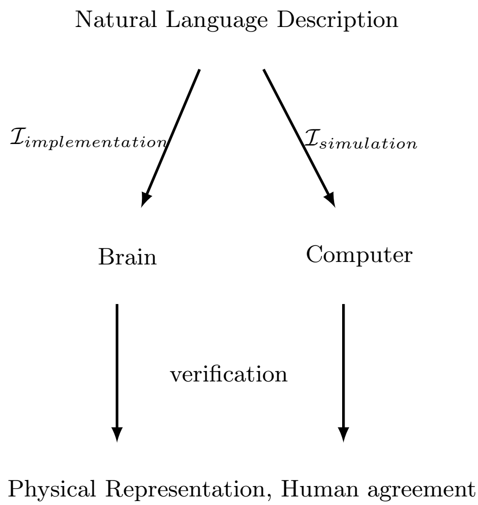

Fig. 1. Interpretation of descriptions in a model: implementation vs. simulation

The author proposes an approach which considers two types of models for interpretation ($\mathcal{I}$) of natural language descriptions of emotions, implementations and simulations, which correspond to answering scientific and engineering questions, respectively, as seen in Fig. 1. It is hypothesized that human understanding of natural language descriptions of emotions that refer to conceptual scales [5,6] are *implemented* by structures of the brain that are specialized in spacialization and wayfinding, such as the hippocampus, and subject to the effects of lateralization that have been observed for emotional words in patients with aphasia [3,4]. The conceptual representation of emotions can be computationally *simulated* using abstract scales in a conceptual space using interval type-2 fuzzy sets which capture inter- and intra-subject uncertainty and a computational implementation called *perceptual computing* [7].

To computationally model emotional concepts, the author uses an algebraic representation where emotions are fuzzy sets in an emotional space. An *emotional variable* ε represents an arbitrary region in this emotional space, i.e. $\varepsilon \subset E$, with the subset symbol $\subset$ used instead of set membership ($\in$) to represent regions in this emotion space in addition to single points. The conceptual meaning of

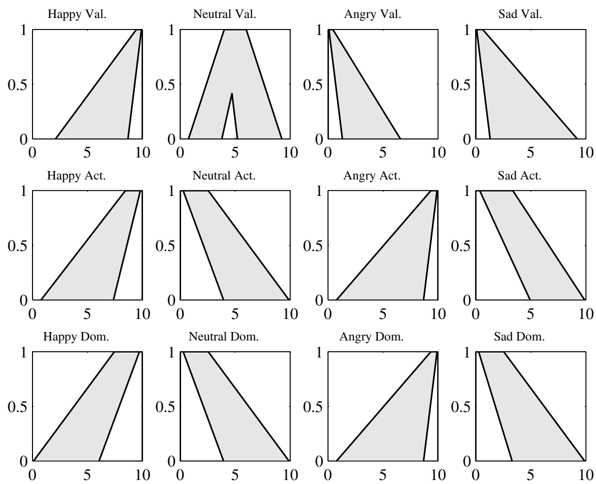

Fig. 2. Example of emotional variables as interval type-2 fuzzy sets that range over valence, activation, and dominance

an emotion word can be represented an by emotional variable that is associated with that word. An *emotion codebook* $V = (W_V, eval_V)$ is a set of words W_V and a function $eval_V$ that maps words of W_V to their corresponding region in the emotional space, $eval_V : W_V \to E$. Thus, an emotional vocabulary can be seen as a dictionary for looking up the meaning of an emotion word, which is represented as an interval type-2 fuzzy set to represent that these sets have fuzzy boundaries (e.g. Fig. 2). Similarity and subsethood are functions $E \times E \to [0, 1]$, where similarity is a measure of the similarity of two emotions and subsethood is a fuzzy logic interpretation of subsethood that allows for partial subsethood values.

The author uses mathematical logic to relate the model of emotion concepts with natural language descriptions. The definition of *theory* from mathematical logic states that a theory Γ is simply a set of sentences in some language $\mathcal{L}$ that is true of a model M [8,9]. In the case of the author's formulation, Γ is the set of natural language statements that are true with respect to particular emotions and negations of the statements that are false for a given emotion, $\mathcal{L}$ is the language of propositional logic, and M is a model of emotion concepts, which is is only accessible through observation of communicative behavior. In this view, the theory of a set of emotions can be seen as an matrix Γ indexed by the emotions in one dimension and the natural language statements in the other dimension. If the theory Γ refers to emotions ε_m for $1 \leq m \leq |E|$ and propositions p_n for $1 \leq n \leq |P|$, then Γ will be an $|E| \times |P|$ matrix.

Ordinarily, Boolean algebra would dictate that this matrix would consist of ones and zeros. Such a representation of general concepts has been explored under the aegis of *formal concept analysis* [10]. However, as the matrix is sparse one must represent the fact that not all of the combinations of questions and emotions have been encountered and that there may be some contradiction among subjects. To this end, we propose that the matrix be a $(1, 0, -1)$-matrix, i.e., a *signed matrix/graph* [11], where 1 indicates that the proposition of column-m is true for the emotion of row-n, -1 indicates that it is false, and 0 indicates that it has not been seen or that a contradiction has been encountered. This representation can be interpreted as an adjacency matrix by which the connectivity of various propositions with various emotions can be assessed using spectral graph theory.

3 Methodology and Experimental Results

The proposed thesis is supported by two main experiments. The first is a game called Emotion Twenty Questions (EMO20Q), which is used to construct a theory of emotions (where a theory is defined as above). The second experiment involved surveys that collected user responses to emotion words on interval-valued slider scales, which were used to construct a fuzzy logic model of emotion words that could account for both inter- and intra-subject variability.

The EMO20Q game is played like the ordinary game of twenty questions, but the object to be guessed in 20 or fewer turns is limited to emotion words. Furthermore, the game was played on an XMPP chat server so that games could be recorded and paralinguistic effects minimized. The questions asked are used to generate propositions about emotions and the answers to the questions determine whether the propositions are true or false of the emotion word. The fact that the emotion word is guessed by the other player demonstrates how shared meaning is established even though both players may have different concepts and different ways of describing emotions. Also, as an experimental methodology, it benefits from not limiting the emotions the players can pick nor the questions they ask. Therefore, it is not biased toward any particular scientific theory of emotions.

Pilot studies of EMO20Q were carried out with 12 players who played 26 matches. All except three matches were successfully completed, and they had a mean and median length of 12 and 15.5 turns, respectively, when unsuccessful games were considered to be 20 turns. This experiment yielded descriptive statistics about how the players used natural language to ask about emotions, which can be seen in Table 1.

When taken as simple text strings, the questions used in the game form a very sparse space of propositions. When the underlying meaning of the questions are represented as a logical formula (by manual annotation), the space becomes less sparse because superficial differences in the same logical proposition are ignored. This annotation process was termed *standardization*.

Although standardization reduced the sparsity of the data, it was still an open question whether the reduction of sparsity was sufficient to provide a coherent knowledge base. The criteria used to determine whether the knowledge was coherent was connectivity of the adjacency graph of the signed-matrix described above.

Table 1. Examples of question categories

Question Categories	Examples
identity (42%)	*is it angry?*
	guilt?
attribute (13%)	*is it something one feels for long periods of time?*
	is it a strong emotion?
similarity/ subsethood (10%)	*is the emotion a type of or related to content or zen contentment (is that a word?_)*
	so it's similar to excited?
situational (14%)	*is the emotion more likely to occur when you are tired?*
	would i feel this if my dog died?
behavior (3%)	*you can express it in an obvious way by sighing?*
	do adults usually try to discourage children from feeling this?
causal (7%)	*yes. mynext question is can it harm anyone besides the feeler?*
	I think I know, but I'll ask one more question...does it ever cause children to wake up and cry?
social (8%)	*are you less likely to experience the emotion when around good firiends?*
	13)would you feel that towards someone who is superior to you?
miscelaneous (3%)	*i dont' know if this is a valid question, but does it start with the letter D?*
	or an aspirational emotion?
	does the word function or can be conjugated as anything eles? i.e. can it be a verb too?

The connectivity was determined through two spectral graph methods: finding the eigenvalues of the graph Laplacian of the absolute-value of the adjacency graph and computing the power series of this graph. To make the matrix Γ above into square, symmetric adjacency matrix, we define the adjacency matrix of Γ, $A = A(\Gamma)$ to be an $M + N \times M + N$ matrix as follows:

$$A(\Gamma) = \begin{bmatrix} zeros(M) & \Gamma^T \\ \Gamma & zeros(N) \end{bmatrix}$$

The absolute value $|A|$ of A describes whether questions have been asked of objects, regardless of whether the answer was yes or no. It is this graph $|A|$ that is used as and indicator about the connectivity of an agent's knowledge. The Laplacian L of a signed graph is calculated by subtracting the absolute adjacency matrix $|A|$ from the diagonal absolute degree matrix $\bar{D}_{ii} = \sum_j |A_{ij}|$, $L = \bar{D} - |A|$. From the matrix L one can tell the number of connected components of A by counting the number of zero eigenvalues. Thus, if there are three eigenvalues that equal zero, the graph is composed of three separate connected components. A graph Laplacian with one zero eigenvalue is a single connected graph. To determine the exact emotion-question pairs that are disconnected, the power series of A is used to count the number of walks between the two nodes. The value of $(A^l)_{ij}$ determines how many walks of length l exist between nodes i and j. If l is greater than the total number of nodes (the combined number of emotions and questions) and the value of $(A^l)_{ij}$ is zero, then there is no path between the nodes. This indicates for an automated agent playing EMO20Q a question that can be asked to improve its knowledge. From this spectral graph analysis, it was determined that overall, the graph was a single connected component except for questions that did not receive a clear yes or no answer.

The second experiment presented in the proposal was a web survey to present subjects with an emotion word and collect their responses in terms of valence, activation, and dominance values. The first experiment showed that, together, dimensional descriptions of emotions and similarity/subsethood descriptions accounted for nearly a fourth of the observed data. Valence, activation, and dominance are also one of the common ways of representing emotions. Similarity and subsethood judgments were not explicitly queried, but rather they were derived from the fuzzy membership functions of the valence, activation, and dominance values, as seen in Fig. 2. The novel aspect of this methodology, known as the *interval approach* and first described in [12], is that instead of giving a response as a value on Likert scale, subjects respond with an interval on the scale, which allows them to indicate their uncertainty so that intra-subject uncertainty to be measured. The interval approach then aggregates the individual intervals from each subject to make a membership function, like those in Fig. 2, for each emotion.

Several sub-experiments were carried out, all with the same methodology, but different emotion vocabularies. One vocabulary was a set of seven emotion words that are commonly used as labels for emotional corpora and the second vocabulary was derived from the first by adding “very” and “sort of” to the first. Another vocabulary was a list of 40 colloquial emotion words from a blogging site that allowed users to indicate their moods [13]. Finally, the fourth vocabulary was a Spanish vocabulary of 30 words that was taken from a mental health promotion.

The accuracy of the resulting fuzzy sets was assessed using a translation task that involved mapping from one vocabulary to another using similarity and subsethood measures. The translation accuracy was thus a proxy for the accuracy of the fuzzy sets themselves. Translation between emotion vocabularies is a need that can arise when different researchers use different categorizations to label emotional data. The accuracy was determined by comparing the model’s translations with human translations. The author found that mapping from one vocabulary to another resulted in performance of up to 86% when a small output vocabulary was used (chance accuracy of 1/7) and approaching 50% when a large output vocabulary was used (chance accuracy of 1/40).

4 Conclusion

The author based his thesis on the following papers, both published and under review. In [14,15,16] attempts at user modeling of human-computer dialogs was undertaken. In particular, [16] examined a scale of user activation level that was obtained by a machine learning technique known as model trees. [17] explored the natural language expressions of common sense psychology concepts, such as concepts like "planning", "execution", "causation", "belief", etc., as they were expressed in text corpora. In [18,19] the first interval type-2 fuzzy logic model of emotions were created and the formal description of this model will be found in [20]. The EMO20Q game was proposed in [21] and the spectral graph analysis is introduced in [22].

References

1. Russell, B.: On denoting. Mind 14, 479–493 (1905)
2. Carnap, R.: The methodological character of theoretical concepts. Minnesota Studies in the Philosophy of Science I, 39–76 (1956)

3. Richardson, J.T.E.: The effect of word imageability in acquired dyslexia. Neuropsychologia 13, 281–288 (1975)
4. Landis, T.: Emotional words: What's so different from just words. Cortex 42, 823–830 (2006)
5. Lakoff, G., Johnson, M.: Metaphors We Live By. University of Chicago Press, Chicago (1980)
6. Gardenfors, P.: Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge (2000)
7. Mendel, J.M., Wu, D.: Perceptual Computing: Aiding People in Making Subjective Judgements. IEEE Press and John Wiley and Sons, Inc. (2010)
8. Enderton, H.B.: A Mathematical Introduction to Logic, 2nd edn. Academic Press, London (2001)
9. Forster, T.: Logic, Induction, and Sets. Cambridge University Press, Cambridge (2003)
10. Ganter, B., Wille, G.S.R. (eds.): Formal Concept Analysis: foundation and applications. Springer, Berlin (2005)
11. Kunegis, J., Lommatzsch, A., Bauckhage, C.: The slashdot zoo: Mining a social network with negative costs. In: World Wide Web Conference (WWW 2009), Madrid, pp. 741–750 (April 2009)
12. Liu, F., Mendel, J.M.: An interval approach to fuzzistics for interval type-2 fuzzy sets. In: Proceedings of Fuzzy Systems Conference, FUZZ-IEEE (2007)
13. Mishne, G.: Applied Text Analytics for Blogs. PhD thesis, University of Amsterdam (2007)
14. Shin, J., Narayanan, S., Gerber, L., Kazemzadeh, A., Byrd, D.: Analysis of user behavior under error conditions in spoken dialogues. In: ICSLP, Denver (2002)
15. Kazemzadeh, A., Lee, S., Narayanan, S.: Acoustic correlates of user response to errors in human-computer dialogues. In: ASRU, St. Thomas, U.S. Virgin, Islands (2003)
16. Kazemzadeh, A., Lee, S., Narayanan, S.: Using model trees for evaluating dialog error conditions based on acoustic information. In: Proceedings of the 1st ACM International Workshop on Human-Centered Multimedia, Santa Barbara, California, USA, pp. 109–114 (2006)
17. Gordon, A., Kazemzadeh, A., Nair, A., Petrova, M.: Recognizing expressions of commonsense psychology in english text. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, ACL 2003 (2003)
18. Kazemzadeh, A., Lee, S., Narayanan, S.: An interval type-2 fuzzy logic system to translate between emotion-related vocabularies. In: Proceedings of Interspeech, Brisbane, Australia (September 2008)
19. Kazemzadeh, A.: Using interval type-2 fuzzy logic to translate emotion words from spanish to english. In: IEEE World Conference on Computational Intelligence (WCCI) FUZZ-IEEE Workshop (2010)
20. Kazemzadeh, A., Lee, S., Narayanan, S.: An interval type-2 fuzzy logic model for the meaning of words in an emotional vocabulary (2011) (under review)
21. Kazemzadeh, A., Georgiou, P.G., Lee, S., Narayanan, S.: Emotion twenty questions: Toward a crowd-sourced theory of emotions. In: D´Mello, S., et al. (eds.) Proceedings of ACII 2011, Part II, vol. 6975, pp. 1–10. Springer, Heidelberg (2011)
22. Kazemzadeh, A., Lee, S., Georgiou, P.G., Narayanan, S.: Determining what questions to ask, with the help of spectral graph theory. In: Proceedings of Interspeech (2011)