Fuzzy Logic Models for the Meaning of Emotion Words

Abe Kazemzadeh, Sungbok Lee, and Shrikanth Narayanan
University of Southern California, USA

I. Introduction

Words and natural language play a central role in how we describe and understand emotions. One can learn about emotions first-hand by observing physiological or behavioral data, but to communicate emotional information to others who are not first-hand observers, one must use natural language descriptions to communicate the emotional information. The field of affective computing deals with creating computer systems that can recognize and understand human emotions. To realize the goals of affective computing, it is necessary not only to recognize and model emotional behavior, but also to understand the language that is used to describe such emotional behavior. For example, a computer system that recognizes a user's emotion from speech should not only recognize the user's emotion from expressive speech acoustics, but also understand when the user says "I am beginning to feel X," where "X" is a variable representing some emotion word or description. The ability to understand descriptions of emotions is important not only for human-computer interaction, but also in deliberative decision making activities where deriving behavioral analytics is based on natural language (for example, in mental health assessments). Such analytics often rely on abstract scales that are defined in terms of natural language.

This paper looks at the problem of creating a computational model for the conceptual meaning of words used to name and describe emotions. To do this, we represent the meaning of emotion words as interval type-2 fuzzy sets (IT2 FSs) that constrain an abstract emotion space. We present two models that represent different views of what this emotion space might be like. The first model consists of the Cartesian product of the abstract scales of valence, activation, and dominance. These scales have been postulated to represent the conceptual meaning of emotion words [1]. The second model is based on scales

Digital Object Identifier 10.1109/MCI.2013.2247824
Date of publication: 11 April 2013

Abstract—This paper presents two models that use interval type-2 fuzzy sets (IT2 FSs) for representing the meaning of words that refer to emotions. In the first model, the meaning of an emotion word is represented by IT2 FSs on valence, activation, and dominance scales. In the second model, the meaning of an emotion word is represented by answers to an open-ended set of questions from the game of Emotion Twenty Questions (EMO20Q). The notion of meaning in the two proposed models is made explicit using the Fregean framework of extensional and intensional components of meaning. Inter- and intra-subject uncertainty is captured by using IT2 FSs learned from interval approach surveys. Similarity and subsethood operators are used for comparing the meaning of pairs of words. For the first model, we apply similarity and subsethood operators for the task of translating one emotional vocabulary, represented as a computing with words (CWW) codebook, to another. This act of translation is shown to be an example of CWW that is extended to use the three scales of valence, activation, and dominance to represent a single variable. We experimentally evaluate the use of the first model for translations and mappings between vocabularies. Accuracy is high when using a small emotion vocabulary as an output, but performance decreases when the output vocabulary is larger. The second model was devised to deal with larger emotion vocabularies, but presents interesting technical challenges in that the set of scales underlying two different emotion words may not be the same. We evaluate the second model by comparing it with results from a single-slider survey. We discuss the theoretical insights that the two models allow and the advantages and disadvantages of each.

derived from answers to yes/no questions, where each scale can be seen as the truth value of a proposition. In each model, the meaning of an emotion word is represented as a fuzzy set in an emotion space, but the two models represent different theoretical organizations of emotion concepts. In the first, a spatial metaphor is used to organize emotion concepts on valence, activation, and dominance scales. In the second model, emotion concepts are represented as lists of propositions and associated truth values.

In both models, the algebraic properties of fuzzy sets can be used as a computational model for the meaning of an emotion word. We outline the properties of these models and describe the methodology that estimates the fuzzy set shape parameters from data collected in *interval approach* surveys [2], [3]. In an interval approach survey, subjects rate words on abstract scales, but instead of picking a single value on the scales (as in a Likert scale survey), users select interval ranges on these scales. In the two models we present, the survey results are aggregated into fuzzy sets for words in an emotion vocabulary. The fuzzy set representation allows one to compute logical relations among these emotion words. By using the relations of similarity and subsethood as measures of mappings between items of two vocabularies, one can translate between these vocabularies. This allows us to use our first model for several applications that involve mapping between vocabularies of emotion words: converting emotion labels from one codebook to another, both when the codebooks are in the same language (for example, when using different emotion annotation schemes) and when they are in different languages, such as when translating emotion words from one language to another (here, Spanish and English). These applications show one way our proposed model may be used and provide experimental evidence by which we can evaluate the model. For evaluation of the first model, we compare the translation applications with human performance in these tasks as a benchmark.

> **Focusing on the conceptual meaning of emotion words allows us to consider cases where emotion is communicated through linguistic meaning, as opposed to paralinguistics or body language.**

Our results show that performance of the first model decreases when the vocabulary size gets larger, which indicates that a three-scale representation for emotions is ideal only for small vocabularies. To address this limitation, our second model uses inspiration from the game of twenty questions, where players can identify a large set of objects using question-asking. Because people's beliefs about emotions can be subjective, many of the answers to questions about emotions are vague and can be represented as fuzzy sets. For evaluation of this model, we test the estimated IT2 FS on data from different subjects who took a single-value survey by finding the membership of these points in the estimated IT2 FS.

Other research has presented related methodologies–using fuzzy logic for affective computing, emotion lexical resource development, and representing emotions using valence, activation, and dominance dimensions. We will commence by describing some of these works and the novelties that will be introduced by our paper.

There are many examples where fuzzy logic has been applied to the task of recognizing and representing observed emotional behavior. [4] gives an example where fuzzy logic is applied to multimodal emotion recognition. Other examples of fuzzy logic in emotion recognition are [5]–[7], which use fuzzy logic rules to map acoustic features to a dimensional representation in valence, activation, and dominance. [8] uses an IT2 FS model for emotion recognition from facial expressions. The model of [9] uses fuzzy logic for emotional behavior generation.

Another related trend of research is the development of lexical resources. Our work can be seen as a lexical resource framework like the Dictionary of Affective Language (DAL) [10]. In this work, 8745 common English words were evaluated for valence and activation (as well as a third dimension, imagery). The methodology for collecting the data in this paper was similar to our survey in presenting subjects with words as stimuli, but in the DAL the values of each word's dimensions are the mean across all subjects, so there is no estimate of the intra-subject variation. Also, compared with the DAL, we focus on words that are names of emotions, rather than words that might have emotional connotation. As such, our approach is more geared toward analyzing the meaning of short utterances explicitly referring to emotions, which we call *natural language descriptions of emotion* [11], while the dictionary of affect would be more appropriate for characterizing the emotional tone at the document-level. Another related research trend outside the domain of affective computing is the study of linguistic description of signals [12], [13], which aims to associate words with the signals they describe.

One of the contrastive traits of this research is that we try to use the dimensional approach and fuzzy logic to model emotion concepts used in natural language descriptions of emotions [11], rather than characterizing data from emotional human behavior [14], [4]–[7]. Focusing on the conceptual meaning of emotion words allows us to consider cases where emotion is communicated through linguistic meaning, as opposed to paralinguistics or body language. The dimensional approach has been used to both describe emotional data and emotion concepts but more often than not this distinction is not made clear. By describing our model of the meaning of emotion words in terminology established by the philosophy of language, we hope to clarify this issue. Furthermore, by rigorously defining the idea of an emotional variable and operations on such variables in terms of fuzzy logic, we can establish general relations such as similarity and subsethood that can be applied even if the underlying representation of valence, activation, and dominance is changed. Another contrast between this work and other research using fuzzy logic to represent emotional dimensions is that we use IT2 FSs [15] and the interval approach [3]. This allows our model to account for both inter- and intra-subject variability. Compared with the earlier developments of [16]–[18], this paper offers a more detailed description of the theoretical framework and analysis of experimental results by incorporating subsethood and applying newer developments to the interval approach [19] (Section III-D). This paper also extends these results by proposing a second model to deal with larger emotion vocabularies (Section IV-C).

By constraining our focus to a conceptual level, we focus on input/output relations whose objects are words, rather than observations of stimuli and behavior. As such, this work can be seen as an instance of Computing with Words (CWW) [20], [21], [22]. CWW is a paradigm that considers words as the input and output objects of computation. Perceptual computing [23], [24] is an implementation of the CWW paradigm that we draw upon in this work.

The rest of the paper is organized as follows. In Section II, we describe what we mean by the "meaning" of emotion words. This is an important topic on its own, but we give an introduction that we deem sufficient for the purposes of this article. In Section III, we describe the fuzzy logic framework and the proposed computational models for emotion words. In Section IV, we describe the experimental implementation and testing of the models. The results are presented in Section V. We discuss advantages and disadvantages of these models in Section VI conclude in Section VII.

II. The Meaning of Meaning

What does it mean to say that our model represents the meaning of emotion words? We believe this is an important question and therefore we will briefly discuss meaning in general in Section II-A and then explain how it relates to the meaning of emotion words in Section II-B.

A. Meaning in General

In an influential paper around the end of the 19th century, the philosopher of language Gottlob Frege described two components of meaning: *extension* and *intension* [25]. The extensional component of meaning is a mapping from words to things in the world, whereas the intensional meaning is a mapping from words to concepts. The stereotypical example of this is illustrated by the terms "morning star," "evening star," and "Venus." The extensional meaning of these three terms is the same, namely the second planet in the solar system. However, the intensional meaning of these three terms is different, which explains why the three terms cannot be freely substituted in an arbitrary sentence without changing the meaning of the sentence. In this paper, we focus on the meaning of individual words, but we touch upon the topic of the meaning of phrases in the second model.

Although the notion of extension and intension are most frequently associated with the field of philosophy of language, the idea can also be described in mathematical terms [26]. One can think of the extension of a function as a set of ordered pairs, where the first item of the pair is an input to the function and the second item in the pair is the corresponding output. The intensions of a function are described by their symbolic or algorithmic representations. Therefore we can have "$f(x) = x^2$" or "$f(x) = x * x$" as intensions of the extensional set of pairs "$\langle 1,1\rangle, \langle 2,4\rangle, \langle 3,9\rangle, \ldots$" Extension and intension have been formally described in the study of formal concept analysis [27].

We believe that by defining meaning in this way, we can describe our model more precisely. Without explicitly describing "meaning," whether in terms of extension and intension or otherwise, this important concept tends to get blurred. Although, this topic is complex, the intuition behind it is rather simple: similar, intuitive distinctions along the lines of intension and extension are common. Extension-related terms include: referent, percept, object, empirical data, Aristotelian world view, or stimulus meaning. Intension-related terms include: signified, concept, subject, knowledge structure, schema, Platonic world view, or linguistic meaning. The process of understanding a word is a mapping, or *interpretation*, from the word itself to the word's meaning, whether it be intensional or extensional. We argue that, when understanding natural language in the absence of first-hand, perceptual evidence, people refer to intensional meaning rather than extensional meaning. It is intensional meaning that we focus on in this paper.

B. The Meaning of Emotion Words

According to the definition of meaning described above, the extensional meaning of an emotion word is the set of human behaviors and states of the world that the word refers to. The intensional meaning of an emotion word is the concept that people have when using it to communicate. Although most other examples of emotion research do not make an explicit distinction between intensional and extensional meaning, it seems that many tend towards extensional meaning, especially those that deal with the analysis of emotional data that has been annotated with emotional labels. In this view, the extensional meaning of an emotion word used as an annotation label refers to the set of all data to which it has been applied. The focus on intensional meaning in this work therefore can be seen as one of its distinguishing features, though it could be said that machine learning that generalizes from training data is in fact a way to infer intentional meaning.

The question then arises about the form of this intensional meaning, in particular, how we can simulate this subjective form of meaning, with respect to emotion words, in a computer. The two computational models we describe mirror two different theoretical views of intensional meaning. One view seeks to represent the intensional meaning of emotion words as points or regions of an abstract, low-dimensional semantic space of valence, activation, and dominance. The other view seeks to represent the intensional meaning of emotion words in relation to other propositions. This latter perspective is exemplified in the Emotion Twenty Question (EMO20Q) game. EMO20Q is played like the normal twenty questions guessing game except that the objects to be guessed are emotions. One player, the answerer, picks an emotion word and the other player, the questioner, tries to guess the emotion word by asking twenty or fewer yes-no questions. Each question can be seen as a proposition about the emotion word, which prompts an answer that ranges on a scale from assent to dissent.

Scale-based models of emotion have an interesting history that goes back to Spearman's attempts to measure general intelligence using factor analysis. At first Spearman hypothesized that there was one underlying scale that could represent a person's intelligence, but later it came to be realized that intelligence was a complex concept that required multiple scales. Factor analysis was the method used to isolate these scales, and in turn factor analysis was used in the pioneering work [28] that first identified valence, activation, and dominance as factors in the connotative meanings of words. In [28], psychologists, aided by one of the early computers, conducted *semantic differential* surveys that tried to measure the meaning of words on Likert scales whose endpoints were defined by thesaurus antonyms. Valence, activation, and dominance were identified as interpretations of the factors that were encountered. Some of the early applications of this emotional model to language are [29], [1], [30], [10]. The pictorial representation of these dimensions, which we use in the interval surveys, was developed by [31]. It should be noted that the valence, dominance, and activation representation is merely a model for emotional meaning and these scales most likely do not exhaustively describe all emotional concepts. In [32] it is argued that four dimensions are required; "unpredictability" in addition to the three scales we use. The approach we advocate here is based on an algebraic model that is generalizable to any scales. Our choice of the three scales for this model was motivated by their wide usage and to balance theoretical and practical concerns.

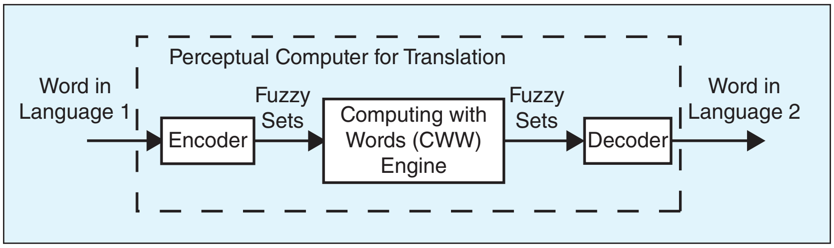

FIGURE 1 Translation as a perceptual computer.

The second model we propose takes a different perspective. Rather than having theoretically motivated scales for various characteristics of emotions, the second model aims to represent the intentional meaning of emotion words in terms of natural language propositions that can be assented to or dissented from. This view could also be construed as an abstract scale of truth with respect to various propositions (which has been considered in the study of veristic fuzzy sets [33]–[35]), but we see this view as qualitatively different from the first model. The reason why we see the propositional model as different from the scale-based model is that, first, the number of propositions about emotions will generally be larger than the number of emotion words, whereas in the case of the scale-based representation the number of scalar dimensions will be smaller than the emotion vocabulary size. Another reason that the propositional model can be considered qualitatively different than the scale-based model is that propositions can be verbally (or orthographically) expressed as linguistic stimuli, whereas abstract scales carry more cognitive implications and are language independent. Some questions from EMO20Q closely correspond to the scales in the first model, e.g., "is it positive?" is similar to valence, "is it a strong emotion?" is similar to activation, and "is it related to another person?" hints at dominance. However, model 2 contains many questions that are very specific, such as "would you feel this emotion on your birthday?".

The models we propose can be seen as an algebraic representation where theoretical entities like emotion concepts are considered virtual objects [36] with abstract scales. In this view, a collection of scales that describe an object can be seen as a *suite of congruence relations*. Recall that a congruence relation $\equiv$ (mod P) is an equivalence relation that holds given some property or function P. A suite of congruence relations is a bundle of equivalence relations $\{\sim_i : i \in I\}$, again, given some property P. In both of the models we present, P are fuzzy sets in an emotion space. In the case of the first model we present, I is a set which can contain valence, activation, and/or dominance. In the case of the second model, I is a set of propositions derived from the EMO20Q game [37]–[40]. For example, for the statement that "ε makes you smile," we can say that happy and amused are congruent given this statement about smiling behavior. In terms of the scales, the equivalence relations on each scale divide the scale space into equivalence classes. In the next section, we describe this space of emotions in more detail.

III. IT2 FS Model for the Meaning of Emotion Words

A. Emotion Space and Emotional Variables

Let E be an *emotion space*, an abstract space of possible emotions (this will be explained later in terms of valence, activation, and dominance, but for the time being we will remain agnostic about the underlying representation). An *emotion variable* ε represents an arbitrary region in this emotion space, i.e., $\varepsilon \subset E$, with the subset symbol $\subset$ used instead of set membership $(\in)$ because we wish to represent regions in this emotion space in addition to single points.

The intensional meaning of an emotion word can be represented by a region of the emotion space that is associated with that word. An *emotion codebook* $C = (W_C, \text{eval}_C)$ is a set of words W_C and a function eval_C that maps words of W_C to their corresponding region in the emotion space, $\text{eval}_C : W_C \rightarrow E$. Thus, an emotion codebook can be seen as a dictionary for looking up the meaning of words in a vocabulary. Words in an emotion codebook can also be seen as constant emotion variables. The region of the emotion space that eval_C maps words to is determined by interval surveys, as described in Section III-D.

We consider two basic relations on emotion variables: similarity and subsethood. Similarity, $\text{sm} : E \times E$, is a binary equivalence relation between two emotion variables (we will see that the fuzzy logic interpretation of similarity will actually be a function, $\text{sm} : E \times E \rightarrow [0,1]$, which measures the amount of similarity between the variables rather than being true or false). Subsethood, $\text{ss} : E \times E$, is a binary relation between two emotion variables that is true if the first variable of the relation is contained in the second. Like similarity, the fuzzy logic interpretation of subsethood is a value between zero and one. Further details are provided in Section III-C, where we will define the fuzzy logic interpretation of these relations.

Finally, a translation is a mapping from the words of one vocabulary to another, as determined by the corresponding codebooks:

$$\text{translate} : W_1 \times C_1 \times C_2 \rightarrow W_2, \tag{1}$$

which is displayed schematically in Figure 1. This can be decomposed by thinking of $C_1 \times C_2$ as a similarity or subsethood matrix, which is denoted as the CWW engine in the figure. Translation can be seen as selecting the word from the output language $w_{\text{output}} \in W_2$ such that the similarity or subsethood is maximized for a given $w_{\text{input}} \in W_1$. In the case of similarity, the translation output is

$$w_{\text{output}} = \arg \max_{w_2 \in W_2} \text{sm}(\text{eval}_{C_2}(w_2), \text{eval}_{C_1}(w_{\text{input}})), \tag{2}$$

where the *argmax* functions as the decoder in Figure 1. The formulation of similarity and subsethood in terms of IT2 FSs will be described in Section III-C and we will empirically evaluate the use of similarity and subsethood for use in translation in Section V.

B. Fuzzy Logic and Emotion Concepts

In Section III-A, the definition of an emotion space E followed a traditional set theoretic formulation. Traditional, non-fuzzy sets have crisp boundaries, which means that we can precisely determine whether a region in the emotion space is a member of any given set representing an emotion word. However, this seems to contradict the intuition and evidence that emotion concepts are somewhat vague and not precisely defined sets [41]. There are several sources of uncertainty that theoretically preclude precise set boundaries in either of the two models we present. There is modeling uncertainty because a computational model is necessarily an approximation of human thought processes. There is measurement uncertainty because the precision on these scales may be limited by perceptual processes of mapping sensory data to concepts and in distinguishing between concepts. Finally, there is uncertainty due to inter- and intra-subject variation. Postulating a blurred boundary between emotion concepts leads us to use fuzzy logic, in particular IT2 FSs.

If we deem that emotion concepts can be represented as fuzzy sets in either of these two models, then how do we determine the shapes of sets in this space? As we describe later in Section III-D, we use the interval approach survey methodology. One can think of a Likert type of survey where the scales represents valence, activation, and dominance and then query users with emotion words as stimuli; however, subjects may be unsure about picking a specific point on the scale due to vagueness in the meaning of emotion words, especially broadly defined emotion words like those typically used as primary emotions. To deal with this intra-subject uncertainty, we turn to interval surveys and IT2 FSs.

Just as type-1 fuzzy sets extend classical sets by postulating set membership grade to be a point in [0,1], type-2 fuzzy sets further extend this generalization by defining a membership function's membership grade at a given point in the domain to be a distribution in [0,1] rather than a single point, which allows for uncertainty in the membership grade [42]. The rationale for type-2 fuzzy logic is that even if a membership function takes a value between 0 and 1, there is still no uncertainty being represented because the membership value is a fixed point. What is represented by type-1 fuzzy sets is partial membership, not uncertainty. Whenever there is uncertainty, type-2 fuzzy logic is motivated on theoretical grounds [21]. The region of uncertainty in the membership grade with respect to the domain is known as the *footprint of uncertainty*.

While general type-2 fuzzy logic systems account for uncertainty, they are more conceptually and computationally complex, and methods to estimate them directly from human input are still ongoing areas of research [43]. IT2 FSs use intervals to capture uncertainty of the membership grade [15].

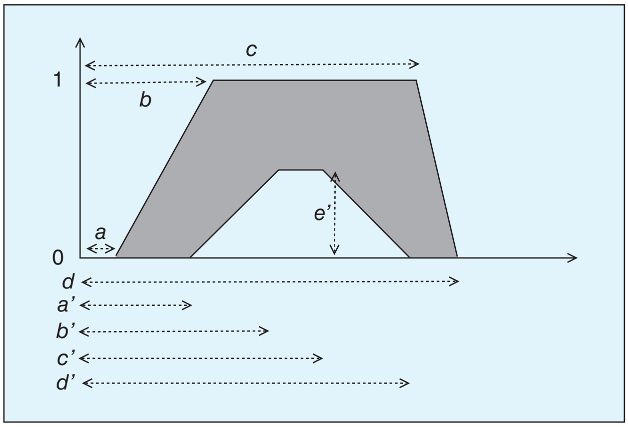

FIGURE 2 Example of a trapezoidal interval type-2 membership function (IT2 MF). A normalized trapezoidal IT2 MF can be specified with nine parameters, ($a, b, c, d, a', b', c', d', e'$). The trapezoidal height of the upper membership function (e), can be omitted in normalized IT2 FSs because it is always equal to 1.

Instead of an arbitrary distribution in [0, 1] as is the case for general type-2 fuzzy sets, IT2 FSs use an interval $[l, u]$ in [0,1] to represent an area of uniform uncertainty in the membership function's value, where $0 \leq l \leq u \leq 1$ are the lower and upper bounds of the uncertainty interval, respectively. IT2 FSs can be regarded as a first-order representation of uncertainty because they are the simplest type of fuzzy set that will account for uncertainty in the membership function. Also, as will be discussed in Section III-D, there is a method for constructing IT2 FSs from human input, which makes the use of IT2 FSs practical for human-computer interaction.

IT2 FSs have been widely used because they approximate the capability to represent the uncertainty of general type-2 fuzzy set models while still using many of the same techniques used for type-1 fuzzy sets. IT2 FSs can be represented as two type-1 membership functions: an upper membership function, which defines the upper bound of membership, and a lower membership function, which represents the lower bound on membership. When these coincide, the IT2 FS reduces to a type-1 fuzzy set [44], [45]. If the difference between the upper and lower membership function is wide, this means that we have much uncertainty about the membership grade.

An example of an interval type-2 membership function can be seen in Fig. 2. The area between the upper and lower membership functions is the footprint of uncertainty. In this paper, as an engineering decision we have restricted ourselves to trapezoidal membership functions, which can be specified in a concise way using a 5-tuple (a, b, c, d, e). The first number of the tuple, a, represents the x-value of the left side point of the base of the trapezoid, b represents the x-value of the left side point of the top of the trapezoid, c represents the x-value of the right side point of the top of the trapezoid, d represents the x-value of the right side point of the base of the trapezoid, and e represents the height of the trapezoid (i.e., the y-value of the top of the trapezoid). Since IT2 FSs consist of an upper and lower membership function, they can be represented as a 10-tuple. However, in the case of normalized interval type-2 membership functions, those whose upper membership function reaches 1, we can leave out the height of the upper membership function and specify the fuzzy set as a 9-tuple consisting of a 4-tuple for the upper membership function with the fifth value assumed to equal be 1, and a 5-tuple for the lower membership function (we must include the fifth value, e' as described above, because in general the height of the lower membership function can be anywhere between 0 and 1).

C. Similarity and Subsethood

Similarity and subsethood form important parts of our model of emotions.

The notion of similarity allows us to indicate that some pairs of emotion concepts are more or less similar. For example, we would say that angry is more similar to frustration than it is to happiness. When we make this judgment, we do not explicitly consider specific experiential examples of angry, frustrated, and happy data. Rather, we argue that one can make similarity judgments based on a mental representations of emotions. Two people could have disjoint sets of formative emotional stimuli, but still largely agree on the emotion concepts which form the intensional meaning of emotion words. In the fuzzy logic interpretation, similarity ranges from 0 to 1, where 1 is equality of two membership functions, and 0 indicates that the membership functions have no overlap.

The notion of subsethood allows us to capture that some general emotions might encompass other emotions. For example, "amused" might be a subset of "happy." The notion of subsethood is defined for traditional sets as being a Boolean value, but for fuzzy sets it takes a value between 0 and 1.

Similarity and subsethood are closely related. For clarity, we present the definitions of similarity and subsethood in terms of crisp sets, then type-1 and type-2 fuzzy sets. The definitions of the fuzzy set similarity and subsethood follow naturally from crisp sets.

The general form of similarity is based on the Jaccard Index, which states that the similarity of two sets is the cardinality of the intersection divided by the cardinality of the union, i.e.,

$$\mathrm{sm}_J(A,B) = \frac{|A \cap B|}{|A \cup B|}. \tag{3}$$

For fuzzy sets, the set operations of intersection and union ($\cup$ and $\cap$) are realized by the min and max functions and the cardinality operator ($|\ |$) is realized by summing along the domain of the variable. Thus for type-1 fuzzy sets,

$$\mathrm{sm}_J(A,B) = \frac{\sum_{i=1}^{N} \min(\mu_A(x_i), \mu_B(x_i))}{\sum_{i=1}^{N} \max(\mu_A(x_i), \mu_B(x_i))}. \tag{4}$$

For IT2 FSs, the right hand side of this equation becomes

$$\frac{\sum_{i=1}^{N} \min(\overline{\mu}_A(x_i), \overline{\mu}_B(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_A(x_i), \underline{\mu}_B(x_i))}{\sum_{i=1}^{N} \max(\overline{\mu}_A(x_i), \overline{\mu}_B(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_A(x_i), \underline{\mu}_B(x_i))}, \tag{5}$$

where $\overline{\mu}(x)$ and $\underline{\mu}(x)$ are the upper and lower membership functions, respectively. The formulas for similarity are symmetric $(\mathrm{sm}_J(A,B) = \mathrm{sm}_J(B,A))$ and reflexive $(\mathrm{sm}_J(A,A) = 1)$ [23].

We also examined a different, earlier similarity method called the Vector Similarity Method (VSM) [46]. This method was used in earlier experiments [16], so we tested it in addition to the newer Jaccard-based method. The VSM uses intuition that similarity of a fuzzy set is based on two notions: similarity of shape and similarity of proximity. Thus, the similarity of two fuzzy sets can be seen as a two element vector: $\mathrm{ss}_V(A,B) = (\mathrm{ss}_{\mathrm{shape}}(A,B), \mathrm{ss}_{\mathrm{proximity}}(A,B))^T$. The similarity measure of proximity is based on the Euclidean distance between the fuzzy set centroids. The similarity measure of shape is based on the Jaccard similarity between the two fuzzy sets once their centroids have been aligned. To convert the vector similarity to a single scalar, the product of $\mathrm{ss}_{\mathrm{shape}}$ and $\mathrm{ss}_{\mathrm{proximity}}$ is taken.

The subsethood measure is closely related to similarity and is based on Kosko's subsethood [47] for type-1 fuzzy sets. The measure of subsethood of a set A in another set B is defined as:

$$\mathrm{ss}_K(A,B) = \frac{|A \cap B|}{|A|}. \tag{6}$$

As with the similarity metric, when the set and cardinality operators are replaced by their fuzzy logic realizations, one obtains

$$\mathrm{ss}_K(A,B) = \frac{\sum_{i=1}^{N} \min(\mu_A(x_i), \mu_B(x_i))}{\sum_{i=1}^{N} \mu_A(x_i)} \tag{7}$$

for the case of type-1 fuzzy sets and for type-2 fuzzy sets the right hand side of the equation becomes

$$\frac{\sum_{i=1}^{N} \min(\overline{\mu}_A(x_i), \overline{\mu}_B(x_i)) + \sum_{i=1}^{N} \min(\underline{\mu}_A(x_i), \underline{\mu}_B(x_i))}{\sum_{i=1}^{N} \overline{\mu}_A(x_i) + \sum_{i=1}^{N} \underline{\mu}_A(x_i)}. \tag{8}$$

As opposed to similarity, subsethood is asymmetrical, i.e., $\mathrm{ss}_K(A,B) \neq \mathrm{ss}_K(B,A)$.

These equations give the similarity and subsethood measures for fuzzy variables of one dimension. To aggregate the similarity of the three dimensions of valence, activation, and dominance, we tried several methods: averaging the similarity of the individual dimensions $\mathbf{sm}_{\mathrm{avg}}(A,B) = 1/3\sum_{i\in\{\mathrm{Val.,Act.,Dom.}\}} \mathrm{sm}_i(A_i,B_i)$, taking the product of the similarity of the individual dimensions $\mathbf{sm}_{\mathrm{prod}}(A,B) = \prod_{i\in\{\mathrm{Val.,Act.,Dom.}\}} \mathrm{sm}_i(A_i,B_i)$, and taking the linguistic weighted average [48] $\mathbf{sm}_{lwa}(A,B) = \sum_{i\in\{\mathrm{Val.,Act.,Dom.}\}} \mathrm{sm}_i(A_i,B_i) w_i / \sum_{i\in\{\mathrm{Val.,Act.,Dom}\}} w_i$. The results of these different choices are described in Section V.

D. Interval Surveys Using the Interval Approach

To estimate the interval type 2 fuzzy sets over the valence, activation, and dominance scales, we used the interval approach [2], [3]. This survey methodology uses a Likert-like scale but the subjects select interval ranges instead of single numbers on the scale, which results in IT2 FSs. One of the novelties of our work that adds to [2], [3] is that we look at modeling a phenomenon where the underlying variable is composed of multiple scales: three separate scales (valence, activation, and dominance) in the case of our first model, and an open-ended number of scales in our second model.

The interval approach assumes that most people will be able to describe words on a scale, similar to a Likert scale. However, while the Likert scale approach allows the subject to choose only a single point on the scale, the interval approach allows the subject to select an interval that encloses the range on the scale that the word applies to. Thus, while a Likert scale can capture direction and intensity on a scale, the interval approach also captures uncertainty. This uncertainty that an individual user has about a word can be thought of as intra-user uncertainty. The user does not need to know about the details of interval type-2 fuzzy logic; they can indicate their uncertainty as an interval which is then aggregated into IT2 FSs by the interval approach, which represent inter-user uncertainty.

After collecting a set of intervals from an interval approach survey, the interval approach estimates an IT2 FS that takes into account the collective uncertainty of a group of subjects. This type of uncertainty can be thought of as inter-user uncertainty. The interval approach consists of a series of steps to learn the fuzzy sets from the survey data which can broadly be grouped into the data part and the fuzzy set part. The data part takes the survey data, preprocesses it, and computes statistics for it. The fuzzy set part creates type-1 fuzzy sets for each subject, and then aggregates them with the union operation to form IT2 FSs. A new version of the interval approach, the enhanced interval approach, was proposed in [19]. This enhancement aims to produce tighter membership functions by placing new constraints on the overlapping of subject-specific membership functions in the reasonable interval processing stage. We tested this method as well as the original interval approach and found that the enhanced interval approach did in fact yield tighter membership functions, but that this did not necessarily improve the overall performance measures when compared with the original method (c.f. Section VI).

IV. Methodology

This section describes the experimental methodologies that were used to create the two models for emotion codebooks. In the first, we use an interval approach survey for emotion words and we adapt the CWW paradigm to account for 3-dimensional fuzzy scales, specifically, by implementing similarity and subsethood measures for fuzzy sets that have 3 dimensions. In the case of the second model, the interval survey is separate from the elicitation of emotional information. The emotional information is collected from the EMO20Q game and thereafter the fuzzy sets are calculated from the answers to the questions in the game.

A. Emotion Vocabularies

In our experiments, we examined four different emotion vocabularies. The first vocabulary consisted of seven emotion category words: *angry*, *disgusted*, *fearful*, *happy*, *neutral*, *sad*, and *surprised*. These are commonly used emotion categories used for labeling emotional data. We refer to this vocabulary as *Emotion Category Words*. These emotions are posited to be basic in that they are reliably distinguishable from facial expressions [49].

TABLE 1 Similarity between words of the Blog Moods vocabulary and the Emotion Category Word vocabulary.

	ANGRY	DISGUSTED	FEARFUL	HAPPY	NEUTRAL	SAD	SURPRISED
AMUSED	0.004	0.003	0.005	**0.060**	0.004	0.005	0.053
TIRED	0.006	0.003	0.034	0.001	0.038	**0.196**	0.001
CHEERFUL	0.003	0.003	0.003	**0.109**	0.001	0.002	0.088
BORED	0.015	0.012	0.075	0.004	0.064	**0.335**	0.004
ACCOMPLISHED	0.015	0.013	0.008	**0.151**	0.006	0.008	0.139
SLEEPY	0.007	0.005	0.018	0.009	**0.172**	0.128	0.010
CONTENT	0.005	0.004	0.007	**0.044**	0.015	0.012	0.040
EXCITED	0.015	0.017	0.006	**0.255**	0.002	0.002	0.213
CONTEMPLATIVE	0.006	0.004	0.012	0.006	**0.161**	0.075	0.007
BLAH	0.014	0.010	0.049	0.005	0.166	**0.359**	0.007
AWAKE	0.020	0.017	0.016	0.061	0.015	0.014	**0.068**
CALM	0.003	0.002	0.011	0.007	**0.137**	0.069	0.008
BOUNCY	0.009	0.012	0.002	**0.361**	0.000	0.001	0.311
CHIPPER	0.002	0.002	0.001	**0.066**	0.002	0.003	0.059
ANNOYED	**0.393**	0.380	0.080	0.041	0.002	0.023	0.076
CONFUSED	0.026	0.020	0.064	0.014	0.046	**0.170**	0.017
BUSY	0.068	0.079	0.049	0.111	0.013	0.012	**0.116**
SICK	0.008	0.004	0.032	0.001	0.023	**0.204**	0.001
ANXIOUS	**0.207**	0.181	0.091	0.028	0.003	0.025	0.038
EXHAUSTED	0.015	0.011	0.048	0.003	0.046	**0.298**	0.004
DEPRESSED	0.008	0.005	0.050	0.001	0.015	**0.218**	0.001
CURIOUS	0.038	0.042	0.014	**0.203**	0.011	0.006	0.176
DRAINED	0.009	0.007	0.039	0.002	0.061	**0.280**	0.003
AGGRAVATED	0.578	**0.618**	0.114	0.047	0.002	0.020	0.087
ECSTATIC	0.000	0.000	0.000	0.108	0.000	0.000	**0.117**
BLANK	0.006	0.004	0.017	0.005	0.133	**0.137**	0.006
OKAY	0.016	0.013	0.035	0.017	**0.076**	0.057	0.020
HUNGRY	**0.084**	0.082	0.029	0.045	0.013	0.034	0.052
HOPEFUL	0.009	0.007	0.007	0.047	0.010	0.009	**0.050**
COLD	0.005	0.003	0.026	0.001	0.047	**0.123**	0.002
CREATIVE	0.027	0.037	0.007	**0.524**	0.001	0.002	0.462
PISSED_OFF	**0.383**	0.363	0.052	0.016	0.000	0.008	0.035
GOOD	0.004	0.003	0.004	**0.067**	0.005	0.006	0.060
THOUGHTFUL	0.005	0.003	0.004	0.011	**0.079**	0.029	0.012
FRUSTRATED	0.186	**0.233**	0.068	0.022	0.001	0.012	0.030
CRANKY	0.325	**0.351**	0.099	0.045	0.002	0.022	0.060
STRESSED	0.288	**0.304**	0.158	0.044	0.003	0.026	0.053

The second vocabulary consisted of 40 words taken from the top 40 emotion mood labels used by the bloggers of LiveJournal (this blogging site lets users label each post with a mood label, which has been used as an annotated corpus for studying emotional text [50]). The words in this vocabulary are: *accomplished*, *aggravated*, *amused*, *angry*, *annoyed*, *anxious*, *awake*, *blah*, *blank*, *bored*, *bouncy*, *calm*, *cheerful*, *chipper*, *cold*, *confused*, *contemplative*, *content*, *cranky*, *crazy*, *creative*, *curious*, *depressed*, *disgusted*, *drained*, *ecstatic*, *excited*, *exhausted*, *fearful*, *frustrated*, *good*, *happy*, *hopeful*, *hungry*, *neutral*, *okay*, *pissed off*, *sad*, *sick*, *sleepy*, *stressed*, *thoughtful*, and *tired*. We refer to this vocabulary as *Blog Moods*.

The third vocabulary was a list of 30 Spanish emotion words that was taken from the mental health initiative of a Southern California medical service provider. The words in the Spanish emotion vocabulary are: *aburrido*, *agobiado*, *agotado*, *ansioso*, *apenado*, *asqueado*, *asustado*, *avergonzado*, *cauteloso*, *celoso*, *cómodo*, *confiado*, *confundido*, *culpable*, *deprimido*, *enamorado*, *enojado*, *esperanzado*, *extático*, *feliz*, *frustrado*, *histérico*, *malicioso*, *pasmado*, *rabioso*, *solitario*, *sorpredido*, *sospechoso*, *timido*, and *triste* (see Table 1 in [17] for glosses of these words from a Spanish-English dictionary). We refer to this vocabulary as *Spanish Emotion Words*.

The fourth vocabulary was elicited from subjects playing EMO20Q, both between two humans and also between a human and computer with the computer in the questioner role. [37], [38], [40]. These data sources resulted in a set of 105 emotion words.

B. Valence, Activation, and Dominance Model (Model 1)

The data collected from the interval surveys for the first model consists of four experiments: three surveys of 32 subjects for English and one survey of eight subjects for Spanish. All surveys had a similar structure. First, the surveys gave the subject instructions. Then the surveys sequentially presented the subject with emotion words, which we will refer to as the stimuli, one word per page. For each stimulus there were sliders for each of the three emotion dimensions. The sliders had two handles, which allowed the subjects to select the lower and upper points of ranges. The range of the sliders was 0–10. The maximum range allowed was 10 and the minimum range was 1 because the steps were integer values and the implementation imposed a constraint that the upper and lower endpoints could not be the same. Above each scale was a pictorial representations known as a *self-assessment manikin* [31] that aimed to illustrate the scale non-verbally.

The overall structure of the Spanish survey was the same as the English one, but special care was required for the translation of the instructions and user interface elements. The first version of the translation was done by a proficient second-language Spanish speaker and later versions were corrected by native Spanish speakers. The subjects of the surveys were native speakers of Spanish with Mexican and Spanish backgrounds.

In the surveys, each subject was presented with a series of randomized stimuli from one of the emotion vocabularies. The description of the stimuli regimen and other implementation details for the experiments can be found in [16] for English and [17] for Spanish. Links to the surveys can be found at http://sail.usc.edu/~kazemzad/emotion_in_text_cgi/.

One final issue was deciding whether similarity or subsethood was best for our task and how to aggregate these metrics for three dimensions. Both similarity and subsethood can be used as an objective function to be maximized by translation. [23, Chapter 4] recommends using subsethood when the output is a classification and similarity if the input and output vocabularies are the same, but it was not immediately clear what would be preferable for our tasks, so we tested the different methods empirically. Also, since this is one of the first studies that uses fuzzy sets that range over more than one dimension, we tested several ways of combining the similarities and subsethoods of the individual scales using the average, product, and linguistic weighted average as described in Section III-C. We also tried leaving dominance out as it is a distinguishing feature in only a few cases.

The mapping from one vocabulary to another is done by choosing the word from the output vocabulary that has the highest similarity or subsethood with the input word. Here, similarity and subsethood are the aggregated scalewise similarities and subsethoods for valence, activation, and dominance.

We examined several different mappings. In [16], we examined mapping from the blog mood vocabulary to the more controlled categorical emotion vocabulary, which simulates the task of mapping from a large, noisy vocabulary to a more controlled one. In this paper, we use mapping tasks that involved translation from Spanish to English to evaluate the estimated IT2 FSs.

To empirically evaluate the performance of the mapping, we used a human translator to complete a similar mapping task. We instructed the translator to choose the best word or, if necessary, two words from the output vocabulary that matched the input word. A predicted result was considered correct if it matched one of the output words chosen by the evaluator.

We also use multidimensional scaling to visualize the derived emotion space. Multidimensional scaling is similar to principal component analysis except that it operates on a similarity matrix instead of a data or covariance matrix. Since it operates directly on a similarity matrix, it is ideal for visualizing the results of aggregating the scale-wise similarities into a single similarity matrix.

C. Propositional Model (Model 2)

We devised the second model to address the results obtained from model 1, described in Section V, where we found that larger vocabulary sizes resulted in lower performance in the translation tasks. Our inspiration for the second model was that people can guess an object from a large, open-ended set by adaptively asking a sequence of questions, as in the game of twenty questions. The sequential questioning behavior thus motivated our representation and experimental design of the EMO20Q. The premiss of EMO20Q is that the

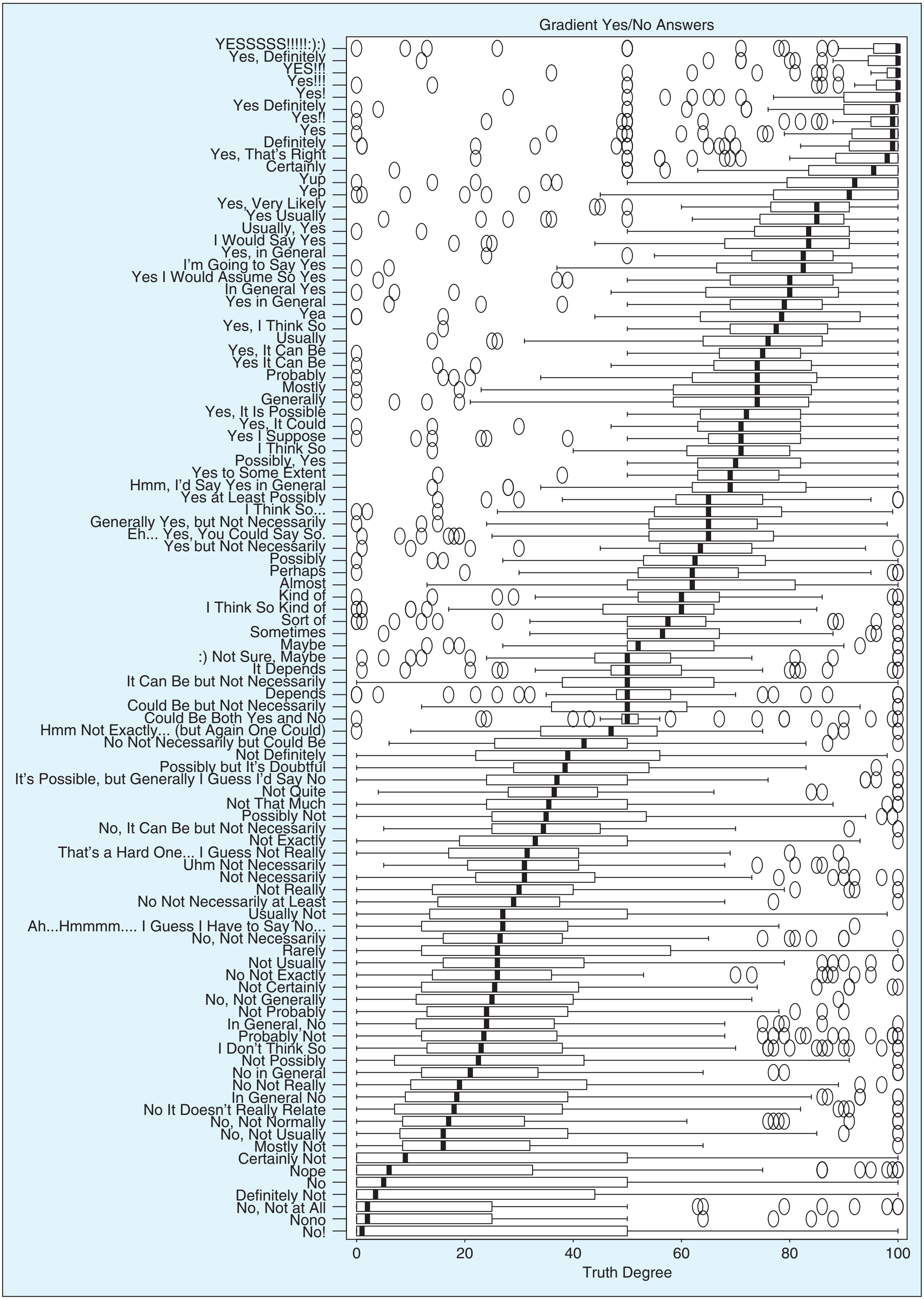

FIGURE 3 Fuzzy answers to yes/no questions obtained by presenting the answer phrase (*x*-axis labels) to users of Amazon Mechanical Turk, who responded by using a slider interface to indicate the truth-degree (*y*-axis). This plot was based on a single handle slider, in contrast to the interval approach surveys, in order to show an overview of the data. The results presented below are for the double handle slider and interval approach analysis.

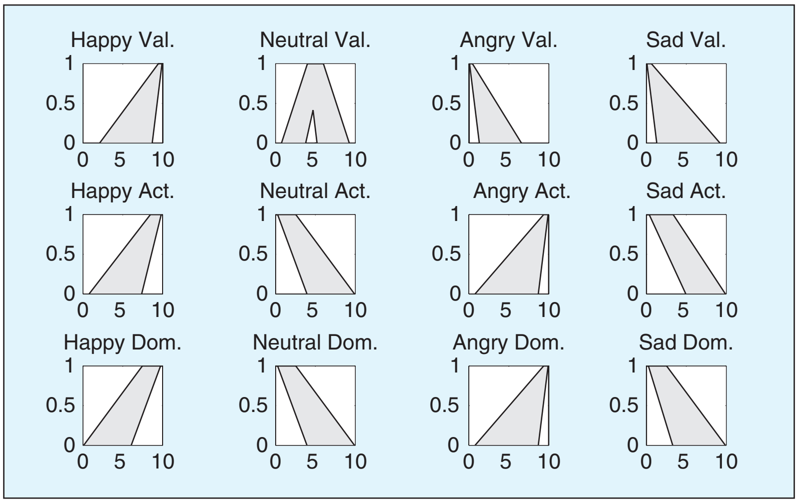

FIGURE 4 Example membership functions (MF's) calculated with the interval approach for happy, neutral, angry, and sad emotions. All the membership functions shown here, except the valence for neutral, are shoulder MF's that model the edges of the domain of μ. The region between the upper and lower MF's, the footprint of uncertainty, is shaded. The variables of Val., Act., and Dom. stand for valence, activation, and dominance.

twenty questions game is a way to elicit human knowledge about emotions and that the game can also be used to test the ability of computer agents to simulate knowledge about emotions. The experimental design of the EMO20Q game was proposed in [37] and since then we have collected data from over 100 human-human and over 300 human-computer EMO20Q games. In this paper we focus on follow-up experiments that aim to understand the answers in the game in terms of fuzzy logic. More information about EMO20Q, including demos, code, and data, can be found at http://sail.usc.edu/emo20q.

Although the questions asked in the EMO20Q game are required to be yes-no questions, the answers are not just "yes" or "no." Often the answer contains some expression of uncertainty. Here we focus on the

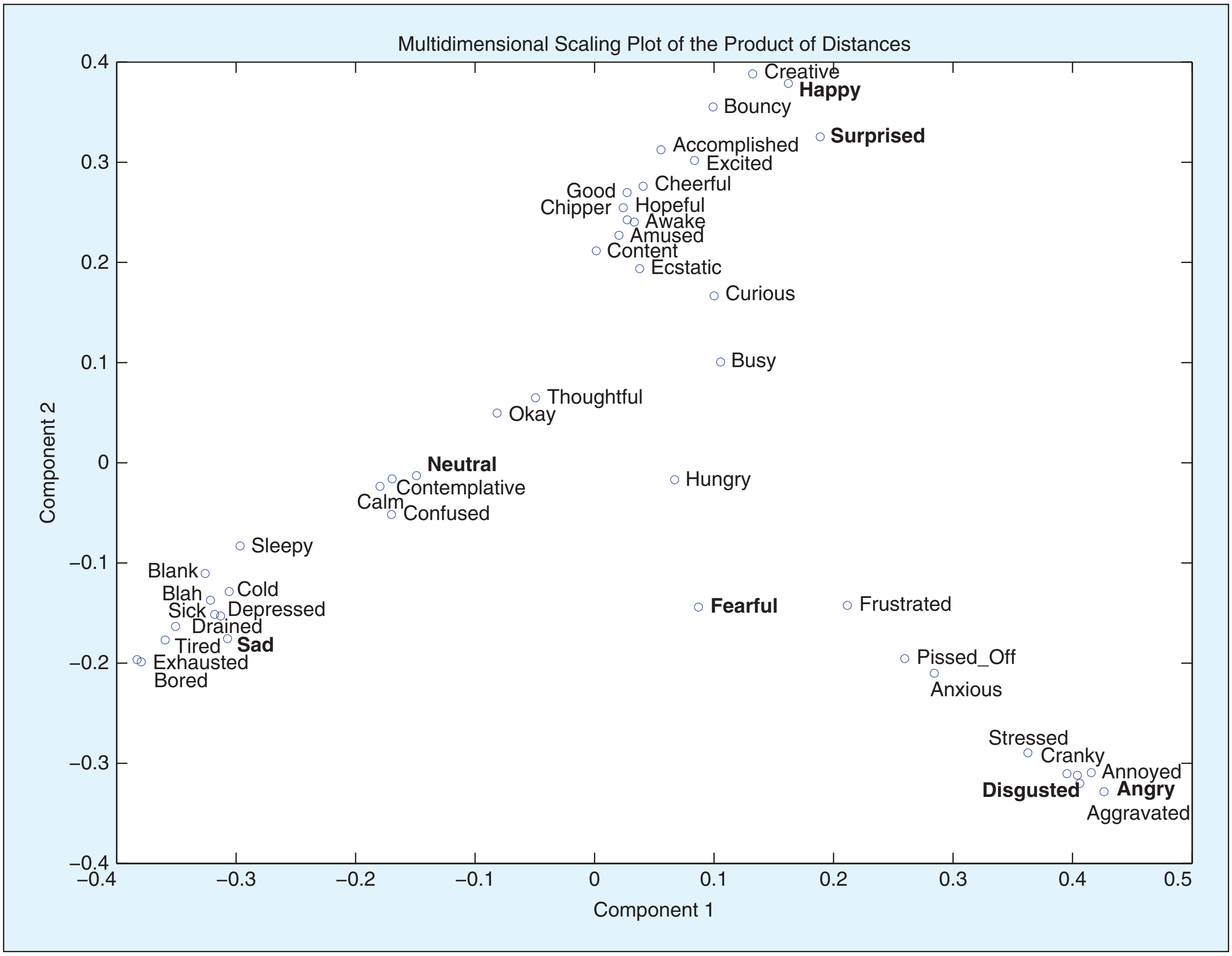

FIGURE 5 Multidimensional scaling (2-D) representation of the emotion words' similarity. This visualizes when the similarity of the individual valence, activation, and dominance dimensions were combined by taking their product. The words in the categorical emotion vocabulary are marked in bold.

fuzzy logical representation of answers to questions in the game. Just as the first model uses valence, activation, and dominance scales to represent emotions, the second model uses the questions from EMO20Q as scales that can be interpreted on axes that range from "yes" to "no." In this case, the interval surveys we performed were not overtly about emotions, but rather to evaluate the answers on the scale from "no" to "yes," which we defined as a domain for fuzzy sets that range from 0 to 100.

Using data from EMO20Q games, we collected a set of questions and answers about emotions. We sampled a set of answers based on frequency of occurrence and how well the set covered the space from affirmative to negative answers. We also included some control stimuli not observed in the data but included to provide insight on how people would interpret negation. For example, we included phrase groups like "certainly," "not certainly" and "certainly not" that would allow us to calibrate how the subjects would interpret phrases that might have a logical interpretation. The final set of stimuli consisted of 99 answers. These were presented to subjects along with either a single or double handle slider. Below in Figure 3, we plot the responses for single sliders, which are easier to visualize than double sliders. In what follows, however, we present the double handle slider results, which form the input to the interval approach methodology described above.

We conducted the interval approach survey on Amazon Mechanical Turk (AMT), an internet marketplace for crowd sourcing tasks that can be completed online. The survey was conducted in sets of 30 stimuli to each of 137 subjects on AMT who were ostensibly English speakers from the U.S. The average amount of ratings per stimulus was 38.5.

V. Experimental Results

In this section, we present the results of experiments that used the two models and the survey methodology described in Sections III-D, IV-B, and IV-C to estimate fuzzy set membership functions for the emotion vocabularies presented in Section IV-A, to calculate similarity and subsethood between emotion words as described in Section III-C, and to map between different emotion vocabularies.

A. Valence, Activation, and Dominance Model (Model 1)

Examples of the membership functions that were calculated for the emotion category vocabulary can be seen in Fig. 4. The distances between these membership functions and those of the blog moods vocabulary can be seen in Table 1, as calculated using the product of the individual scale-wise similarities as the aggregation method. In Fig. 5 we display the results of calculating a similarity matrix between the words of both vocabularies using *multidimensional scaling* (MDS) [51]. MDS is a statistical approach in the same family as principal components analysis (PCA) and factor analysis. We use MDS in this case because factor analysis has unwanted assumptions (namely, a multivariate normal distribution with linear relationships) and because PCA operates on feature vectors as opposed to similarity matrices (and also assumes linear relationships). We performed MDS on the aggregated similarity measurements to qualitatively visualize the emotion space as derived from the similarity matrix. The result of combining the similarities of the valence, activation, and dominance dimensions was slightly different using sum versus product aggregation. The sum aggregation produced a more spread out distribution of the words in the space induced by MDS, while the product aggregation produced a space where the emotions are more tightly clustered. This was because the product aggregation method was less sensitive to small dissimilarities. The multidimensional scaling plot also allows one to see which emotions are close and potentially confusable. For example, "happy" and "surprised" are very close, as are "angry" and "disgusted." Since mapping between vocabularies, like MDS, is done using similarities, this implies that these pairs are confusable. Since the components derived from MDS are calculated algorithmically, they are not directly interpretable as in the case of factor analysis.

TABLE 2 Similarity between Spanish and English emotion words.

	ANGRY	DISGUSTED	FEARFUL	HAPPY	NEUTRAL	SAD	SURPRISED
ABURRIDO	0.2284	0.2335	**0.6370**	0.1965	0.3196	0.4610	0.1230
AGOBIADO	0.4762	**0.5696**	0.4611	0.3122	0.1495	0.2895	0.2175
AGOTADO	0.2250	0.2344	0.4883	0.1425	0.4081	**0.5135**	0.1012
ANSIOSO	0.4579	**0.4748**	0.2837	0.3655	0.2703	0.1728	0.3598
APENADO	0.2915	0.2928	**0.7711**	0.3128	0.1219	0.4065	0.1211
ASQUEADO	0.5445	**0.5969**	0.3885	0.4538	0.2045	0.2784	0.3199
ASUSTADO	0.4610	**0.5324**	0.3209	0.3508	0.2213	0.2141	0.3489
AVERGONZADO	0.2701	0.2663	**0.6345**	0.2393	0.0660	0.4737	0.0713
CAUTELOSO	0.0918	0.0957	**0.5357**	0.1848	0.3784	0.3126	0.0958
CELOSO	**0.7396**	0.6880	0.3335	0.1832	0.0515	0.2444	0.2390
CÓMODO	0.0436	0.0510	0.3363	0.3686	**0.3963**	0.3518	0.2240
CONFIADO	0.2835	0.3307	0.2382	**0.4753**	0.1393	0.0562	0.2821
CONFUNDIDO	0.2488	0.2531	**0.7690**	0.2202	0.1286	0.4498	0.0878
CULPABLE	0.3275	0.3445	**0.7051**	0.2916	0.1375	0.3921	0.1401
DEPRIMIDO	0.2893	0.2914	0.5585	0.1529	0.3380	**0.7058**	0.0978
ENAMORADO	0.4371	0.5611	0.0942	0.4572	0.1055	0.0351	**0.5774**
ENOJADO	**0.8732**	0.7125	0.3596	0.1940	0.1054	0.2654	0.3494
ESPERANZADO	0.0929	0.0987	0.4023	**0.5903**	0.1798	0.1625	0.3270
EXTÁTICO	0.3140	0.3108	0.0611	0.4305	0.1337	0.0268	**0.7222**
FELIZ	0.1329	0.1655	0.2293	**0.6020**	0.1796	0.0770	0.5046
FRUSTRADO	0.6414	**0.7271**	0.3003	0.3021	0.1677	0.3026	0.3337
HISTÉRICO	0.6522	**0.6566**	0.2804	0.2340	0.1550	0.1874	0.4272
MALICIOSO	0.3347	**0.4270**	0.3427	0.3540	0.2273	0.1322	0.2325
PASMADO	0.3102	0.3480	**0.3910**	0.2544	0.1931	0.3231	0.2654
RABIOSO	**0.5416**	0.4616	0.2190	0.0945	0.0018	0.1402	0.3598
SOLITARIO	0.2657	0.2672	**0.6091**	0.0904	0.2549	0.5565	0.0396
SORPRENDIDO	0.3405	**0.3803**	0.1229	0.3336	0.1706	0.0746	0.3675
SOSPECHOSO	0.3026	0.3497	**0.5129**	0.3883	0.2084	0.2425	0.2900
TIMIDO	0.0844	0.0857	0.3925	0.1092	0.3578	**0.4436**	0.0515
TRISTE	0.3376	0.3396	**0.6502**	0.1477	0.2389	0.5882	0.0852

To check the mapping induced by the similarity matrices, we show in Table 1 the similarity matrix for the product aggregation of the dimension-wise similarity measures of the valence, activation, and dominance scales. The location of the maximum of each row (bold) shows the final translation from the larger vocabulary (rows) to the smaller vocabulary (columns). The most glaring error is that "fearful" is not in the range of the mapping from large vocabulary to small vocabulary due to relatively low similarity to any word in the blog mood vocabulary. Cases where one would expect to have a mapping to "fearful" (e.g., "anxious," "stressed") do show elevated similarity to "fearful" but "angry" or "disgusted" are higher. The observation that most of the values in the "fearful" column are lower than the other columns, we normalized each column by its maximum value. Doing this does in fact produce the intuitive mapping of "anxious" and "stressed" to "fearful," but also changed other values.

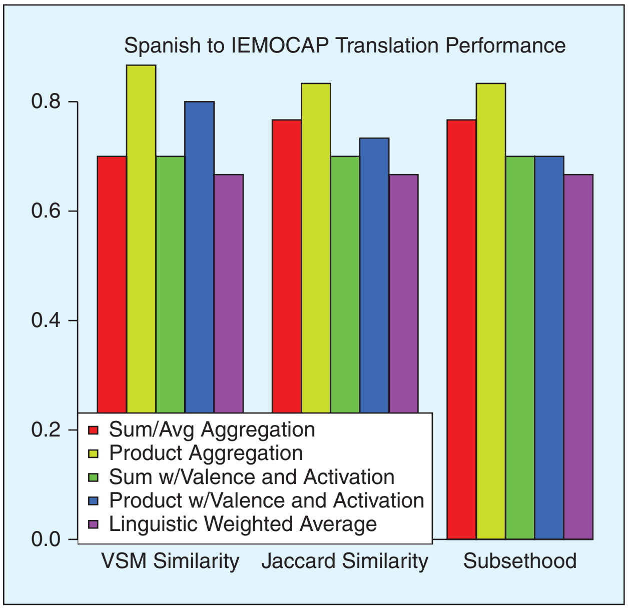

FIGURE 6 Performance of translating from the Spanish emotion vocabulary to the categorical emotion vocabulary, which was the set of emotion labels used for annotating the IEMOCAP corpus [52].

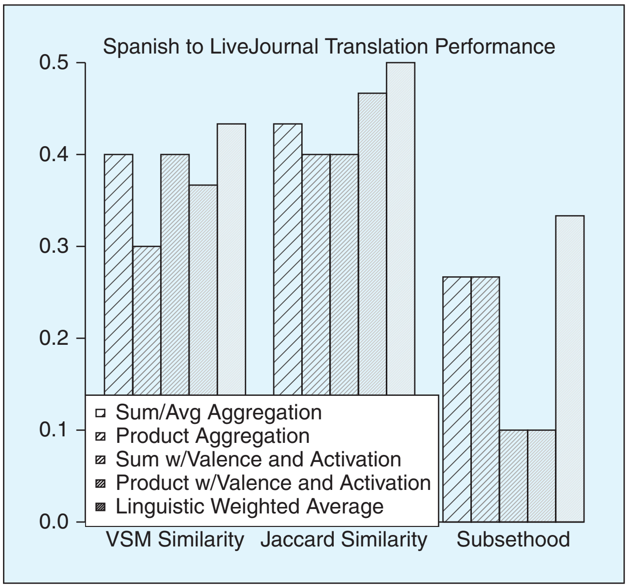

FIGURE 7 Performance of translating Spanish emotion words to liveJournal mood labels (colloquial emotion words).

To better quantify the intuitive goodness of the mapping from one vocabulary to another, we undertook an evaluation based on human performance on the same mapping task. We found that at least one of the subject's choices matched the predicted mapping except in the following five cases (i.e., performance of approximately 84%): "confused," "busy," "anxious," "hungry," and "hopeful." Filtering out clearly nonemotion words like "hungry" may have improved the results here, but our aim was to use a possibly noisy large vocabulary, since the data came from the web.

To see if the fuzzy logic approach agreed with a simpler approach, we converted the survey interval end-points to single points by taking the midpoints of the subjects' intervals and then averaging across all subjects. As points in the 3-D emotion space, the mapping performance of Euclidean distance was essentially the same as those determined by the fuzzy logic similarity measures. However, a simple Euclidean distance metric loses some of the theoretical benefits we have argued for, as it does not account for the shape of the membership functions and cannot account for subsethood.

Based on the membership functions from the Spanish survey and the previous English surveys, we constructed similarity matrices between the Spanish words as input and the English words as output. The similarity matrix of the Spanish words and the Emotion Category Word vocabulary are shown in Table 2. Overall, the best performance of 86.7% came from mapping from the Spanish vocabulary to the Emotion Category Word vocabulary using similarity (rather than subsethood), and aggregating the scale-wise similarities using the multiplicative product of the three scales. The performance of mapping from Spanish to the Blog Mood vocabulary was worse that with the Emotion Category Word vocabulary as output because the much larger size of the Blog Mood vocabulary resulted in more confusability. The best performance for this task was 50% using similarity and linguistic weighted average for aggregating the similarities. A comparison of the different similarity and aggregation methods can be seen in Fig. 6 for mapping from Spanish to the Emotion Category Word vocabulary and Fig. 7 for mapping from Spanish to the Blog Moods vocabulary.

B. Propositional Model (Model 2)

For the propositional model, we collected a set of 1228 question-answer pairs from 110 human-human EMO20Q matches, in which 71 unique emotion words were chosen. In these matches, the players successfully guessed the other players' emotion words in 85% of the matches, requiring on average 12 turns.

In the set of question-answer pairs there were 761 unique answer strings. We selected a set of 99 answers based on frequency of occurrence and how well the set covered the space from affirmative to negative answers. We used the interval approach to obtain fuzzy sets for the answers to yes/no questions.

A sample of these are shown in Figure 8. To evaluate these, we determined the extent to which the medians from the single handle slider survey were a full or partial members in the fuzzy sets derived from interval approach's double handle slider survey, which used different subjects but the same stimuli. We found that the IT2 FSs from the interval approach surveys corresponded well with the single-slider data. All of the estimated IT2 FSs except one contained the median of the single-slider values, i.e., 99%. This word, "NO!", was a singleton IT2 FS at zero, while the median from the single slider was at one (on the scale from 0 to 100). The average value of the IT2 FS membership functions (which is an interval-valued range) at points corresponding to the median of the single-slider values was (0.41,0.84). To evaluate the enhanced interval approach (EIA), we found that the EIA-derived IT2 FSs performed nearly as well. The IT2 FSs contained all but two of the median single-slider (~98%) and the average membership of the median single-slider values was (0.12,0.89).

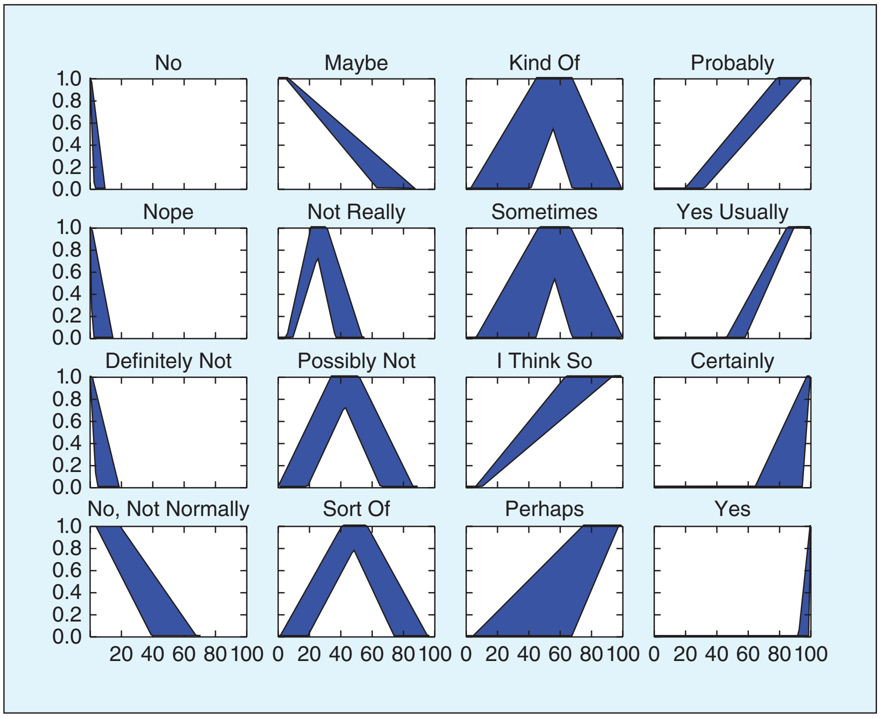

FIGURE 8 Example IT2 FSs calculated with the enhanced interval approach for answers to yes/no questions.

Beyond these quantitative measurements, the membership functions from model 2 are qualitatively tighter than those of model 1, especially with the enhanced interval approach. Though some of the membership functions span large portions of the domain, these are answers that signify uncertainty (such as "kind of," "I think so," and "perhaps" in Figure 8). This was in contrast to model 1, which more frequently resulted in broad membership functions with wide footprints of uncertainty. The data and code for the experiments of model 2 can be accessed at http://code.google.com/p/cwwfl/.

VI. Discussion

Variables that range over sets and functions rather than individual numbers are important developments for modern mathematics, and further, variables that range over proofs, automata languages, and programs further add to the richness of objects that can be represented with variables. This paper looked at expanding the domain of variables to include emotions. To model a seemingly non-mathematical object in such a way, we use fuzzy sets, another relatively new type of variable. This paper proposed two models for emotion variables, one that represented the meaning of emotion words on a three dimensional axis of valence, activation, and dominance, and another that represented emotions as a sparse vector of truth values over propositions about emotions.

First, we examine the relative benefits and drawbacks of the two models we proposed: the first model based on valence, activation, and dominance scales, and the second model based on questions about emotions whose answers are rated on a scale from true to false.

The first model captures intuitive gradations between emotions. For example, the relation of "ecstatic" and "happy" can be seen in their values on the scales: "ecstatic" will a subset of "happy" with valence and activation values more to the extreme periphery. Also, the scales used by the first model are language-independent, iconic representations of emotion, which enables researchers to use the same scales for multiple languages.

However, for the first model, each word needs an interval survey on the three scales to calculate the membership function for the word, which is laborious and limits the model to words whose membership functions have been calculated already. Also, as we have seen, performance degrades with the size of the vocabulary. Some of the performance degradation can be expected due to the inherent difficulty of making a decision with more choices. However, limiting the representation to three scales does also limit the resolution and expressiveness of the model.

The second model, on the other hand, gives a better resolution when there is a large number of emotions. With more emotions, more expressivity is needed than just valence, activation, and dominance. To give examples of some of the emotion words from EMO20Q that are difficult to represent with only valence,

We consider two basic relations on emotional variables: similarity and subsethood.

activation, and dominance, we can see that "pride," "vindication," and "confidence," might all have similar valence, activation, and dominance values, so it would be hard to distinguish these on the basis of only the three scales. By representing emotions with propositions based on questions from EMO20Q, we can use a single fuzzy scale for any arbitrary proposition: once the scales are established the bulk of the data can be collected purely in natural language. Moreover, the propositional truth-value scale can be used for other domains besides emotions.

However, with the second model there is no clear way to compare emotions that were not asked the same set of questions. In the EMO20Q game, the questions are seen as they occur in the game. It will be necessary to collect more data outside of the game to make sure that all the prevalent questions are asked about each emotion. Even though we can use a single fuzzy scale for each proposition's truth-value the set of all propositions about emotions is a vast open set, so data collection is still an issue. Since the propositions are based on a specific human language, the equivalence of different propositions in different languages is not as apparent as in the first model.

There were several modifications that we made to the interval approach to make it more robust for when all intervals are discarded by the preprocessing. It was determined that the final removal of all intervals took place in the reasonable interval processing stage. The modification to the original interval approach involved keeping the intervals in this stage if all would have been removed. This had the effect of creating a very broad membership function with a lower membership function that was zero at all points. The enhanced interval approach improved the rejection of intervals in various stages by separately considering interval endpoint criteria and interval length criteria. For the first model, the enhanced interval approach yielded worse results when using the translation task as a evaluation metric. This was due to the narrower membership functions that the enhanced interval approach was designed to produce. In the case of similarity and subsethood calculation, the narrower membership function led to more zero entries in the calculation of similarity and subsethood. In the translation task, this resulted in a less robust translation because small variations in the membership function would yield a disproportionate change in similarity and subsethood values. However, in the case of the second model, where the fuzzy sets are used in a more traditional fashion, i.e., as propositional truth quantifiers, the enhanced interval approach did in fact yield membership functions that appeared to more tightly contain the single slider results and performed as well on the evaluation metric we used for this task.

The different models both use IT2 FSs, but beyond that, they present different approaches in the representation of emotion descriptions. Because of the difference in approach and the resulting format of the model, they were difficult to evaluate in the same way. For the first model, because the fuzzy scales of valence, activation, and dominance are directly tied to the emotion representation and because the scales are nonlinguistic in nature (they are labeled with a cartoon manikin), the cross-language translation task was a possible evaluation metric. However, the fuzzy scales used in the second model are indirectly linked to emotions via linguistic propositions about emotions. Since the propositions about emotions are specific to a given language, the translation task is not directly facilitated by this model.

From the comments given by the subjects of the survey, for model 1, we found that subjects reported confusion with the scale of dominance, despite the pictorial representation in the survey. For model 2, we found that the interpretation of linguistic truth values was a source of reflection for the subjects and this provided insight into the variation that may have otherwise been attributed to lack of cooperation on the part of the Amazon Mechanical Turkers. For example, the stimulus "definitely," from a logical point of view would be assumed to be a strong "yes." However, several Turkers mentioned that they realized that, when they use the word "definitely," they do not mean "definitely" in the logical sense, but rather that the colloquial meaning is somewhat more relaxed. From the fuzzy set representation point of view, it may be advantageous to recognize distinct senses for the meaning of words and phrases. In the case mentioned, the word "definitely" could have colloquial sense and a logical sense. Another example of this was in the control phrases we used in the second model. For example "not certainly" was often confused with "certainly not." This is not to say that all the Turkers were cooperative and took the time to understand the task, but it shows that there are many factors involved with measuring uncertainty. From Figure 3, we can see that the default value of the slider (in this case, a single slider at the middle of the scale) was a salient point of outliers. Modeling the effects of uncooperative users who may click through as quickly as possible is one possible improvement that could be made to the interval approach from the data processing point of view.

Our conclusion in comparing the two models is that for basic emotions the valence, activation, and dominance scales of model 1 would suffice. Examples of a use-case for the first model would be for converting a large, expressive set of emotion labels to a smaller set for the purpose of training a statistical classifier. However, for the class of all words used to describe emotions in natural language, the representational power of first model's valence, activation, and dominance scales is not sufficient. To fully understand what a given emotion word means to someone, our work indicates that the second model is a better model if the modeling goal is to represent a larger vocabulary and finer shades of meaning.

VII. Conclusions

In this paper we presented two models to represent the meaning of emotion words. We gave an explicit description of meaning in our models. The first model involved interpreting the emotion words as three-dimensional IT2 FSs on the dimensions of valence, activation, and dominance. This model allowed us to map between emotion vocabularies of different sizes and different languages. The mapping was induced by picking the most similar word of the output vocabulary given the input vocabulary word. The similarity used for this mapping was derived from similarity or subsethood measures of the individual dimensions that were aggregated into a single measure for each pair of input and output vocabulary words. We devised a second model that addresses the challenges that arise when the vocabulary of emotion words is large. Instead of the lower dimensional representation in terms of valence, activation, and dominance scales, the second model used a high dimensional representation where the emotion words were represented in terms of answers to questions about emotions, as determined from data from the EMO20Q game. In the second model, IT2 FSs were used to represent the truth values of answers to questions about emotions. We found that the second model was necessary to capture more highly nuanced meaning when the vocabulary of emotion words was large.

Acknowledgment

The authors would like to thank Jerry Mendel, Dongrui Wu, Mohammad Reza Rajati, Ozan Cakmak, and Thomas Forster for their discussion. We would also like to thank Rebeka Campos Astorkiza, Eduardo Mendoza Ramirez, and Miguel Ángel Aijón Oliva for helping to translate the Spanish version of our experiment.

References

[1] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *J. Res. Personality*, vol. 11, pp. 273–294, Sept. 1977.
[2] F. Liu and J. M. Mendel, "An interval approach to fuzzistics for interval type-2 fuzzy sets," in *Proc. Fuzzy Systems Conf.*, 2007, pp. 1–6.
[3] F. Liu and J. M. Mendel, "Encoding words into interval type-2 fuzzy sets using an interval approach," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 6, pp. 1503–1521, 2008.
[4] D. W. Massaro and M. M. Cohen, "Fuzzy logical model of bimodal emotion perception: Comment on 'The perception of emotions by ear and by eye' by de Gelder and Vroomen," *Cogn. Emotion*, vol. 14, no. 3, pp. 313–320, 2000.
[5] M. Grimm, K. Kroschel, E. Mower, and S. Narayanan, "Primitives-based evaluation and estimation of emotions in speech," *Speech Commun.*, vol. 49, pp. 787–800, Dec. 2006.
[6] C. M. Lee and S. Narayanan, "Emotion recognition using a data-driven inference system," in *Proc. Eurospeech*, Geneva, Switzerland, 2003, pp. 157–160.
[7] D. Wu, T. D. Parsons, E. Mower, and S. Narayanan, "Speech parameter estimation in 3D space," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2010, pp. 737–742.
[8] A. Konar, A. Chakraborty, A. Halder, R. Mandal, and R. Janarthanan, "Interval type-2 fuzzy model for emotion recognition from facial expression," in *Proc. Perception Machine Intelligence*, 2012, pp. 114–121.
[9] M. El-Nasr, J. Yen, and T. R. Ioerger, "Flame: Fuzzy logic adaptive model of emotions," *Auton. Agents Multi-Agent Syst.*, vol. 3, no. 3, pp. 219–257, 2009.
[10] C. M. Whissell, *The Dictionary of Affect in Language*. New York: Academic Press, 1989, pp. 113–131.
[11] A. Kazemzadeh, "Précis of dissertation proposal: Natural language descriptions of emotions," in *Proc. ACII (Doctoral Consortium)*, 2011, pp. 216–223.
[12] S. Kim, P. G. Georgiou, S. S. Narayanan, and S. Sundaram, "Supervised acoustic topic model for unstructured audio information retrieval," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, 2010, pp. 243–246.
[13] S. Sundaram and S. S. Narayanan, "Classification of sound clips by two schemes: Using onomatopoeia and semantic labels," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2008, pp. 1341–1344.
[14] M. Grimm and K. Kroschel, "Rule-based emotion classification using acoustic features," in *Proc. Int. Conf. Telemedicine Multimedia Communication*, 2005.
[15] Q. Liang and J. Mendel, "Interval type-2 fuzzy logic systems: theory and design," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 5, pp. 535–550, 2000.
[16] A. Kazemzadeh, S. Lee, and S. Narayanan, "An interval type-2 fuzzy logic system to translate between emotion-related vocabularies," in *Proc. Interspeech*, pp. 2747–2750, 2008.
[17] A. Kazemzadeh, "Using interval type-2 fuzzy logic to translate emotion words from Spanish to English," in *Proc. IEEE World Conf. Computational Intelligence FUZZ-IEEE Workshop*, 2010, pp. 1–8.
[18] O. Cakmak, A. Kazemzadeh, and S. Yildirim, and S. Narayana, "Using interval type-2 fuzzy logic to analyze Turkish emotion words," in *Proc. APSIPA Annu. Summit Conf.*, 2012, pp. 1–4.
[19] S. Coupland, J. M. Mendel, and D. Wu, "Enhanced interval approach for encoding words into interval type-2 fuzzy sets and convergence of the word FOUs," in *FUZZ-IEEE World Cong. Computational Intelligence*, 2010, pp. 1–8.
[20] J. M. Mendel, R. I. John, and F. Liu, "Computing with words and its relations with fuzzistics," *Inform. Sci.*, vol. 177, no. 4, pp. 988–1006, 2007.
[21] J. M. Mendel, "Computing with words: Zadeh, Turing, Popper and Occam," *IEEE Comput. Intell. Mag.*, vol. 2, no. 4, pp. 10–17, 2007.
[22] L. A. Zadeh, "Fuzzy logic = computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 4, pp. 103–111, May 1996.
[23] J. M. Mendel and D. Wu, *Perceptual Computing: Aiding People in Making Subjective Judgements*. Piscataway, NJ: IEEE Press, 2010.
[24] J. M. Mendel and D. Wu, "Challenges for perceptual computer applications and how they were overcome," *IEEE Comput. Intell. Mag.*, vol. 7, pp. 36–47, Aug. 2012.
[25] G. Frege, "Über sinn und bedeutung," in *Zeitschrift für Philosophie und Philosophische Kritik*, 1892, pp. 25–50.
[26] T. Forster, *Logic, Induction, and Sets*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[27] B. Ganter, G. Stumme, and R. Wille, Eds., *Formal Concept Analysis: Foundation and Applications*. Berlin, Germany: Springer-Verlag, 2005.
[28] C. E. Osgood, G. J. Suci, and P. H. Tannenbaum, *The Measurement of Meaning*. Urbana, IL: Univ. Illinois Press, 1957.
[29] D. E. Heise, "Semantic differential profiles for 1000 most frequent English words," *Psychol. Monographs*, vol. 79, no. 8, pp. 1–31, 1965.
[30] J. A. Russell, "A circumplex model of affect," *J. Personality Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.
[31] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Therapy Exp. Psych.*, vol. 25, pp. 49–59, Mar. 1994.
[32] J. R. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, "The world of emotions is not two-dimensional," *Psychol. Sci.*, vol. 18, pp. 1050–1057, Dec. 2007.
[33] I. B. Türksen, "Computing with descriptive and veristic words," in *Proc. Int. Conf. North American Fuzzy Information Processing Society*, 1999, pp. 13–17.
[34] L. A. Zadeh, "From search engines to question answering systems–the problems of world knowledge, relevance, deduction and precisiation," in *Fuzzy Logic and the Semantic Web*, E. Sanchez, Ed. The Netherlands: Elsevier, 2006, ch. 9, pp. 163–211.
[35] M. R. Rajati, H. Khaloozadeh, and W. Pedrycz, "Fuzzy logic and self-referential reasoning: A comparative study with some new concepts," *Artificial Intell. Rev.*, pp. 1–27, Mar. 2012.
[36] T. Forster, *Reasoning About Theoretical Entities*. Singapore: World Scientific, 2003.
[37] A. Kazemzadeh, P. G. Georgiou, S. Lee, and S. Narayanan, "Emotion twenty questions: Toward a crowd-sourced theory of emotions," in *Proc. ACII'11*, 2011, pp. 1–10.
[38] A. Kazemzadeh, J. Gibson, P. Georgiou, S. Lee, and S. Narayanan, "EMO20Q questioner agent," in *Proc. ACII (Interactive Event)*, 2011, pp. 313–314.
[39] A. Kazemzadeh, S. Lee, P. G. Georgiou, and S. Narayanan, "Determining what questions to ask, with the help of spectral graph theory," in *Proc. Interspeech*, pp. 2053–2056, 2011.
[40] A. Kazemzadeh, J. Gibson, J. Li, S. Lee, P. G. Georgiou, and S. Narayanan, "A sequential Bayesian agent for computational ethnography," in *Proc. Interspeech*, Portland, OR, 2012.
[41] L. F. Barrett, "Are emotions natural kinds?" *Perspectives Psychol. Sci.*, vol. 1, pp. 28–58, Mar. 2006.
[42] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-I," *Inform. Sci.*, vol. 8, no. 3, pp. 199–249, 1975.
[43] R. John and S. Coupland, "Type-2 fuzzy logic: Challenges and misconceptions," *IEEE Comput. Intell. Mag.*, vol. 7, pp. 47–52, Aug. 2012.
[44] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Upper Saddle River, NJ: Prentice Hall, pp. 451–453, 2001.
[45] J. M. Mendel, R. I. John, and F. Liu, "Interval type-2 fuzzy logic systems made simple," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 6, pp. 808–821, 2006.
[46] D. Wu and J. Mendel, "A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets," *Inform. Sci.*, vol. 178, no. 2, pp. 381–402, 2008.
[47] B. Kosko, "Fuzzyness vs. probability," *Int. J. General Syst.*, vol. 17, nos. 2–3, pp. 211–240, 1990.
[48] D. Wu and J. M. Mendel, "The linguistic weighted average," in *FUZZ-IEEE,* Vancouver, BC, pp. 566–573, 2006.
[49] P. Ekman, "Facial expression and emotion," *Amer. Psychol.*, vol. 48, no. 4, pp. 384–392, 1993.
[50] G. Mishne, "Applied text analytics for blogs," Ph.D. dissertation, Univ. Amsterdam, Amsterdam, The Netherlands, 2007.
[51] T. F. Cox and M. A. A. Cox, *Multidimensional Scaling*, 2nd ed. Boca Raton, FL: CRC Press, 2000.
[52] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. Chang, S. Lee, and S. Narayanan, "IEMOCAP: Interactive emotional dyadic motion capture database," *J. Lang. Resour. Eval.*, vol. 42, no. 4, pp. 335–359, 2008.