ISCA Archive
http://www.isca-speech.org/archive

7th ISCA Workshop on
Speech Synthesis (SSW-7)
Kyoto, Japan
September 22-24, 2010

Toward Naturally Expressive Speech Synthesis: Data–Driven Emotion Detection Using Latent Affective Analysis

Jerome R. Bellegarda

Speech & Language Technologies,
Apple Inc.
Cupertino, California 95014, USA
jerome @ apple.com

Abstract

A necessary step in the generation of expressive speech synthesis is the automatic detection and classification of emotions most likely to be present in textual input. Though increasingly data-driven, emotion analysis still relies on critical expert knowledge to isolate the emotional keywords or keysets necessary to the construction of affective categories. This makes it vulnerable to any discrepancy between affective states and domain of discourse. This paper proposes a more general strategy, which leverages two separate semantic levels: one encapsulates the foundations of the domain considered, while the other specifically accounts for the overall affective fabric of the language. Exposing the emergent relationship between these two levels advantageously informs the emotion classification process. Empirical evidence suggests that this approach is effective for automatic emotion analysis in text. This bodes well for its deployability toward naturally expressive speech synthesis.

Index Terms: expressive speech synthesis, affective congruence, detection and classification of emotional states, latent semantic analysis.

1. Introduction

The critical influence of emotion on human communication makes it indispensable to take account of emotional states as an integral part of human-computer interaction, at both input and output levels [1]. If a spoken dialog system could reliably determine that a user is upset or annoyed, for instance, it could switch to a potentially more adequate mode of interaction [2]. Likewise, expressive speech synthesis is expected to play a pivotal role in the widespread deployment and acceptance of future natural language interfaces [3].

Imparting emotional quality to synthetic speech requires making certain speech segments sound happy or subdued, angry or scornful, authoritative or uncertain, etc., as appropriate in the given communicative situation considered [4]. In order to do so, it is necessary to solve two complementary sub-problems: (i) the appropriate emotion must be identified from the given input text, and (ii) the corresponding signal modifications, if any, must be effected in speech generation. A number of techniques, often closely tied to the synthesis framework adopted, have been explored to address (ii): cf., e.g., [5]. Comparatively less work has been done on (i): typical applications focus on emotion identification in a social media context, mostly to understand and leverage how people react to content on the Internet: see, e.g., [6].

In this work, we focus on the automatic detection and classification of emotions in text with the specific perspective of an eventual integration into a text-to-speech (TTS) system. Emotion detection can thus be viewed as the necessary first step in the generation of *naturally expressive* synthetic speech, where (ideally) any emotion conveyed would be congruent with the subject matter and discourse context under consideration.

Identifying emotions in textual input presupposes the existence of some suitable taxonomy of emotional states. Emphasis has traditionally been placed on the set of six "universal" emotions [7]: ANGER, DISGUST, FEAR, JOY, SADNESS, and SURPRISE [8] – [10]. Emotion analysis is typically carried out using a simplified description of such emotional states in a low-dimensional space, which normally comprises dimensions such as valence (positive/negative evalution), activation (stimulation of activity), and/or control (dominant/submissive power) [11] – [13]. Classification proceeds based on an underlying emotional knowledge base, which strives to provide adequate distinctions between different emotions. This affective information can either be built entirely upon manually selected vocabulary as in [14], or derived automatically from data based on expert knowledge of the most relevant features that can be extracted from the input text [8]. In both cases, the net effect is to rely on a few thousand annotated "emotional keywords" (or, more generally, "emotional keysets"), the presence of which is the trigger for the associated emotional label(s).

The drawback of such confined lexical affinity is that it only supports simplified relationships between affective words and emotional categories. As a result, emotion analysis tends to be hampered by the bias inherent in the relatively few core terms explicitly taken into account in taxonomy construction. This has sparked interest in data-driven approaches based on latent semantic analysis (LSA), a paradigm originally developed for information retrieval [15]. Upon suitable training using a large corpus of texts, LSA allows a similarity score to be computed between generic terms and affective categories [16]. This way, every word can automatically be assigned some fractional affective influence. Still, the affective categories themselves are usually specified with the help of a reference lexical database like WordNet [17].

This paper aims at more broadly exploiting the principle of latent semantics in emotion analysis. We cast the problem as a general application of *latent semantic mapping* (LSM), an extrapolation of LSA for modeling global relationships implicit in large volumes of data [18], [19]. More specifically, we use the LSM framework to expose two distinct semantic levels: one that encapsulates the foundations of the domain considered (e.g., broadcast news, email messages, SMS conversations, etc.), and one that specifically accounts for the overall affective fabric of the language. Then, we leverage these two descriptions to appropriately relate domain and affective levels, and thereby in-

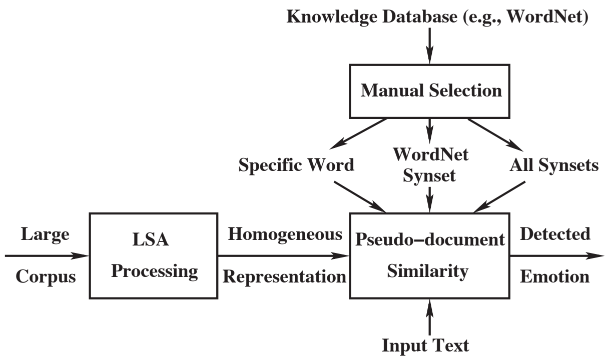

Figure 1: *Emotion Detection using Conventional Analysis.*

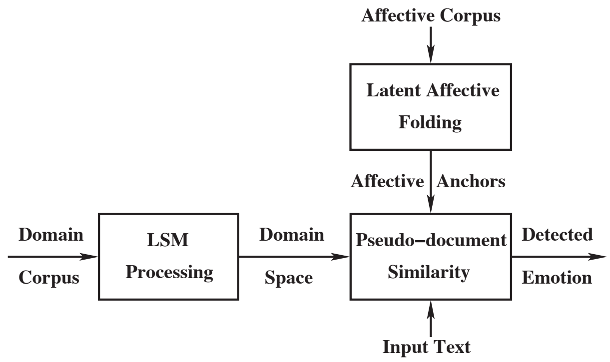

Figure 2: *Emotion Detection using Latent Affective Analysis.*

form the emotion classification process. This *de facto* bypasses the need for any explicit external knowledge in the form of emotional keywords or keysets.

The paper is organized as follows. The next section provides some further motivation for latent semantics and discusses the limitations of conventional solutions. Section 3 gives a general overview of latent affective analysis. In Section 4, we describe in detail the mechanics of the proposed approach and subsequent emotion detection. Finally, Section 5 reports the outcome of experimental evaluations conducted on the "Affective Text" portion of the SemEval-2007 corpus [20].

2. Conventional Analysis

As alluded to above, lexical affinity alone fails to provide sufficient distinction between different emotions, in large part because only relatively few words have inherently clear, unambiguous emotional meaning. For example, *happy* and *sad* encapsulate JOY and SADNESS, respectively, in all conceivable scenarios. But is *thrilling* a marker of JOY or SURPRISE? Does *awful* capture SADNESS or DISGUST? It largely depends on contextual information: *thrilling* as a synonym for *uplifting* conveys JOY (as in *a thrilling speech*), while *thrilling* as a synonym for *amazing* may well mark SURPRISE (as in *a thrilling waterfall ride*); similarly, *awful* as a synonym for *grave* reflects SADNESS (as in *an awful car accident*), while *awful* as a synonym for *foul* is closer to DISGUST (as in *an awful smell*). The vast majority of words likewise carry multiple potential emotional connotations, with the degree of affective polysemy tightly linked to the granularity selected for the underlying taxonomy of emotions.

Data-driven approaches based on LSA purport to "individuate" such indirect affective words via inference mechanisms automatically derived in an unsupervised way from a large corpus of texts, such as the British National Corpus (BNC) [16]. By looking at document-level co-occurrences, contextual information is exploited to encapsulate semantic information into a relatively low dimensional vector space. Suitable affective categories are then constructed in that space by "folding in" either the specific word denoting the emotion, or its associated synset (say, from WordNet), or even the entire set of words in all synsets that can be labelled with that emotion [13]. This is typically done by placing the relevant word(s) into a "pseudo-document," and mapping it into the space as if it were a real one [15]. Finally, the global emotional affinity of a given input text is determined by computing its similarity to all pseudo-documents. The resulting procedure is depicted in Fig. 1.

This solution is attractive, if for no other reason than it allows every word to automatically be assigned some fractional affective influence. However, it suffers from two limitations which may well prove deleterious in practical situations. First, the inherent lack of supervision routinely leads to a latent semantic space which is not particularly representative of the underlying domain of discourse. And second, the construction of the affective categories still relies heavily on pre-defined lexical affinity, potentially resulting in an unwarranted bias in the taxonomy of affective states.

The first limitation impinges on the effectiveness of any LSA-based approach, which is known to vary substantially based on the size and quality of the training data [19], [21]. In the present case, any discrepancy between underlying space and domain of discourse may distort the position of certain words in the space, which could in turn lead to subsequent sub-optimal affective weight assignment. For instance, in the examples above, the word *smell* is a great deal more critical to the resolution of *awful* as a marker of DISGUST than the word *car*. But that fact may never be uncovered if the only pertinent documents in the training corpus happen to be about high-end fragrances and wrecked automobiles. Thus, it is highly desirable to derive the latent semantic space using data representative of the application at hand. This points to a modicum of supervision.

The second limitation is tied to the difficulty of coming up with an *a priori* affective description that will work universally. Stipulating the affective categories using only the specific word denoting the emotion is likely to be less robust than using the set of words in all synsets labelled with that emotion. On the other hand, the latter may well expose some inherent ambiguities resulting from affective polysemy. This is compounded by the relatively small number of words for which an affective distribution is even available. For example, the well-known General Inquirer content analysis system [22] lists only about 2000 words with positive outlook and 2000 words with negative outlook. There are exactly 1281 words inventoried in the affective extension of WordNet [13], and the affective word list from [23] comprises less than 1000 words. This considerably complicates the construction of reliable affective categories in the latent space.

3. Latent Affective Analysis

To address the two limitations above, we propose to more broadly leverage the LSM paradigm [18, 19], following the overall procedure depicted in Fig. 2. Compared to Fig. 1, we inject some supervision at two separate levels: not only regarding

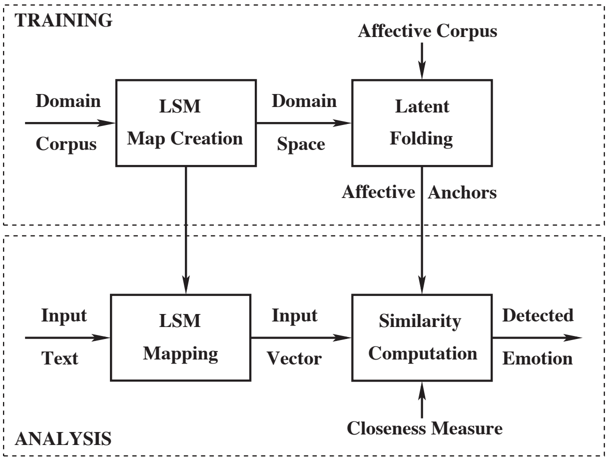

Figure 3: *Implementation of Latent Affective Analysis.*

the particular domain considered, but also how the affective categories themselves are defined. The specification of a suitable training collection centered on the application at hand allows LSM to encapsulate the general foundations of the domain into a representative latent semantic space. At the same time, the specification of a separate affective training collection, such as mood-annotated blog entries from LiveJournal.com [13], serves as a descriptive blueprint for the data-driven construction of affective categories.

We exploit this blueprint by invoking the same "folding" approach used in standard LSA. Affective categories emerge in the domain space via a process called latent affective folding, which leverages the fractional influence of each training document encountered within a particular emotion to infer the overall *affective anchor* for this emotion in the latent semantic space. The goal is to superimpose these data-driven affective anchors (which each can be viewed as the centroid of a particular affective category) on top of conventional domain information.

This is analogous to what happens in Fig. 1, with a crucial difference regarding the reliability of affective categories: in latent affective folding, they are derived from a corpus of texts as opposed to manually selected keywords or keysets. This reduces any potential discrepancy between affective states and discourse vocabulary, which in turn helps resolve ambiguities in emotional connotations.

Finally, the input text is mapped into the latent semantic space as before. Emotion classification then follows from assessing how closely it aligns with each of the affective anchors just produced.

4. Implementation

Expanding the basic framework of Fig. 2 to take into account the two separate phases of training and analysis, latent affective analysis is implemented as illustrated in Fig. 3. In what follows, we will use the subscript $_1$ for domain-related entities, and the subscript $_2$ for entities associated with the affective realm.

4.1. LSM Space Construction

Let $\mathcal{T}_1$, $|\mathcal{T}_1| = N_1$, be a collection of training texts (be they sentences, paragraphs, or documents) reflecting the domain of interest, and $\mathcal{V}_1$, $|\mathcal{V}_1| = M_1$, the associated set of all words (possibly augmented with some strategic word pairs, triplets, etc., as appropriate) observed in this collection. Generally, M_1 is on the order of several tens of thousands, while N_1 may be as high as a million.

We first construct a $(M_1 \times N_1)$ matrix W_1, whose elements w_{ij} suitably reflect the extent to which each word $w_i \in \mathcal{V}_1$ appeared in each text $t_j \in \mathcal{T}_1$. From [19], a reasonable expression for w_{ij} is:

$$w_{i,j} = (1 - \varepsilon_i)\, \frac{c_{i,j}}{n_j}\, , \tag{1}$$

where $c_{i,j}$ is the number of times w_i occurs in text t_j, n_j is the total number of words present in this text, and ε_i is the normalized entropy of w_i in $\mathcal{V}_1$. The global weighting implied by $1 - \varepsilon_i$ reflects the fact that two words appearing with the same count in a particular text do not necessarily convey the same amount of information; this is subordinated to the distribution of words in the entire set $\mathcal{V}_1$.

If we denote by $f_{i,j} = c_{i,j} / \sum_j c_{i,j}$ the frequency of occurrence of w_i in text t_j relative to its occurrence in the entire collection $\mathcal{T}_1$, the expression for ε_i is easily seen to be:

$$\varepsilon_i = -\frac{1}{\log N} \sum_{j=1}^{N} f_{i,j} \log f_{i,j}\, . \tag{2}$$

By definition, $0 \le \varepsilon_i \le 1$, with equality if and only if $f_{i,j} = 1$ and $f_{i,j} = 1/N$, respectively. A value of ε_i close to 1 indicates a word distributed across many texts throughout the corpus, while a value of ε_i close to 0 means that the word is present only in a few specific texts. The global weight $1 - \varepsilon_i$ is therefore a measure of the indexing power of the word w_i.

We then perform a singular value decomposition (SVD) of W_1 as [19]:

$$W_1 = U_1\, S_1\, {V_1}^T\, , \tag{3}$$

where U_1 is the $(M_1 \times R_1)$ left singular matrix with row vectors $u_{1,i}$ $(1 \le i \le M_1)$, S_1 is the $(R_1 \times R_1)$ diagonal matrix of singular values $s_{1,1} \ge s_{1,2} \ge \ldots \ge s_{1,R_1} > 0$, V_1 is the $(N_1 \times R_1)$ right singular matrix with row vectors $v_{1,j}$ $(1 \le j \le N_1)$, $R_1 \ll M_1, N_1$ is the order of the decomposition, and ^{T} denotes matrix transposition. On Fig. 3, this process is called *LSM map creation*.

As is well known, both left and right singular matrices U_1 and V_1 are column-orthonormal, i.e., ${U_1}^T U_1 = {V_1}^T V_1 = I_{R_1}$ (the identity matrix of order R_1). Thus, the column vectors of U_1 and V_1 each define an orthornormal basis for the space of dimension R_1 spanned by the $u_{1,i}$'s and $v_{1,j}$'s. We refer to this space as the *latent semantic space* $\mathcal{L}_1$. The (rank-R_1) decomposition (3) encapsulates a mapping between the set of words w_i and texts t_j and (after apropriate scaling by the singular values) the set of R_1-dimensional vectors $y_{1,i} = u_{1,i} S_1$ and $z_{1,j} = v_{1,j} S_1$.

The basic idea behind (3) is that the rank-R_1 decomposition captures the major structural associations in W_1 and ignores higher order effects. Hence, the relative positions of the input words in the space $\mathcal{L}_1$ reflect a parsimonious encoding of the semantic concepts used in the domain under consideration. This means that any new text mapped onto a vector "close" (in some suitable metric) to a particular set of words can be expected to be closely related to the concept encapsulated by this set. If each of these words is then scored in terms of their affective affinity, this offers a way to automatically predict the overall emotional affinity of the text.

4.2. Affective Anchors

In order to do so, we need to isolate regions in that space which are representative of the underlying taxonomy of emotions adopted. The centroid of each such region is the *affective anchor* associated with that basic emotion. Affective anchors are superimposed onto the space $\mathcal{L}_1$ on the basis of the affective corpus available.

Let $\mathcal{T}_2$, $|\mathcal{T}_2| = N_2$, represent a separate training collection of mood-annotated texts (again they could be sentences, paragraphs, or documents), representative of the desired categories of emotions (such as JOY and SADNESS). We denote by $\mathcal{V}_2$, $|\mathcal{V}_2| = M_2$, the associated set of words or expressions observed in this collection. As such affective data may be more difficult to gather than regular texts (especially in annotated form), in practice $N_2 < N_1$.

Next, consider the representation in the LSM space $\mathcal{L}_1$ of the texts in $\mathcal{T}_2$. Using standard folding, each text can be treated as a regular pseudo-document and mapped into the space in the usual manner. This results in a set of vectors $\nu_{2,k}$ $(1 \leq k \leq N_2)$ which can be viewed as functionally equivalent to the set of vectors $z_{1,j}$ $(1 \leq j \leq N_1)$, albeit for the affective rather than the domain data.

Now let L be the number of distinct emotions selected. It is possible to group all N_2 texts in $\mathcal{T}_2$ into L subsets $\mathcal{T}_2^{(\ell)}$, one for each affective category.[1] For each $1 \leq \ell \leq L$, the representation of the subset $\mathcal{T}_2^{(\ell)}$ in the space $\mathcal{L}_1$ is therefore prototypical of that particular emotion.

This forms the basis for computing the region associated with each emotion ℓ in the domain space. By gathering all relevant $\nu_{2,k}$, we can compute:

$$\hat{z}_{1,\ell} = \frac{1}{|\mathcal{T}_2^{(\ell)}|} \sum_{\mathcal{T}_2^{(\ell)}} \nu_{2,k} \, , \tag{4}$$

as the centroid of each pertinent region in $\mathcal{L}_1$. This is turn defines the affective anchor representing each emotion ℓ $(1 \leq \ell \leq L)$ in the semantic space. Referring back to Fig. 3, this process is known as *latent folding*.

The notation $\hat{z}_{1,\ell}$ is chosen to underscore the connection with $z_{1,j}$: in essence, $\hat{z}_{1,\ell}$ represents the (fictitious) text in the domain space that would be perfectly aligned with emotion ℓ, had it been seen the training collection $\mathcal{T}_1$. Comparing the representation of an input text to each of these anchors therefore leads to a quantitative assessment for the overall emotional affinity of the text.

4.3. Emotion Classification

In order to do so, we first need to specify how to represent in that space an input text not seen in the training corpus, say t_p (where $p > N_1$). For each entry in $\mathcal{T}_1$, we compute for the new text the weighted counts (1) with $j = p$. The resulting feature vector, a column vector of dimension M_1, can be thought of as an additional column of the matrix W_1. Assuming the matrices U_1 and S_1 do not change appreciably, the SVD expansion (3) therefore implies:

$$t_p = U_1 \, S_1 \, v_{1,p}^T \, , \tag{5}$$

where the R_1-dimensional vector $v_{1,p}^T$ acts as an additional column of the matrix $V_1{}^T$. Thus, the represention of the new text in the domain space can be obtained from $z_{1,p} = v_{1,p} S_1$. On Fig. 3, this process is called *LSM mapping*.

All is needed now is a suitable closeness measure to compare this representation to each affective anchor $\hat{z}_{1,\ell}$ $(1 \leq \ell \leq L)$. From [19], a natural metric to consider is the cosine of the angle between them. This yields:

$$\mathcal{C}(z_{1,p}, \hat{z}_{1,\ell}) = \frac{z_{1,p} \, \hat{z}_{1,\ell}^T}{\|z_{1,p}\| \, \|\hat{z}_{1,\ell}\|} \, , \tag{6}$$

for any $1 \leq \ell \leq L$. Using (6), it is a simple matter to directly compute the relevance of the input text to each emotional category. Referring back to Fig. 3 one last time, this process is known as *similarity computation*. It is important to note that word weighting is now implicitly taken into account by the LSM formalism.

5. Experimental Evaluation

We assessed the benefits of the latent affective analysis approach proposed above on a standard semantic evaluation database via a systematic comparison with conventional techniques based on expert knowledge of emotional keywords and keysets.

5.1. Test Database

The experimental setup used the data set that was developed for the SemEval 2007 task on "Affective Text" [20]. This task was focused on the emotion classification of news headlines. Headlines normally consist of a few words and are often written by creative people with the intention to "provoke" emotions, and consequently attract the readers' attention. These characteristics make this kind of data particularly suitable for use in an automatic emotion recognition setting, as the affective/emotional features (if present) are guaranteed to appear in these typically short sentences.

The test data accordingly consisted of 1,250 short news headlines[2] extracted from news web sites (such as Google news, CNN, etc.) and/or newspapers, and annotated along $L = 6$ emotions (ANGER, DISGUST, FEAR, JOY, SADNESS, and SURPRISE) by 6 different annotators. The reader is referred to [20] for detailed information on data annotation, including studies on inter-annotator agreement.

5.2. Baseline Systems

For baseline purposes, we selected two different kinds of systems: (i) based entirely upon manually selected vocabulary as in [14], and (ii) based on LSA trained on a large corpus of texts as in [16]. For (i), we used a simple word accumulation strategy, which annotates the emotions in a text based on the presence of words from the WordNet-Affect lexicon [13].

For (ii), we used all three LSA-based systems implied in Fig. 1. They only differ in the way each emotion is represented in the LSA space: either based on a specific word only (e.g., JOY), or the word plus its WordNet synset, or the word plus all WordNet synsets labelled with that emotion in WordNet-Affect (cf. [20]). In all three cases, the large corpus used for LSA processing was the Wall Street Journal (WSJ) text collection [24], comprising about 86,000 articles.

[1] Note that, in the case of multiple emotional annotations, each text could conceivably contribute to several such subsets.

[2] Development data was merged into the original SemEval 2007 test set to produce a larger test set.

Table 1: *Results on SemEval-2007 Test Corpus.*

Approach Considered	Precision	Recall	F-Measure
Baseline Word Accumulation	44.7	2.4	4.6
LSA (Specific Word Only)	11.5	65.8	19.6
LSA (With WordNet Synset)	12.2	77.5	21.1
LSA (With All WordNet Synsets)	11.4	89.6	20.3
Latent Affective Analysis	18.8	90.1	31.1

5.3. Latent Affective Training

For latent affective analysis, we needed to select two separate training corpora. For the "domain" corpus, we selected a collection of about $N_1 = 8,500$ relatively short English sentences (with a vocabulary of roughly $M_1 = 12,000$ words) originally compiled for the purpose of a building a concatenative text-to-speech voice. Though not completely congruent with news headlines, we felt that the type and range of topics covered was close enough to serve as a good proxy for the domain.

For the "affective" corpus, we relied on about $N_2 = 5,000$ mood-annotated blog entries from LiveJournal.com, with a filtered[3] vocabulary of about $M_2 = 20,000$ words. The indication of mood being explicitly specified when posting on LiveJournal, without particular coercion from the interface, mood-annotated posts are likely to reflect the true mood of the blog authors [13]. The moods were then mapped to the $L = 6$ emotions considered in the classification.

Finally, we formed the domain matrix W_1 and processed it as in (3). We used $R_1 = 100$ for the dimension of the domain space $\mathcal{L}_1$, and computed the affective anchors as per (4).

5.4. Experimental Results

We adopted the standard criteria of precision, recall, and F-measure to compare latent affective analysis to the four baseline systems mentioned above. The results are summarized in Table 1.

For reference purposes, comparable F-measure results by participants in the SemEval-2007 Task 14 were in the 10 to 20 range [20]. We also note that our LSA results are comfortingly close to those reported in [13], even though the LSA training corpus was different (WSJ versus BNC).

Consistent with the observations in [13], word accumulation secures the highest precision at the cost of the lowest recall, while LSA-based systems achieve high recall but significantly lower precision. Encouragingly, the F-measure obtained with latent affective analysis is substantially higher than with all baseline approaches. This suggests an improved ability to resolve distinctions between emotional connotations.

6. Conclusion

We have proposed a new strategy for the data-driven analysis of emotion in text. This strategy articulates around two coupled phases: (i) separately encapsulate both the foundations of the domain considered and the overall affective fabric of the language, and (ii) exploit the emergent relationship between these two semantic levels of description in order to inform the emotion classification process. We address (i) by leveraging the latent topicality of two distinct corpora, as uncovered by a global LSM analysis. Domain and affective descriptions are then superimposed to produce the desired connection between all terms and emotional categories. Because this connection automatically takes into account the influence of the entire training corpora, it is more encompassing than that based on the relatively few affective terms typically selected in conventional processing.

Empirical evidence gathered on the "Affective Text" portion of the SemEval-2007 corpus [20] shows the effectiveness of the proposed strategy: latent affective analysis outperforms affectively weighted word accumulation, as well as standard LSA approaches based on expert knowledge of emotional keywords or keysets. It thus appears to be a promising solution for automatic emotion analysis in text. This in turn bodes well for its general deployability as a critical pre-processing step in the generation of naturally expressive speech synthesis, where any emotion conveyed would be congruent with the textual input under consideration.

Future efforts will concentrate on expanding the basic premise underlying latent affective analysis into a more general framework which supports different mapping instantiations. Indeed, a potential drawback of the implementation described in Section 4 is that the computation (4) is patently sensitive to the distribution of words within $\mathcal{T}_2$, which may be quite different from the distribution of words within $\mathcal{T}_1$. In such a case, latent affective folding may well introduce a bias in the position of the affective anchors in the domain space. In [25], we derive an alternative mapping technique to alleviate this problem.

[3] Extensive text pre-processing is usually required on blog entries, to address typos and assorted creative license.

7. References

[1] R.W. Picard, *Affective Computing*, Cambridge, MA: MIT Press, 1997.

[2] J. Liscombe, G. Riccardi, and D. Hakkani–Tür, "Using Context to Improve Emotion Detection in Spoken Dialog Systems," *Proc. Interspeech*, Lisbon, Portugal, 1845–1848, 2005.

[3] E. Cosatto, J. Ostermann, H.P. Graf, and J. Schroeter, "Lifelike talking faces for interactive services," in *Proc. IEEE*, **91**(9), 1406–1429, 2003.

[4] M. Schröder, "Approaches to Emotional Expressivity in Synthetic Speech," in *The Emotion in the Human Voice*, Vol. 3, K. Izdebski [Ed], San Diego, CA: Plural, 2008.

[5] M. Schröder, "Expressing Degree of Activation in Synthetic Speech," *IEEE Trans. Audio, Speech, and Language Processing*, **14**(4):1128–1136, 2006.

[6] http://www.moodbase.com, 2010.

[7] P. Ekman, "Facial Expression and Emotion", *American Psychologist*, **48**(4), 384–392, 1993.

[8] C. Ovesdotter Alm, D. Roth, and R. Sproat, "Emotions from Text: Machine Learning for Text–Based Emotion Prediction," in *Proc. Conf. Human Language Technology and Empirical Methods in NLP*, Vancouver, BC, 579–586, 2005.

[9] H. Liu, H. Lieberman, and T. Selker, "A Model of Textual Affect Sensing Using Real-World Knowledge," in *Proc. Intelligent User Interfaces (IUI)*, Miami, FL, 125–132, 2003.

[10] P. Subasic and A. Huettner, "Affect Analysis of Text Using Fuzzy Semantic Typing," *IEEE Trans. Fuzzy Systems*, **9**(4):483–496, 2001.

[11] A. Mehrabian, "Framework for a Comprehensive Description and Measurement of Emotional States," *Genetic, Social, and General Psychology Monographs*, **121**(3):339–361, 1995.

[12] J.A. Russell, "A Circumplex Model of Affect," *J. Personality and Social Psychology*, **39**:1161–1178, 1980.

[13] C. Strapparava and R. Mihalcea, "Learning to Identify Emotions in Text," in *Proc. 2008 ACM Symposium on Applied Computing*, New York, NY, 1556–1560, 2008.

[14] C.M. Whissell, "The Dictionary of Affect in Language," in *Emotion: Theory, Research, and Experience*, R. Plutchik and H. Kellerman [Eds], New York, NY: Academic Press, 13–131, 1989.

[15] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman, "Indexing by Latent Semantic Analysis," *J. Amer. Soc. Information Science*, **41**:391–407, 1990.

[16] C. Strapparava, A. Valitutti, and O. Stock, "The Affective Weight of Lexicon," in *Proc. 5th Int. Conf. Language Resources and Evaluation (LREC)*, Lisbon, Portugal, 2006.

[17] C. Fellbaum [Ed], *WordNet: An Electronic Lexical Database*, Cambridge, MA: MIT Press, 1998.

[18] J.R. Bellegarda, "Latent Semantic Mapping: A Data–Driven Framework for Modeling Global Relationships Implicit in Large Volumes of Data," *IEEE Signal Processing Magazine*, **22**(5):70–80, 2005.

[19] J.R. Bellegarda, *Latent Semantic Mapping: Principles & Applications*, Synthesis Lectures on Speech and Audio Processing Series, Fort Collins, CO: Morgan & Claypool, 2008.

[20] C. Strapparava and R. Mihalcea, "SemEval-2007 Task 14: Affective Text," in *Proc. 4th Int. Workshop on Semantic Evaluations (SemEval 2007)*, Prague, Czech Republic, 2007.

[21] M. Mohler and R. Mihalcea, "Text–to–Text Semantic Similarity for Automatic Short Answer Grading," in *Proc. 12th Conf. European Chap. ACL*, Athens, Greece, 567–575, 2009.

[22] P.J. Stone, "Thematic Text Analysis: New agendas for Analyzing Text Content," in *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*, C.W. Roberts [Ed], Mahwah, NJ: Lawrence Erlbaum Assoc. Publishers, 35–54, 1997.

[23] P. Johnson–Laird and K. Oatley, "The Language of Emotions: An Analysis of a Semantic Field," *Cognition and Emotion*, **3**:81–123, 1989.

[24] D. Graff, R. Rosenfeld, and D. Paul, "CSR-III Text," Linguistic Data Consortium, #LDC95T6, 1995.

[25] J.R. Bellegarda, "Latent Affective Mapping: A Novel Framework for the Data–Driven Analysis of Emotion in Text," in *Proc. Interspeech*, Makuhari, Japan, Sept. 2010.