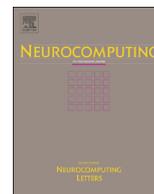




ELSEVIER

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

An ELM-based model for affective analogical reasoning

Erik Cambria^{a,*}, Paolo Gastaldo^b, Federica Bisio^b, Rodolfo Zunino^b^a School of Computer Engineering, Nanyang Technological University, Singapore^b Department of Naval, Electrical, Electronic and Telecommunication Engineering, University of Genoa, Italy

ARTICLE INFO

Article history:

Received 22 July 2013

Received in revised form

11 January 2014

Accepted 19 January 2014

Available online 5 October 2014

Keywords:

Artificial intelligence

Extreme learning machine

Analogical reasoning

Natural language processing

Opinion mining

ABSTRACT

Between the dawn of the Internet through year 2003, there were just a few dozens exabytes of information on the Web. Today, that much information is created weekly. The opportunity to capture the opinions of the general public about social events, political movements, company strategies, marketing campaigns, and product preferences has raised increasing interest both in the scientific community, for the exciting open challenges, and in the business world, for the remarkable fallouts in marketing and financial prediction. Keeping up with the ever-growing amount of unstructured information on the Web, however, is a formidable task and requires fast and efficient models for opinion mining. In this paper, we explore how the high generalization performance, low computational complexity, and fast learning speed of extreme learning machines can be exploited to perform analogical reasoning in a vector space model of affective common-sense knowledge. In particular, by enabling a fast reconfiguration of such a vector space, extreme learning machines allow the polarity associated with natural language concepts to be calculated in a more dynamic and accurate way and, hence, perform better concept-level sentiment analysis.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The ways people express their opinions and sentiments have radically changed in the past few years thanks to the advent of social networks, web communities, blogs, wikis, and other online collaborative media. Actually, the availability of these new tools allow people to create and share, in a time and cost efficient way, their own contents, ideas, and opinions with virtually the millions of people connected to the World Wide Web. This has made available by click a huge source of information and opinions and has provided a powerful communication medium to share knowledge and to get advantage from others' experiences. As a major consequence, the distillation of knowledge from this huge amount of unstructured information can be a key factor for marketers who want to create an image or identity in the minds of their customers for their product, brand, or organization.

On the other hand, these online social data remain hardly accessible to computers, as they are specifically meant for human consumption. Online information retrieval is still mainly based on algorithms relying on the textual representation of web pages. Such algorithms are very good at retrieving texts, splitting them

into parts, checking the spelling, and counting their words. But when it comes to interpreting sentences and extracting useful information for users, their capabilities are still very limited.

Indeed, such scenario has led to the emerging fields of opinion mining and sentiment analysis [1–3], which deal with information retrieval and knowledge discovery from text using data mining and natural language processing (NLP) techniques to distill knowledge and opinions from the huge amount of information on the World Wide Web. Mining opinions and sentiments from natural language, though, is an extremely difficult task as it involves a deep understanding of most of the explicit and implicit, regular and irregular, syntactical and semantic rules proper of a language. Sentic computing [4] tackles these crucial issues by exploiting affective common-sense reasoning, i.e., the intrinsically human capacity to interpret the cognitive and affective information associated with natural language and, hence, to infer new knowledge and make decisions, in connection with one's social and emotional values, sensors, and ideals. Thus, common-sense computing techniques are applied to bridge the semantic gap between word-level natural language data and the concept-level opinions conveyed by these. To achieve this goal, the sentic computing framework takes advantage of AffectNet,¹ a semantic network in which common-sense concepts (e.g., 'read book', 'payment', 'play music') are linked to a hierarchy of affective domain labels

* Corresponding author.

E-mail addresses: cambria@ntu.edu.sg (E. Cambria), paolo.gastaldo@unige.it (P. Gastaldo), federica.bisio@edu.unige.it (F. Bisio), rodolfo.zunino@unige.it (R. Zunino).

¹ <http://sentic.net/affectnet.zip>

(e.g., ‘joy’, ‘amazement’, ‘fear’, ‘admiration’). In particular, the vector space representation of such a semantic network, termed AffectiveSpace,² enables affective analogical reasoning on natural language concepts. In practice, concepts conveying similar semantic and affective information, e.g., ‘enjoy conversation’ and ‘chat with friend’, tend to fall near each other in the multi-dimensional space that characterizes AffectiveSpace.

The present research fits within the sentic computing framework and aims at exploiting machine learning to develop a cognitive model that can effectively support emotion recognition in natural language text. In this regard, the issue addressed in this paper is the design of an emotion categorization architecture that is able to remap any concept represented according to the AffectiveSpace into a suitable space defined by four affective dimensions: Pleasantness, Attention, Sensitivity, and Aptitude. Previous works already published in the literature [4] proved that a categorization model based on these four affective dimensions can potentially synthesize the full range of emotional experiences. Moreover, one can take advantage of such representation based on a four-dimensional sentic vector to effectively support cognitive tasks such as polarity detection and emotion recognition. However, the task of predicting the four-dimensional sentic vector to be associated with a generic concept involves non-linear, complex mechanisms [5]; this in turn means that one cannot address such problem by designing an explicit model. Therefore, in this work extreme learning machine (ELM) is adopted as a powerful tool to tackle this challenging task by exploiting inductive learning methodologies.

The paper shows the proposed ELM-based architecture for emotion categorization compares favorably with other approaches previously published in the literature that addressed the same problem [5]. Indeed, the most interesting outcome of the present research concerns the eventual dimensionality of the AffectiveSpace, which defines the input space to categorization model. The experimental results prove that by adopting ELM-based predictors one can attain consistent performance in terms of emotion categorization while using a 50-dimensional AffectiveSpace, while previous approaches exploited a 100-dimensional space [4]. This in turn demonstrates that the proposed architecture is able to significantly reduce the overall complexity of the categorization model system.

The rest of the paper is organized as follows: Section 2 presents related work in the field of opinion mining and sentiment analysis; Section 3 describes the multi-dimensional vector space model of affective common-sense knowledge; Section 4 illustrates the adopted emotion categorization model; Section 5 presents the ELM-based model for affective analogical reasoning; Section 6 proposes an evaluation of the cognitive-inspired model; finally, Section 7 concludes the paper and suggests directions for future work.

2. Background

Most existing approaches to opinion mining and sentiment analysis rely on the extraction of a vector representing the most salient and important text features, which is later used for classification purposes. Some of the most commonly used features are term frequency and presence. The latter is a binary-valued feature vectors in which the entries merely indicate whether a term occurs (value 1) or not (value 0) formed a more effective basis for review polarity classification. This is indicative of an interesting difference between typical topic-based text

categorization and polarity classification. While a topic is more likely to be emphasized by frequent occurrences of certain keywords, overall sentiment may not usually be highlighted through repeated use of the same terms. Other term-based features are often added to the features vector. Position is one of these, in consideration of how the position of a token in a text unit can affect the way in which the token affects the sentiment of the text. Also presence n-grams, typically bi-grams and tri-grams, are often taken into account as useful features. Some methods also rely on the distance between terms. Part-of-speech (POS) information (nouns, adjectives, adverbs, verbs, etc.) is also commonly exploited in general textual analysis as a basic form of word-sense disambiguation. Certain adjectives, in particular, have been proved to be good indicators of sentiment and sometimes have been used to guide feature selection for sentiment classification. In other works, finally, the detection of sentiments was performed through selected phrases, which were chosen via a number of pre-specified POS patterns, most including an adjective or an adverb. Several other approaches have been developed for the general task of mapping a given piece of text to a label belonging to a predefined set of categories, or to a real number representative of a polarity degree. However, such approaches and their performance are strictly bound to the considered domain of application and to the related topics. Moreover, most of the literature on sentiment analysis has focused on text written in English and, consequently, most resources developed, e.g., sentiment lexicons, are in English. Adapting such resources to other languages can be considered as a domain adaptation problem.

The evolution of sentiment analysis research can be studied in terms of the different tokens, or building blocks, of the analysis, and the implicit information associated with them. In this sense, existing approaches can be grouped into four main categories: keyword spotting, lexical affinity, statistical methods, and concept-based techniques. Keyword spotting is the most naïve approach and probably also the most popular because of its accessibility and economy. Text is classified into affect categories based on the presence of fairly unambiguous affect words like ‘happy’, ‘sad’, ‘afraid’, and ‘bored’. Elliott’s Affective Reasoner [6], for example, watches for 198 affect keywords, e.g., ‘distressed’, ‘enraged’, plus affect intensity modifiers, e.g., ‘extremely’, ‘somewhat’, ‘mildly’, plus a handful of cue phrases, e.g., ‘did that’, ‘wanted to’. Other popular sources of affect words are Ortony’s Affective Lexicon [7], which groups terms into affective categories, and Wiebe’s linguistic annotation scheme [8]. The weaknesses of this approach lie in two areas: poor recognition of affect when negation is involved and reliance on surface features. In relation to its first weakness, while the approach can correctly classify the sentence “today was a happy day” as being happy, it is likely to fail on a sentence like “today was not a happy day at all”. Regarding its second weakness, the approach relies on the presence of obvious affect words, which are only surface features of the prose. In practice, a lot of sentences convey affect through underlying meaning rather than affect adjectives. For example, the text “My husband just filed for divorce and he wants to take custody of my children away from me” certainly evokes strong emotions, but uses no affect keywords, and therefore, cannot be classified using a keyword spotting approach.

Lexical affinity is slightly more sophisticated than keyword spotting as rather than simply detecting obvious affect words, it assigns arbitrary words a probabilistic ‘affinity’ for a particular emotion. For example, ‘accident’ might be assigned a 75% probability of indicating a negative affect, as in ‘car accident’ or ‘hurt by accident’. These probabilities are usually learnt using linguistic corpora [9–11]. Though often outperforming pure keyword spotting, there are two main problems with the approach. First, lexical affinity, operating solely on the word-level, can easily be tricked by sentences like “I avoided an accident” (negation) and “I met my

² <http://sentic.net/affectivespace.zip>

girlfriend by accident” (other word senses). Second, lexical affinity probabilities are often biased towards text of a particular genre, dictated by the source of the linguistic corpora. This makes it difficult to develop a reusable, domain-independent model.

Statistical methods, such as Bayesian inference and support vector machine (SVM), have been popular for affect classification of texts and have been used by researchers on projects such as Pang’s movie review classifier [12] and many others [13–16]. By feeding a machine learning algorithm a large training corpus of affectively annotated texts, it is possible for the system to not only learn the affective valence of affect keywords (as in the keyword spotting approach), but also to take into account the valence of other arbitrary keywords (like lexical affinity), punctuation, and word co-occurrence frequencies. However, statistical methods are generally semantically weak, meaning, with the exception of obvious affect keywords, other lexical or co-occurrence elements in a statistical model have little predictive value individually. As a result, statistical text classifiers only work with acceptable accuracy when given a sufficiently large text input. So, while these methods may be able to affectively classify user’s text on the page- or paragraph-level, they do not work well on smaller text units such as sentences or clauses.

Concept-based approaches [17–20], in turn, focus on a semantic analysis of text through the use of web ontologies or semantic networks, which allow the handling of conceptual and affective information associated with natural language opinions. By relying on large semantic knowledge bases, such approaches step away from blind use of keywords and word co-occurrence counts, but rather rely on the implicit meaning/features associated with natural language concepts. Unlike purely syntactical techniques, concept-based approaches are also able to detect sentiments that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey any emotion, but are implicitly linked to other concepts that do so. Moreover, concept-level sentiment analysis enables comparative fine-grained feature-based sentiment analysis. Rather than gathering isolated opinions about a whole item (e.g., iPhone5), users are generally more interested in comparing different products according to their specific features (e.g., iPhone5’s vs Galaxy S3’s touchscreen), or even sub-features (e.g., fragility of iPhone5’s vs Galaxy S3’s touchscreen). In this context, the construction of comprehensive common and common-sense knowledge bases is key for feature-spotting and polarity detection, respectively. Such common-sense bases, in particular, allow natural language text to be properly deconstructed into sentiments– for example, to appraise the concept ‘long battery life’ as positive for a laptop review and ‘long queue’ as negative for a post office, or the concept ‘go read the book’ as positive for a book review but negative for a movie review.

3. The vector space model

The best way to solve a problem is to already know a solution for it. But, if we have to face a problem we have never encountered before, we need to use our intuition. Intuition can be explained as the process of making analogies between the current problem and the ones solved in the past to find a suitable solution. Marvin Minsky attributes this property to the so called ‘difference-engines’ [21]. This particular kind of an agent operates by recognizing differences between the current state and the desired state, and acting to reduce each difference by invoking K-lines that turn on suitable solution methods. This kind of thinking may be the essence of our supreme intelligence since in everyday life no two situations are ever the same and we have to perform this action continuously.

Human mind constructs intelligible meanings by continuously compressing over vital relations [22]. The compression principles aim to transform diffuse and distended conceptual structures to more focused versions so as to become more congenial for human understanding. To this end, principal component analysis (PCA) has been applied on the matrix representation of AffectNet. In particular, truncated singular value decomposition (TSVD) has been preferred to other dimensionality reduction techniques for its simplicity, relatively low computational cost, and compactness. TSVD is particularly suitable for measuring cross-correlations between affective common-sense concepts as it uses an orthogonal transformation to convert the set of possibly correlated common-sense features associated with each concept into a set of values of uncorrelated variables (the principal components of the SVD).

By using Lanczos’ method [23], the generalization process is relatively fast (a few seconds), despite the size and the sparseness of AffectNet. Applying TSVD on AffectNet causes it to describe other features that could apply to known affective concepts by analogy: if a concept in the matrix has no value specified for a feature owned by many similar concepts, then by analogy the concept is likely to have that feature as well. In other words, concepts and features that point in similar directions and, therefore, have high dot products, are good candidates for analogies.

After performing TSVD on AffectNet, hereby termed A for the sake of conciseness, a low-rank approximation of it is obtained, that is, a new matrix $\tilde{A} = U_M \Sigma_M V_M^T$. This approximation is based on minimizing the Frobenius norm of the difference between A and \tilde{A} under the constraint $rank(\tilde{A}) = M$. For the Eckart–Young theorem [24], it represents the best approximation of A in the least-square sense, in fact:

$$\min_{\tilde{A}|rank(\tilde{A})=M} |A - \tilde{A}| = \min_{\tilde{A}|rank(\tilde{A})=M} |\Sigma - U^* \tilde{A} V| = \min_{\tilde{A}|rank(\tilde{A})=M} |\Sigma - S| \quad (1)$$

assuming that \tilde{A} has the form $\tilde{A} = USV^*$, where S is diagonal. From the rank constraint, i.e., S has M non-zero diagonal entries, the minimum of the above statement is obtained as follows:

$$\min_{\tilde{A}|rank(\tilde{A})=M} |\Sigma - S| = \min_{s_i} \sqrt{\sum_{i=1}^n (\sigma_i - s_i)^2} \quad (2)$$

$$\min_{s_i} \sqrt{\sum_{i=1}^n (\sigma_i - s_i)^2} = \min_{s_i} \sqrt{\sum_{i=1}^M (\sigma_i - s_i)^2 + \sum_{i=M+1}^n \sigma_i^2} = \sqrt{\sum_{i=M+1}^n \sigma_i^2} \quad (3)$$

Therefore, \tilde{A} of rank M is the best approximation of A in the Frobenius norm sense when $\sigma_i = s_i (i = 1, \dots, M)$ and the corresponding singular vectors are the same as those of A . If all but the first M principal components are discarded, common-sense concepts and emotions are represented by vectors of M coordinates. These coordinates can be seen as describing concepts in terms of ‘eigenmoods’ that form the axes of AffectiveSpace, i.e., the basis e_0, \dots, e_{M-1} of the vector space (Fig. 1). For example, the most significant eigenmood, e_0 , represents concepts with positive affective valence. That is, the larger a concept’s component in the e_0 direction is, the more affectively positive it is likely to be. Concepts with negative e_0 components, then, are likely to have negative affective valence. Thus, by exploiting the information sharing property of TSVD, concepts with the same affective valence are likely to have similar features – that is, concepts conveying the same emotion tend to fall near each other in AffectiveSpace. Concept similarity does not depend on their absolute positions in the vector space, but rather on the angle they make with the origin. For example, concepts such as ‘beautiful day’, ‘birthday party’, and ‘make person happy’ are found very close in direction in the vector space, while concepts like ‘feel guilty’, ‘be laid off’,

the function $G(x) = -1/\sigma\sqrt{2\pi}e^{-x^2/2\sigma^2}$, for its symmetric inverted bell curve shape that quickly rises up towards the unit value (Fig. 2). In particular, the function models how valence or intensity of an affective dimension varies according to different values of arousal or activation (x), spanning from null value (emotional void) to the unit value (heightened emotionality). Justification for assuming that the Gaussian function (rather than a step or simple linear function) is appropriate for modeling the variation of emotion intensity is based on research into the neural and behavioral correlates of emotion, which are assumed to indicate emotional intensity in some sense. Nobody genuinely knows what function subjective emotion intensity follows, because it has never been truly or directly measured [27].

For example, the so-called Duchenne smile (a genuine smile indicating pleasure) is characterized by smooth onset, increasing to an apex, and a smooth, relatively lengthy offset [28]. More generally, Klaus Scherer has argued that emotion is a process characterized by non-linear relations among its component elements - especially physiological measures, which typically look Gaussian [29]. Emotions, in fact, are not linear [25]: the stronger the emotion, the easier it is to be aware of it. Mapping this space of possible emotions leads to a hourglass shape (Fig. 3). It is worth to note that, in the model, the state of ‘emotional void’ is a-dimensional, which contributes to determine the hourglass shape. Total absence of emotion can be associated with the total absence of reasoning (or, at least, consciousness) [30], which is not an envisaged mental state as, in human mind, there is never nothing going on. Each affective dimension of the Hourglass model is characterized by six levels of activation (measuring the strength of an emotion), termed ‘sentic levels’, which represent the intensity thresholds of the expressed or perceived emotion.

These levels are also labeled as a set of 24 basic emotions [25], six for each of the affective dimensions, in a way that allows the model to specify the affective information associated with text

both in a dimensional and a discrete form (Table 1). The dimensional form, in particular, is termed ‘sentic vector’ and it is a four-dimensional float vector that can potentially synthesize the full range of emotional experiences in terms of Pleasantness, Attention, Sensitivity, and Aptitude. In the model, the vertical dimension represents the intensity of the different affective dimensions, i.e., their level of activation, while the radial dimension represents K-lines [21] that can activate configurations of the mind, which can either last just a few seconds or years. The model follows the pattern used in color theory and research in order to obtain judgments about combinations, i.e., the emotions that result when two or more fundamental emotions are combined, in the same way that red and blue make purple.

Hence, some particular sets of sentic vectors have special names, as they specify well-known compound emotions (Fig. 4). For example, the set of sentic vectors with a level of Pleasantness $\in [G(2/3), G(1/3)]$, i.e., joy, a level of Aptitude $\in [G(2/3), G(1/3)]$, i.e., trust, and a minor magnitude of Attention and Sensitivity, are termed ‘love sentic vectors’ since they specify the compound emotion of love (Table 2). More complex emotions can be synthesized by using three, or even

Table 1

The sentic levels of the Hourglass model. Labels are organized into four affective dimensions with six different levels each, whose combined activity constitutes the ‘total state’ of the mind.

Interval	Pleasantness	Attention	Sensitivity	Aptitude
$[G(1), G(2/3)]$	Ecstasy	Vigilance	Rage	Admiration
$[G(2/3), G(1/3)]$	Joy	Anticipation	Anger	Trust
$[G(1/3), G(0)]$	Serenity	Interest	Annoyance	Acceptance
$(G(0), -G(1/3)]$	Pensiveness	Distraction	Apprehension	Boredom
$(-G(1/3), -G(2/3))]$	Sadness	Surprise	Fear	Disgust
$(-G(2/3), -G(1))]$	Grief	Amazement	Terror	Loathing

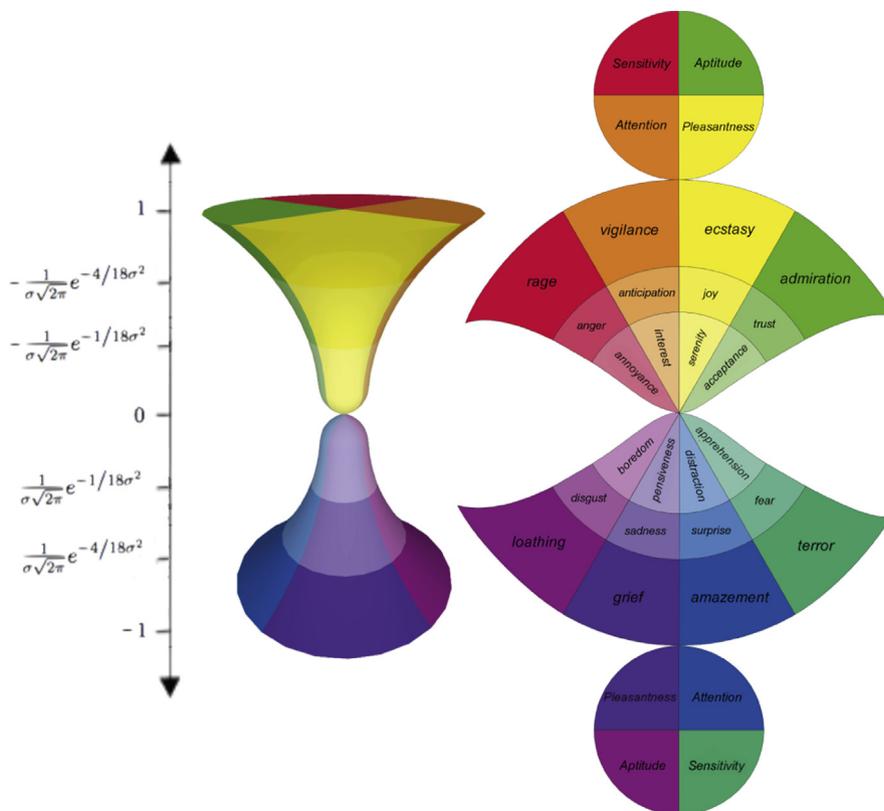


Fig. 3. The 3D model and the net of the Hourglass of Emotions. Since affective states go from strongly positive to null to strongly negative, the model assumes a hourglass shape.

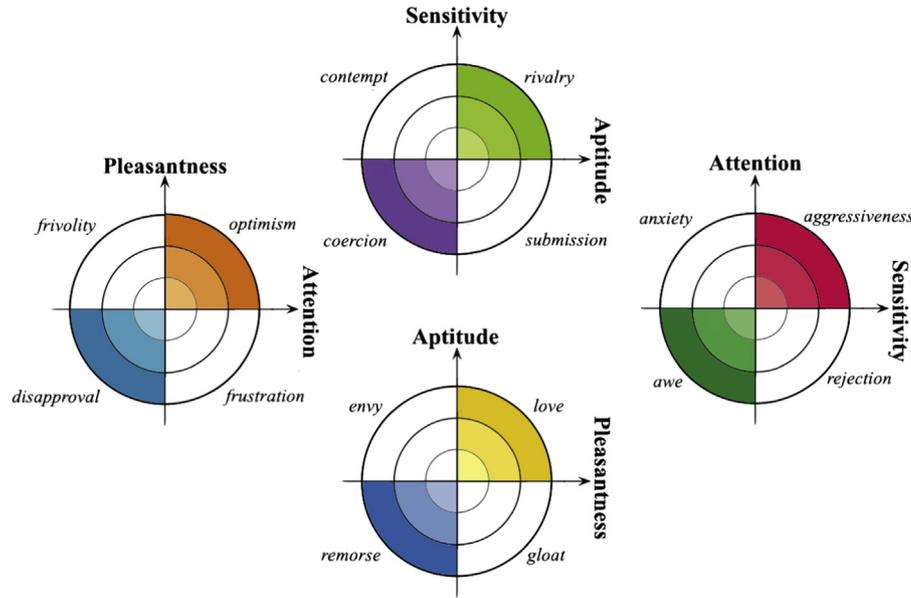


Fig. 4. Hourglass compound emotions of second level. By combining basic emotions pairwise, it is possible to obtain complex emotions.

four, sentic levels, e.g., joy + trust + anger=jealousy. Therefore, analogous to the way primary colors combine to generate different color gradations (and even colors we do not have a name for), the primary emotions of the Hourglass model can blend to form the full spectrum of human emotional experience. Beyond emotion detection, the Hourglass model is also used for polarity detection tasks. Since polarity is strongly connected to attitudes and feelings, it is defined in terms of the four affective dimensions, according to the formula:

$$p = \frac{\sum_{i=1}^N \text{Pleasantness}(c_i) + |\text{Attention}(c_i)| - |\text{Sensitivity}(c_i)| + \text{Aptitude}(c_i)}{3N} \quad (4)$$

where c_i is an input concept, N the total number of concepts, and 3 the normalization factor (as the Hourglass dimensions are defined as float $\in [-1, +1]$). In the formula, Attention is taken as absolute value since both its positive and negative intensity values correspond to positive polarity values (e.g., 'surprise' is negative in the sense of lack of Attention, but positive from a polarity point of view). Similarly, Sensitivity is taken as negative absolute value since both its positive and negative intensity values correspond to negative polarity values (e.g., 'anger' is positive in the sense of level of activation of Sensitivity, but negative in terms of polarity). Besides practical reasons, the formula is important because it shows a clear connection between polarity (opinion mining) and emotions (sentiment analysis).

5. ELM-based affective analogical reasoning

Affective analogical reasoning consists in processing the cognitive and affective information associated with natural language concepts, in order to compare the similarities between new and understood concepts and, hence, use such similarities to gain understanding of the new concept. It is a form of inductive reasoning because it strives to provide understanding of what is likely to be true, rather than deductively proving something as fact. The reasoning process begins by determining the target concept to be learned or explained. It is then compared to a general matching concept whose semantics and sentics (that is, the conceptual and affective information associated with it) are already well-understood. The two concepts must be similar

Table 2

The second-level emotions generated by pairwise combination of the sentic levels of the Hourglass model. The co-activation of different levels gives birth to different compound emotions.

	Attention > 0	Attention < 0	Aptitude > 0	Aptitude < 0
Pleasantness > 0	Optimism	Frivolity	Love	Gloat
Pleasantness < 0	Frustration	Disapproval	Envy	Remorse
Sensitivity > 0	Aggressiveness	Rejection	Rivalry	Contempt
Sensitivity < 0	Anxiety	Awe	Submission	Coercion

enough to make a valid, substantial comparison. Affective analogical reasoning is based on the brain's ability to form semantic patterns by association. The brain may be able to understand new concepts more easily if they are perceived as being part of a semantic pattern. If a new concept is compared to something the brain already knows, it may be more likely that the brain will store the new information more readily.

Such a semantic association needs *high generalization performance*, in order to better match conceptual and affective patterns. Because of the dynamic nature of AffectiveSpace, moreover, affective analogical reasoning should be characterized by *fast learning speed*, in order for concept associations to be recalculated every time a new multi-word expression is inserted in AffectNet. Finally, the process should be of *low computational complexity*, in order to perform big social data analysis [31]. All such features are those typical of ELM, a machine learning technique that, in recent years, has proved to be a powerful tool to tackle challenging modeling problems [32–39].

5.1. Extreme learning machine

The ELM approach [40] was introduced to overcome some issues in back-propagation network [41] training, specifically, potentially slow convergence rates, the critical tuning of optimization parameters [42], and the presence of local minima that call for multi-start and re-training strategies. The ELM learning problem settings require a training set, X , of N labeled pairs, (\mathbf{x}_i, y_i) , where $\mathbf{x}_i \in \mathcal{R}^m$ is the i -th input vector and $y_i \in \mathcal{R}$ is the associate expected 'target' value; using a scalar output implies that the network has one output unit, without loss of generality.

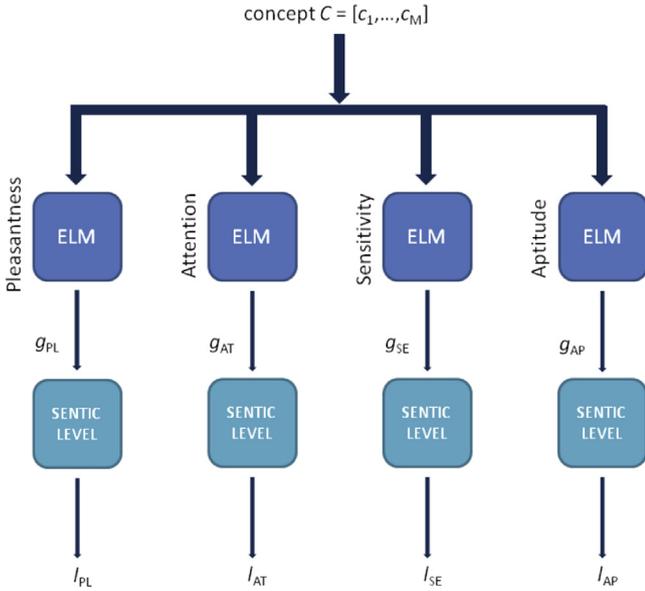


Fig. 5. The ELM-based framework for describing common-sense concepts in terms of the four Hourglass model's dimensions.

The input layer has m neurons and connects to the 'hidden' layer (having N_h neurons) through a set of weights $\{\hat{\mathbf{w}}_j \in \mathcal{R}^m; j = 1, \dots, N_h\}$. The j -th hidden neuron embeds a bias term, \hat{b}_j , and a nonlinear 'activation' function, $\varphi(\cdot)$; thus the neuron's response to an input stimulus, \mathbf{x} , is

$$a_j(\mathbf{x}) = \varphi(\hat{\mathbf{w}}_j \cdot \mathbf{x} + \hat{b}_j) \quad (5)$$

Note that (5) can be further generalized to a wider class of functions [43] but for the subsequent analysis this aspect is not relevant. A vector of weighted links, $\bar{\mathbf{w}}_j \in \mathcal{R}^{N_h}$, connects hidden neurons to the output neuron without any bias [44]. The overall output function, $f(\mathbf{x})$, of the network is

$$f(\mathbf{x}) = \sum_{j=1}^{N_h} \bar{\mathbf{w}}_j a_j(\mathbf{x}) \quad (6)$$

It is convenient to define an 'activation matrix', \mathbf{H} , such that the entry $\{h_{ij} \in \mathbf{H}; i = 1, \dots, N; j = 1, \dots, N_h\}$ is the activation value of the j -th hidden neuron for the i -th input pattern. The \mathbf{H} matrix is:

$$\mathbf{H} \equiv \begin{bmatrix} \varphi(\hat{\mathbf{w}}_1 \cdot \mathbf{x}_1 + \hat{b}_1) & \cdots & \varphi(\hat{\mathbf{w}}_{N_h} \cdot \mathbf{x}_1 + \hat{b}_{N_h}) \\ \vdots & \ddots & \vdots \\ \varphi(\hat{\mathbf{w}}_1 \cdot \mathbf{x}_N + \hat{b}_1) & \cdots & \varphi(\hat{\mathbf{w}}_{N_h} \cdot \mathbf{x}_N + \hat{b}_{N_h}) \end{bmatrix} \quad (7)$$

In the ELM model, the quantities $\{\hat{\mathbf{w}}_j, \hat{b}_j\}$ in (5) are set randomly and are not subject to any adjustment, and the quantity $\bar{\mathbf{w}}_j$ in (6) is the only degree of freedom. The training problem reduces to the minimization of the convex cost:

$$\min_{(\bar{\mathbf{w}}, \hat{\mathbf{b}})} \|\mathbf{H}\bar{\mathbf{w}} - \mathbf{y}\|^2 \quad (8)$$

A matrix pseudo-inversion yields the unique L_2 solution, as proven in [40]:

$$\bar{\mathbf{w}} = \mathbf{H}^+ \mathbf{y} \quad (9)$$

The simple, efficient procedure to train an ELM therefore involves the following steps:

1. Randomly set the input weights $\hat{\mathbf{w}}_i$ and bias \hat{b}_i for each hidden neuron.
2. Compute the activation matrix, \mathbf{H} , as per (7).

3. Compute the output weights by solving a pseudo-inverse problem as per (9).

Despite the apparent simplicity of the ELM approach, the crucial result is that even random weights in the hidden layer endow a network with a notable representation ability [40]. Moreover, the theory derived in [45] proves that regularization strategies can further improve its generalization performance. As a result, the cost function (8) is augmented by an L_2 regularization factor as follows:

$$\min_{\bar{\mathbf{w}}} \{\|\mathbf{H}\bar{\mathbf{w}} - \mathbf{y}\|^2 + \lambda \|\bar{\mathbf{w}}\|^2\} \quad (10)$$

5.2. The emotion categorization framework

The proposed framework is designed to receive as input a natural language concept represented according to an M -dimensional space and to predict the corresponding sentic levels for the four affective dimensions involved: Pleasantness, Attention, Sensitivity, and Aptitude. The dimensionality M of the input space stems from the specific design of AffectiveSpace. As for the outputs, in principle each affective dimension can be characterized by an analog value in the range $[-1, 1]$, which represents the intensity of the expressed or received emotion. Indeed, as discussed in Section 4, those analog values are eventually remapped to obtain six different sentic levels for each affective dimension.

The research presented in this paper spans each affective dimension separately, under the reasonable assumption that the various dimensions map perceptual phenomena that are mutually independent [4]. As a result, each affective dimension is handled by a dedicated ELM, which addresses a regression problem. Thus, each ELM-based predictor is fed by the M -dimensional vector describing the concept and yields as output the analog value that would eventually lead to the corresponding sentic level. Fig. 5 provides the overall scheme of the framework; here, g_x is the level of activation predicted by the ELM and l_x is the corresponding sentic level.

In theory, one might also implement the framework showed in Fig. 5 by using four independent predictors based on a multi-class classification schema. In such a case, each predictor would directly yield as output a sentic level out of the six available. However, two important aspects should be taken into consideration. First, the design of a reliable multi-class predictor is not straightforward, especially when considering that several alternative schemata have been proposed in the literature without a clearly established solution. Second, the emotion categorization scheme based on sentic levels stem from an inherently analog model, i.e., the Hourglass of Emotions. This ultimately motivates the choice of designing the four prediction systems as regression problems.

In fact, the framework schematized in Fig. 5 represents an intermediate step in the development of the final emotion categorization system. One should take into account that every affective dimension can in practice take on seven different values: the six available sentic levels plus a 'neutral' value, which in theory correspond to the value $G(0)$ in the emotion categorization model discussed in Section 4. In practice, though, the neutral level is assigned to those concepts that are characterized by a level activation that lies in an interval around $G(0)$ in that affective dimension. Therefore, the final framework should properly manage the eventual seven-level scale. To this end, the complete categorization system is set to include a module that is able to predict if an affective dimension is present or absent in the description of a concept. In the latter case, no sentic level should be associated with that affective dimension (i.e., $l_x = \text{null}$). In the present paper, this task is addressed by exploiting the hierarchical

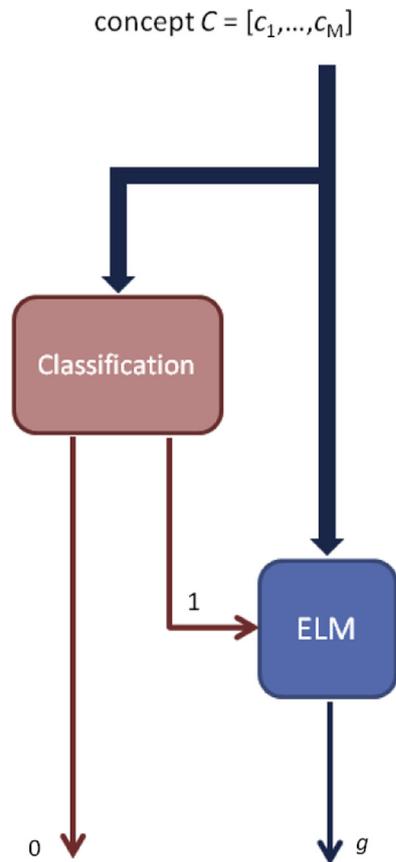


Fig. 6. The hierarchical scheme in which an SVM-based classifier first filters out unemotional concepts and an ELM-based predictor then classifies emotional concepts in terms of the involved affective dimension.

approach presented in Fig. 6. Hence, given a concept and an affective dimension, first a SVM-based binary classifier is entitled to decide if a sentic level should be assessed. Accordingly, the ELM-based predictor is asked to assess the level of activation only if the SVM-based classifier determines that a sentic level should be associated with that concept. Otherwise, it is assumed that the neutral level should be associated with that concept (i.e., the corresponding affective dimension is not involved in the description of that concept). Obviously, such structure is replicated for each affective dimension. Fig. 7 schematizes the complete framework.

6. Experimental results

The proposed emotion categorization framework has been tested both on a benchmark of 6,813 common-sense concepts and on the real-world dataset of 2000 patient opinions used in [5]. As for the benchmark, the publicly available Sentic API³ was used to obtain for each concept the corresponding sentic vector, i.e., the level of activation of each affective dimension. According to the emotion categorization model presented in Section 4, the Sentic API expresses the level of activation as an analog number in the range $[-1, 1]$, which are eventually mapped into sentic levels by adopting the Gaussian mapping function. Indeed, the neutral sentic level is codified by the value '0'. The format adopted by the Sentic API to represent the levels of activation actually prevents one to approach the prediction problem as an authentic

regression task, as per Fig. 5. The neutral sentic level corresponds to a single value in the analog range used to represent activations.

Therefore, experimental results are presented as follows: firstly, the performance of the system depicted in Fig. 5 is analyzed (according to that set up, the ELM-based predictors are not designed to assess the neutral sentic level); secondly, the performance of the complete framework (Fig. 7) is discussed; lastly, a use-case evaluation on the patient opinion dataset is proposed.

6.1. Accuracy in the prediction of the sentic levels

The emotion categorization framework proposed in Fig. 5 exploits four independent ELM-based predictors to estimate the levels of activation of as many affective dimensions. In this experiment, it is assumed that each ELM-based predictor can always assess correctly a level of activation set to '0'. A cross-validation procedure has been used to robustly evaluate the performance of the framework. As a result, the experimental session involved ten different experimental runs. In each run, 800 concepts randomly extracted from the complete benchmark provided the test set; the remaining concepts were heavily split into a training set and a validation set. The validation set was designed to support the model selection phase, i.e., the selection of the best parameterization for the ELM predictors. In the present configuration, two quantities were involved in the model selection phase: the number of neurons N_h in the hidden layer and the regularization parameter λ . The following parameters were used for model selection:

- $N_h \in [100, 1000]$ by steps of 100 neurons;
- $\lambda = \{1 \times 10^{-6}, 1 \times 10^{-5}, 1 \times 10^{-4}, 1 \times 10^{-3}, 1 \times 10^{-2}, 1 \times 10^{-1}, 1\}$.

In each run the performance of the emotion categorization framework was measured by using only the patterns included in the test set, i.e., the patterns that were not involved in the training phase or in the model selection phase. Table 3 reports the performance obtained by the emotion categorization framework over the ten runs. The table actually compares the results of three different sets up, which differs in the dimensionality M of AffectiveSpace that describe the concepts. Thus, Table 3 provides the results achieved with $M=100$, $M=70$, and $M=50$. The results refer to a configuration of the ELM predictors characterized by the following parameterization: $N_h=200$ and $\lambda=1$; such configuration was obtained by exploiting the model selection phase. The performance of each setting is evaluated according to the following quantities (expressed as average values over the ten runs):

- *Pearson's correlation coefficient*: the measure of the linear correlation between predicted levels of activation and expected levels of activation for the four predictors.
- *Strict accuracy*: the percentage of patterns for which the framework correctly predicted the four sentic levels; thus, a concept is assumed to be correctly classified only if the predicted sentic level corresponds to the expected sentic level for every affective dimension.
- *Smooth accuracy*: the percentage of patterns for which the framework correctly predicted three sentic levels out of four; thus, a concept is assumed to be correctly classified even when one among the four predictors fails to assign the correct sentic level.
- *Relaxed accuracy*: in this case, one relaxes the definition of correct prediction of the sentic level. As a result, given an affective dimension, the prediction is assumed correct even when the assessed sentic level and the expected sentic level are contiguous in Table 1. As an example, let suppose that the

³ <http://sentic.net/api>

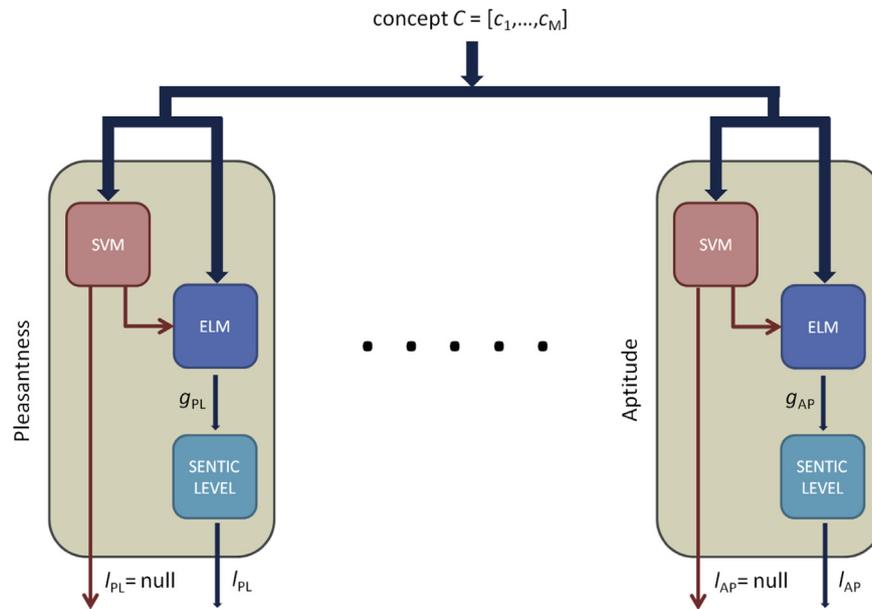


Fig. 7. The final framework: a hierarchical scheme is adopted to classify emotional concepts in terms of Pleasantness, Attention, Sensitivity, and Aptitude.

Table 3

Performance obtained by the emotion categorization framework over the ten runs with three different sets up of AffectiveSpace's dimensionality.

M	Correlation				Accuracy		
	Pleasantness	Attention	Sensitivity	Aptitude	Strict	Smooth	Relaxed
100	0.69	0.67	0.78	0.72	39.4	73.4	87.0
70	0.71	0.67	0.78	0.72	41.0	75.4	88.4
50	0.66	0.66	0.77	0.71	40.9	75.3	86.4

expected sentic level in the affective dimension Sensitivity for the incoming concept is 'annoyance'. Then, the prediction is assumed correct even when the assessed sentic level is 'anger' or 'apprehension'. Therefore, the relaxed accuracy gives the percentage of patterns for which the framework correctly predicted the four sentic levels according to such criterion.

In practice, the smooth accuracy and the relaxed accuracy allow one to take into account two crucial issues: the dataset can include noise and entries may incorporate a certain degree of subjectiveness. The results provided in Table 3 lead to the following comments:

- Emotion categorization is in fact a challenging problem; in this regard, the gap between strict accuracy and smooth/relaxed accuracies confirms that the presence of noise is a crucial issue.
- The ELM-based framework can attain satisfactory performance in terms of smooth accuracy and relaxed accuracy. Actually, the proposed framework scored a 75% accuracy in correctly assessing at least three affective dimension for an input concept.
- Reliable performance can be achieved even when a 50-dimensional AffectiveSpace is used to characterize concepts. The latter result indeed represents a very interesting outcome, as previous approaches to the same problem in general exploited a 100-dimensional AffectiveSpace [4,5]. In this respect, the present work shows that the use of ELM-based predictors can reduce the overall complexity of the framework by shrinking the feature space.

6.2. Accuracy of the complete emotion categorization system

As anticipated above, the complete categorization system exploits the hierarchical approach presented in Fig. 6 to assess the level of activation of a concept. According to such set up, the accuracy of the SVM-based classifier is critical to the whole system's performance, as it handles the preliminary filtering task before that actual sentic description is evaluated. In principle, one might analyze the performance of the two components separately and assess the run-time generalization accuracy accordingly. Nevertheless, in the present context, the system performance has been measured as a whole, irrespectively of the internal structure of the evaluation scheme. On the other hand, one should also consider that, given a concept and a sentic dimension in which such concept should be assessed as neutral, to predict a low activation value is definitely less critical than predicting a large activation value. Therefore, the system performance has been evaluated by avoiding considering as an error the cases in which the expected sentic level is 'neutral' and the assessed sentic level is the less intense (either positive or negative). As an example, given the sentic dimension Attention, to classify a neutral sentic level either as 'interest' or 'distraction' would not be considered an error.

The performance of the framework has been evaluated by exploiting the same cross-validation approach already applied in the previous experimental session. In the present case, though, the model selection approach involved both the SVM-based classifiers and the ELM-based predictors. For the SVM classifiers, two

Table 4

Performance obtained by the overall framework over the ten runs with a 50-dimensional AffectiveSpace.

M	Accuracy		
	Strict	Smooth	Relaxed
50	38.3	72	79.8

quantities were set with model selection: the regularization parameter C and the width σ of the Gaussian kernel. The following parameters were used for model selection:

- $C = \{1, 10, 100, 1000\}$;
- $\sigma = \{0.1, 0.25, 0.5, 0.75, 1, 1.5, 2, 5, 10\}$.

Table 4 reports the performance obtained by the framework over the ten runs. In this case, the experimental session involved only the set up $M=50$, which already proved to attain a satisfactory trade-off between accuracy and complexity. The table provides the average value over the ten runs of the following quantities: strict accuracy, smooth accuracy, relaxed accuracy. The results refer to a configuration of the SVM classifiers characterized by the following parameterization: $C=1$ and $\sigma=1.5$; As expected, the accuracy of the complete framework is slightly inferior to that of the system presented in Section 6.1. Indeed, the results confirm that the proposed approach can attain satisfactory accuracies by exploiting a 50-dimensional AffectiveSpace. In this regard, one should also notice that the estimated performance of the proposed methodology appears quite robust, as it is estimated on ten independent runs involving different compositions of the training and the test set.

6.3. Use-case evaluation

In order to test the performance of the proposed approach in a real-world environment, the ELM-based framework was embedded into an opinion mining engine [4] for the inference of cognitive and affective information associated with natural language. Such an engine consists of four main components: a pre-processing module, which performs a first skim of text; a semantic parser, whose aim is to extract concepts from the opinionated text; a target spotting module, which identifies opinion targets; an affect interpreter, for emotion recognition and polarity detection.

The pre-processing module firstly interprets all the affective valence indicators usually contained in opinionated text such as special punctuation, complete upper-case words, cross-linguistic onomatopoeias, exclamation words, negations, degree adverbs and emoticons. Secondly, it converts text to lower-case and, after lemmatizing it, splits the opinion into single clauses according to grammatical conjunctions and punctuation.

Then, the semantic parser deconstructs text into concepts using a lexicon based on sequences of lexemes that represent multiple-word concepts extracted from the Sentic API. These n-grams are not used blindly as fixed word patterns but exploited as reference for the module, in order to extract multiple-word concepts from information-rich sentences. So, unlike other shallow parsers, the module can recognize complex concepts also when irregular verbs are used or when these are interspersed with adjective and adverbs, e.g., the concept 'buy christmas present' in the sentence "I bought a lot of very nice Christmas gifts". The semantic parser, additionally, provides, for each retrieved concept, the relative frequency, valence and status, that is the concept's occurrence in the text, its positive or negative connotation and the degree of intensity with which the concept is expressed. For each clause, the

Table 5

Structured output example of opinion mining engine.

Opinion target	Category	Moods	Polarity
'iphone4'	'phones', 'electronics'	'ecstasy', 'interest'	+0.71
'speaker'	'electronics', 'music'	'annoyance'	-0.34
'touchscreen'	'electronics'	'ecstasy', 'anticipation'	+0.82
'camera'	'photography', 'electronics'	'acceptance'	+0.56

module outputs a small bag of concepts (SBoC), which is later on analyzed separately by the target spotting module and the affect interpreter to infer the cognitive and affective information associated with the input text, respectively.

The target spotting module aims to individuate one or more opinion targets, such as people, places, events and ideas, from the input concepts. This is done by projecting the concepts of each SBoC into the graph representation of AffectNet, in order to assign these to a specific conceptual class. The categorization does not consist in simply labeling each concept but also in assigning a confidence score to each category label, which is directly proportional to the value of belonging to a specific conceptual cluster (number of steps in the AffectNet graph). The affect interpreter, in turn, projects the concepts of each SBoC into AffectiveSpace and feeds their coordinates to the four ELM-based predictors, in order to assign such concepts to a specific affective class and, hence, calculate polarity in terms of the Hourglass dimensions, as specified in formula (4).

As an example of how the opinion mining engine works, intermediate and final outputs obtained when a natural language opinion is given as input to the system can be examined. The tweet "I think iPhone4 is the top of the heap! OK, the speaker is not the best i hv ever seen bt touchscreen really puts me on cloud 9... camera looks pretty good too!" is selected. After the pre-processing and semantic parsing operations, the following SBoCs are obtained:

SBoC#1:

< Concept : 'think' >
 < Concept : 'iphone4' >
 < Concept : 'topheap' >

SBoC#2:

< Concept : 'ok' >
 < Concept : 'speaker' >
 < Concept : '!good' ++ >
 < Concept : 'see' >

SBoC#3:

< Concept : 'touchscreen' >
 < Concept : 'putcloudnine' ++ >

SBoC#4:

< Concept : 'camera' >
 < Concept : 'lookgood' - >

These are then concurrently processed by the target spotting module and the affect interpreter, which detect the opinion targets and output, for each of them, the relative affective information both in a discrete way, with one or more emotional labels, and in a dimensional way, with a polarity value $\in [-1, +1]$ (as shown in Table 5). In order to evaluate the resulting opinion mining engine, the patient opinion database used in [5] is adopted, and results obtained using k-NN, k-medoids, and the two single layer feedforward networks (SLFNs) developed in [5] (DNN and CNN, which use a 100-dimensional AffectiveSpace) are compared with those obtained using the ELM-based predictors exploiting a 50-dimensional AffectiveSpace.

Table 6
F-measure values relative to PatientOpinion evaluation.

	k-NN (%)	k-medoids (%)	DNN (%)	CNN (%)	ELM (%)
Clinical service	70.1	72.9	77.2	81.8	82.7
Communication	69.8	75.3	75.5	78.1	78.1
Food	79.4	79.6	83.1	79.7	85.2
Parking	71.0	72.5	74.0	76.7	76.8
Staff	76.1	76.1	82.1	77.9	83.0
Timeliness	72.3	73.0	77.4	79.6	81.9

The resource is a dataset obtained from PatientOpinion,⁴ a social enterprise pioneering an online feedback service for users of the UK national health service to enable people to share their recent experience of local health services online. It is a manually tagged dataset of 2000 patient opinions that associates to each post a category (namely, clinical service, communication, food, parking, staff, and timeliness) and a positive or negative polarity. There are no ethical issues involved in the data used in the experimentation as tweets, blogposts, and patient opinions were all anonymized. In order to guarantee full anonymity, moreover, the text associated with tweets, blogposts, and patient opinions has never been wholly reported in the proposed tables and examples. The dataset is hereby used to test the combined detection of opinion targets and the polarity associated with these. Results show that the ELM-based framework outperforms k-medoids, k-NN, and the two SLFNs in all cases (Table 6). Besides achieving higher generalization performance, the ELM-based predictors have lower computational complexity, in which they use just 50 dimensions (in stead of 100), and faster learning speed, in which they were trained in the order of milliseconds (rather than seconds).

7. Conclusion and future work

In a world in which millions of people express their opinions about commercial products and services everywhere on the Web, the distillation of knowledge from this huge amount of unstructured information is a key factor for tasks such as social media marketing, product positioning, and financial market prediction. However, all the sentiment analysis tasks are highly challenging. Our understanding and knowledge of the problem and its solution are still very limited. A possible reason for this limitedness may be due to our popular ways of doing research. So far, researchers have probably relied too much on machine learning algorithms. Some of the most effective machine learning algorithms, produce no human understandable results such that, although they may achieve improved accuracy, little about how and why is known, apart from some superficial knowledge gained in the manual feature engineering process. All such approaches, moreover, rely on syntactical structure of text, which is far from the way human mind processes natural language.

Sentic computing and similar concept-based approaches aim to bridge this gap by giving machines means to grasp the semantics associated with natural language concepts through common-sense reasoning, rather than blindly processing text through a limited set of affect words and their co-occurrence frequencies. Sentic computing, moreover, enables finer-grained opinion mining. Opinions and sentiments do not occur only at document- or paragraph-level, nor are they limited to a single valence or target. Contrary or complementary attitudes toward the same topic or multiple topics can be present across the span of the same document, paragraph, and even sentence. To this end, sentic

computing pioneers the concepts of affective common-sense reasoning and concept-level sentiment analysis.

In this work, we explored how an ensemble of concept-based sentiment analysis and machine learning techniques could emulate the cognitive process of affective analogical reasoning, in order to quickly, dynamically, and effectively infer semantics and sentics associated with natural language concepts. For its intrinsic properties, ELM turned out to be a better option than state-of-the-art approaches. In particular, we focused on demonstrating the higher generalization performance of the ELM-based predictors over standard SLFNs. However, the ELM-based framework is also characterized by lower computational complexity and faster learning speed. Because of the current dimensions of AffectiveSpace, it is difficult to thoroughly prove the performance of such features (although it was already proved that the ELM predictors used half of the dimensions needed by SLFNs and were trained three orders of magnitude faster). To this end, we plan a more extensive evaluation using a bigger AffectiveSpace and showing how the ELM-based framework is a better solution for online learning, i.e., for an AffectiveSpace that is dynamically expanding (in which new common-sense concept are being continuously learnt).

The use of multiple ELMs was inspired by Minsky's theory of the mind, according to which brain activity consists of different independent resources and that emotional states result from turning some set of these resources on and turning another set of them off. Along these lines, we plan to further explore how the ensemble application of sentic computing and ELM can enable multi-level affective common-sense reasoning. Current thinking in cognitive psychology suggests that humans process information at multiple levels.

Most commonly, a dual-mode model [46,47] has one system employing slow, effortful, serial, conscious processing, and another which involves fast, effortless, parallel, unconscious processing. While for the former mode (conscious reasoning) traditional techniques [4,5] may be applied, for the latter mode (unconscious reasoning) ELM is clearly more suitable, especially for cases in which affective analogical reasoning has to be performed on common-sense concepts for which few semantic features are available, e.g., newly learnt concepts.

In all such cases, a good-guess would be preferable than a no-answer caused by inconsistency and sparseness of the common-sense data. An important difference between traditional AI systems and human intelligence is our ability to harness common-sense knowledge to inform our decision-making strategies and behavior on the fly. This allows humans to easily and quickly adapt to novel situations where traditional AI fails catastrophically for lack of situation-specific rules and generalization capabilities, which instead ELM-based models can potentially master for the fast learning speed, low computational complexity, and high generalization performance that characterize them.

References

- [1] B. Pang, L. Lee, *Opinion mining and sentiment analysis*, *Found. Trends Inf. Retr.* 2 (2008) 1–135.
- [2] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan and Claypool, 2012.
- [3] E. Cambria, A. Hussain, C. Havasi, C. Eckl, *SenticSpace: visualizing opinions and sentiments in a multi-dimensional vector space*, in: R. Setchi, I. Jordanov, R. Howlett, L. Jain (Eds.), *Knowledge-Based and Intelligent Information and Engineering Systems, Lecture Notes in Artificial Intelligence*, Vol. 6279, Springer, Berlin, 2010, pp. 385–393.
- [4] E. Cambria, A. Hussain, *Sentic Computing: Techniques, Tools, and Applications*, Springer, Dordrecht, Netherlands, 2012.
- [5] E. Cambria, T. Mazzocco, A. Hussain, *Application of multi-dimensional scaling and artificial neural networks for biologically inspired opinion mining*, *Biol. Inspir. Cogn. Archit.* 4 (2013) 41–53.
- [6] C.D. Elliott, *The Affective Reasoner: A Process Model of Emotions in a Multi-Agent System*, Northwestern University, 1992.

⁴ <http://patientopinion.org.uk>

- [7] A. Ortony, G. Clore, A. Collins, *The Cognitive Structure of Emotions*, Cambridge University Press, Cambridge, 1988.
- [8] J. Wiebe, T. Wilson, C. Cardie, Annotating expressions of opinions and emotions in language, *Lang. Resour. Eval.* 39 (2) (2005) 165–210.
- [9] R. Stevenson, J. Mikels, T. James, Characterization of the affective norms for english words by discrete emotional categories, *Behav. Res. Methods* 39 (2007) 1020–1024.
- [10] S. Somasundaran, J. Wiebe, J. Ruppenhofer, Discourse level opinion interpretation, in: COLING, Manchester, 2008, pp. 801–808.
- [11] D. Rao, D. Ravichandran, Semi-supervised polarity lexicon induction, in: EACL, Athens, 2009, pp. 675–682.
- [12] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, in: EMNLP, Philadelphia, 2002, pp. 79–86.
- [13] A. Weichselbraun, S. Gindl, A. Scharl, Extracting and grounding context-aware sentiment lexicons, *IEEE Intell. Syst.* 28 (2) (2013) 39–46.
- [14] L. Garcia-Moya, H. Anaya-Sanchez, R. Berlanga-Llavori, A language model approach for retrieving product features and opinions from customer reviews, *IEEE Intell. Syst.* 28 (3) (2013) 19–27.
- [15] R. Xia, C. Zong, X. Hu, E. Cambria, Feature ensemble plus sample selection: a comprehensive approach to domain adaptation for sentiment classification, *IEEE Intell. Syst.* 28 (3) (2013) 10–18.
- [16] G. Di Fabbriozio, A. Aker, R. Gaizauskas, Summarizing on-line product and service reviews using aspect rating distributions and language modeling, *IEEE Intell. Syst.* 28 (3) (2013) 28–37.
- [17] A. Tsai, R. Tsai, J. Hsu, Building a concept-level sentiment dictionary based on commonsense knowledge, *IEEE Intell. Syst.* 28 (2) (2013) 22–30.
- [18] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, S. Bandyopadhyay, Enhanced SenticNet with affective labels for concept-based opinion mining, *IEEE Intell. Syst.* 28 (2) (2013) 31–38.
- [19] C. Hung, H.-K. Lin, Using objective words in SentiWordNet to improve sentiment classification for word of mouth, *IEEE Intell. Syst.* 28 (2) (2013) 47–54.
- [20] C. Bosco, V. Patti, A. Bolioli, Developing corpora for sentiment analysis and opinion mining: a survey and the Senti-TUT case study, *IEEE Intell. Syst.* 28 (2) (2013) 55–63.
- [21] M. Minsky, *The Society of Mind*, Simon and Schuster, New York, 1986.
- [22] G. Fauconnier, M. Turner, *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*, Basic Books, 2003.
- [23] C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. Natl. Bureau Stand.* 45 (4) (1950) 255–282.
- [24] C. Eckart, G. Young, The approximation of one matrix by another of lower rank, *Psychometrika* 1 (3) (1936) 211–218.
- [25] R. Plutchik, The nature of emotions, *Am. Sci.* 89 (4) (2001) 344–350.
- [26] M. Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, Simon & Schuster, New York, 2006.
- [27] L. Barrett, Solving the emotion paradox: categorization and the experience of emotion, *Pers. Soc. Psychol. Rev.* 10 (1) (2006) 20–46.
- [28] E. Krumbhuber, A. Kappas, Moving smiles: the role of dynamic components for the perception of the genuineness of smiles, *J. Nonverb. Behav.* 29 (1) (2005) 3–24.
- [29] M. Lewis, I. Granic, Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development, Cambridge University Press, 2002.
- [30] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*, Harper Perennial, 1991.
- [31] E. Cambria, H. Wang, B. White, Guest Editorial: Big Social Data Analysis, *Knowledge-Based Systems* 69 (2014) 1–2.
- [32] N.-Y. Liang, P. Saratchandran, G.-B. Huang, N. Sundararajan, Classification of mental tasks from eeg signals using extreme learning machine, *Int. J. Neural Syst.* 16 (01) (2006) 29–38.
- [33] Y. Xu, Z. Dong, K. Meng, R. Zhang, K. Wong, Real-time transient stability assessment model using extreme learning machine, *IET Gener. Trans. Distrib.* 5 (3) (2011) 314–322.
- [34] W. Zong, G.-B. Huang, Face recognition based on extreme learning machine, *Neurocomputing* 74 (16) (2011) 2541–2551.
- [35] X. Chen, Z.Y. Dong, K. Meng, Y. Xu, K.P. Wong, H. Ngan, Electricity price forecasting with extreme learning machine and bootstrapping, *IEEE Trans. Power Syst.* 27 (4) (2012) 2055–2062.
- [36] Y. Song, J. Zhang, Automatic recognition of epileptic eeg patterns via extreme learning machine and multiresolution feature extraction, *Expert Syst. Appl.* 40 (14) (2013) 5477–5489.
- [37] S. Decherchi, P. Gastaldo, R. Zunino, E. Cambria, J. Redi, Circular-ELM for the reduced-reference assessment of perceived image quality, *Neurocomputing* 102 (2013) 78–89.
- [38] P. Gastaldo, R. Zunino, E. Cambria, S. Decherchi, Combining ELM with random projections, *IEEE Intell. Syst.* 28 (6) (2013) 46–48.
- [39] E. Principi, S. Squartini, E. Cambria, F. Piazza, Acoustic template-matching for automatic emergency state detection: an elm based algorithm, *Neurocomputing*. <http://dx.doi.org/10.1016/j.neucom.2014.01.067>.
- [40] G.-B. Huang, D.H. Wang, Y. Lan, Extreme learning machines: a survey, *Int. J. Mach. Learn. Cybern.* 2 (2) (2011) 107–122.
- [41] S. Ridella, S. Rovetta, R. Zunino, Circular backpropagation networks for classification, *IEEE Trans. Neural Netw.* 8 (1) (1997) 84–97.
- [42] T.P. Vogl, J. Mangis, A. Rigler, W. Zink, D. Alkon, Accelerating the convergence of the back-propagation method, *Biol. Cybern.* 59 (4–5) (1988) 257–263.
- [43] G.-B. Huang, L. Chen, C.-K. Siew, Universal approximation using incremental constructive feedforward networks with random hidden nodes, *IEEE Trans. Neural Netw.* 17 (4) (2006) 879–892.
- [44] G.-B. Huang, An insight into extreme learning machines: random neurons, random features and kernels, *Cogn. Comput.* <http://dx.doi.org/10.1007/s12559-014-9255-2>.
- [45] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, *IEEE Trans. Syst. Man Cybern.* 42 (2) (2012) 513–529.
- [46] M. Lieberman, *Social cognitive neuroscience: a review of core processes*, *Annu. Rev. Psychol.* 58 (2007) 259–289.
- [47] D. Kahneman, *Thinking, Fast and Slow*, Farrar, Straus and Giroux, New York, 2011.



Erik Cambria received his BEng and MEng with honours in Electronic Engineering from the University of Genoa in 2005 and 2008, respectively. In 2012, he was awarded his PhD in Computing Science and Mathematics, following the completion of a Cooperative Awards in Science and Engineering (CASE) project born from the collaboration between the University of Stirling, the MIT Media Lab, and Sitekit Solutions Ltd., which included internships at HP Labs India, the Chinese Academy of Sciences, and Microsoft Research Asia. From August 2011 to May 2014, Erik was a research scientist at the National University of Singapore (Cognitive Science Programme) and an associate researcher at the Massachusetts Institute of Technology (Synthetic Intelligence Project). Today, Erik is an assistant professor at Nanyang Technological University (School of Computer Engineering), where he teaches natural language processing and data mining. His research interests include concept-level sentiment analysis, affective common-sense reasoning, noetic natural language processing, and intention awareness. Erik is editorial board co-chair of Cognitive Computation, associate editor of Knowledge-Based Systems, and guest editor of many other top-tier AI journals, including three issues of IEEE Intelligent Systems, two issues of IEEE CIM, and one issue of Neural Networks. He is also involved in several international conferences as workshop series organizer, e.g., ICDM SENTIRE since 2011, program chair, e.g., ELM since 2012, PC member, e.g., UAI in 2014, tutorial organizer, e.g., WWW in 2014, special track chair, e.g., AAAI FLAIRS in 2015, and keynote speaker, e.g., CICling in 2015.



Paolo Gastaldo obtained the “Laurea” degree in Electronic Engineering and a PhD in Space Sciences and Engineering (2004), both from Genoa University, Italy. Since 2004 he is with the Department of Biophysical and Electronics Engineering (DIBE) of Genoa University; he is currently with DIBE as an Assistant Professor. His main research area includes innovative systems for visual signal understanding, neural network-based methods for nonlinear information processing, and DSP-based architectures for advanced signal interpretation, such as intelligent object tracking for video surveillance and cryptography.



Federica Bisio obtained her BEng and MEng in Electronic Engineering from the University of Genoa in 2010 and 2012, respectively. She is currently pursuing a PhD in Electronic Engineering, Information Technology, Robotics and Telecommunications in the Department of Electrical, Electronics, Telecommunications Engineering and Naval Architecture (DITEN) at the University of Genoa. Her main research interests are machine learning and reasoning, both with respect to theory and to applications in different areas, including network security, social data analysis, and sentic computing.



Rodolfo Zunino (born Genoa, Italy, 1961) obtained the “Laurea” degree cum laude in Electronic Engineering from Genoa University in 1985. From 1986 to 1995 he was a research consultant with the Department of Biophysical and Electronic Engineering (DIBE) of Genoa University. He is currently with DIBE as an Associate Professor, teaching Electronics for Embedded Systems and Electronics for Security. His main scientific interests include intelligent systems for Computer Security, network security and Critical Infrastructure Protection, embedded electronic systems for neural networks, efficient models for data representation and learning, massive-scale text-mining and text-clustering methods,

and advanced techniques for multimedia data processing. Rodolfo Zunino coauthored more than 170 scientific papers in International Journals and Conferences; he has been the Co-Chairman of the two Editions of the International Workshop on Computational Intelligence for Security in Information Systems (CISIS'08 and CISIS'09). Since 2001 he is contributing as an Associate Editor of the IEEE

Transactions on Neural Networks, and has participated in the Scientific Committees of several International Events (ICANN'02, ICANN'09, IWPAAMS2004, IWPAAMS2005, Applied Computing 2006). Rodolfo Zunino is a Senior Member of IEEE (CIS—Computational Intelligence Society).