Random Features and Random Neurons for Brain-Inspired Big Data Analytics

Mandar Gogate, Amir Hussain
School of Computing
Edinburgh Napier University
Scotland, UK
{m.gogate,a.hussain}@napier.ac.uk

Kaizhu Huang
Department of Electrical and Electronic Engineering
Xi'an Jiaotong-Liverpool University
Suzhou, China
kaizhu.huang@xjtlu.edu.cn

***Abstract*—With the explosion of Big Data, fast and frugal reasoning algorithms are increasingly needed to keep up with the size and the pace of user-generated contents on the Web. In many real-time applications, it is preferable to be able to process more data with reasonable accuracy rather than having higher accuracy over a smaller set of data. In this work, we leverage on both random features and random neurons to perform analogical reasoning over Big Data. Due to their big size and dynamic nature, in fact, Big Data are hard to process with standard dimensionality reduction techniques and clustering algorithms. To this end, we apply random projection to generate a multi-dimensional vector space of commonsense knowledge and use an extreme learning machine to perform reasoning on it. In particular, the combined use of random multi-dimensional scaling and randomly-initialized learning methods allows for both better representation of high-dimensional data and more efficient discovery of their semantic and affective relatedness.**

***Index Terms*—Dimensionality reduction; neural networks.**

I. Introduction

Although fundamental in many areas of science, randomness is really native to computer science [1]. Its computational nature was clarified by Kolmogorov [2]. He and his followers built in the 1960s- 1970s the first successful theory of random objects, defined roughly as those that cannot be computed from short descriptions. Kolmogorov also suggested in the 1960s that randomness may have an important relationship with nondeterminism; namely, that the task of finding a 'nonrandomness' witness (i.e., short fast program generating a given string) may be a good candidate to prove that exhaustive search cannot be avoided.

In the context of big data analytics [3], randomness can be key to address emerging needs such as fast learning speed and big dimensionality reduction [4]. When dealing with highly-dynamic and highly-dimensional data, minimal human intervention and efficient data representation are important factors for making sense of big data streams. Because of Big Data's volume, velocity, and variety, in fact, standard data representation techniques and learning methods are bound to fail. In this work, we propose to exploit randomness to address the issue of scalability of knowledge representation and reasoning within sentic computing [5], a framework that has been used for several sentiment analysis applications, including healthcare [6], [7], intention awareness [8] and finance [9], [10].

In the past, graph mining techniques and multi-dimensionality reduction techniques [11] have been employed on a knowledge base obtained by blending ConceptNet [12], a directed graph representation of commonsense knowledge [13], with WordNet-Affect (WNA) [14], a linguistic resource for the lexical representation of affect. In this work, a novel cognitive model based on the combined use of random projections (RP) [15] and extreme learning machines (ELM) [16] is exploited on the same knowledge base to further improve the way multi-word expressions are organized in a brain-like universe of natural language concepts. Results demonstrate noticeable enhancements in emotion recognition from natural language text with respect to previously adopted strategies and pave the way for future development of more biologically inspired approaches to the emulation of affective commonsense reasoning.

The rest of this paper is organized as follows: Section 2 introduces related works; Section 3 illustrates how the affective commonsense knowledge base is constructed; next, Section 4 and 5 describe the multi-dimensional scaling techniques adopted to perform reasoning on such a knowledge base; then, Section 6 presents an evaluation of the proposed cognitive architecture; finally, Section 7 offers some concluding remarks and future work recommendations.

II. Related Work

Concept-level sentiment analysis is a NLP task that has recently raised growing interest both within the scientific community, leading to many exciting open challenges, as well as in the business world, due to the remarkable benefits to be had from marketing and financial market prediction. The potential applications of concept-level sentiment analysis, in fact, are countless and span interdisciplinary areas such as stock market prediction, political forecasting, social network analysis, social stream mining, and human-robot interaction.

For example, Li et al. [17] implemented a generic stock price prediction framework and plugged in six different models with different analyzing approaches. They used Harvard psychological dictionary and Loughran-McDonald financial sentiment dictionary to construct a sentiment space. Textual news articles were then quantitatively measured and projected onto such a sentiment space.

2375-9259/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDMW.2019.00080

The models' prediction accuracy was evaluated on five years historical Hong Kong Stock Exchange prices and news articles and their performance was compared empirically at different market classification levels. Rill et al. [18] proposed a system designed to detect emerging political topics in Twitter sooner than other standard information channels. For the analysis, authors collected about 4 million tweets before and during the parliamentary election 2013 in Germany, from April until September 2013. It was found that new topics appearing in Twitter can be detected right after their occurrence. Moreover, authors compared their results to Google Trends, observing that the topics emerged earlier in Twitter than in Google Trends.

Jung and Segev [19] analyzed how communities change over time in the citation network graph without additional external information and based on node and link prediction and community detection. The identified communities were classified using key term labeling. Experiments showed that the proposed methods can identify the changes in citation communities multiple years in the future with performance differing according to the analyzed time span.

Montejo-Raez et al. [20] introduced an approach for sentiment analysis in social media environments. Similar to explicit semantic analysis, microblog posts were indexed by a predefined collection of documents. In the proposed approach, performed by means of latent semantic analysis, these documents were built up from common emotional expressions in social streams. Bell et al. [21] proposed a novel approach to social data analysis, exploring the use of microblogging to manage interaction between humans and robots, and evaluating an architecture that extends the use of social networks to connect humans and devices. The approach used NLP techniques to extract features of interest from textual data retrieved from a microblogging platform in real-time and, hence, to generate appropriate executable code for the robot. The simple rule-based solution exploited some of the 'natural' constraints imposed by microblogging platforms to manage the potential complexity of the interactions and to create bi-directional communication.

III. Building the Knowledge Base

The affective commonsense knowledge base developed within this research work is built upon ConceptNet, the graph representation of the Open Mind corpus, which structurally similar to WordNet [22], but whose scope of contents is general world knowledge, in the same vein as Cyc [23]. Instead of insisting on formalizing commonsense reasoning using mathematical logic [24], ConceptNet uses a new approach: it represents data in the form of a semantic network and makes it available to be used in natural language processing (NLP). The prerogative of ConceptNet, in fact, is contextual commonsense reasoning: while WordNet is optimized for lexical categorization and word-similarity determination, and Cyc is optimized for formalized logical reasoning, ConceptNet is optimized for making practical context-based inferences over real-world texts.

In ConceptNet, WordNet's notion of node in the semantic network is extended from purely lexical items (words and simple phrases with atomic meaning) to include higher-order compound concepts, e.g., 'satisfy hunger' and 'follow recipe', to represent knowledge around a greater range of concepts found in everyday life. Moreover WordNet's repertoire of semantic relations is extended from the triplet of synonym, *IsA* and *PartOf*, to a repertoire of twenty semantic relations including, for example, *EffectOf* (causality), *SubeventOf* (event hierarchy), *CapableOf* (agent's ability), *MotivationOf* (affect), *PropertyOf*, and *LocationOf*. ConceptNet's knowledge is also of a more informal, defeasible, and practically valued nature.

For example, WordNet has formal taxonomic knowledge that 'dog' is a 'canine', which is a 'carnivore', which is a 'placental mammal'; but it cannot make the practically oriented member-to-set association that 'dog' is a 'pet'. ConceptNet also contains a lot of knowledge that is defeasible, i.e., it describes something that is often true but not always, e.g., EffectOf('fall off bicycle', 'get hurt'), which is something that cannot be left aside in commonsense reasoning. Most of the facts interrelating ConceptNet's semantic network are dedicated to making rather generic connections between concepts. Hence, ConceptNet alone is not enough for sentiment analysis tasks as it specifies how concepts are semantically related to each other but often lacks connections between concepts that convey the same kind of emotion or similar polarity.

To overcome such a hurdle, WNA, a linguistic resource for the lexical representation of affective knowledge developed starting from WordNet, is used. WNA is built by assigning to a number of WordNet synsets one or more affective labels (a-labels). In particular, the affective concepts representing emotional states are identified by synsets marked with the a-label 'emotion', but there are also other a-labels for concepts representing moods, situations eliciting emotions, or emotional responses. WNA was developed in two stages. The first consisted of the identification of a first core of affective synsets. The second step consisted of the extension of the core with the relations defined in WordNet. ConceptNet and WNA are blended together by combining the matrix representations of the two knowledge bases linearly into a single matrix, in which the information between the two initial sources is shared.

The first step to create the affective blend is to transform the input data so that it can all be represented in the same matrix. To do this, the lemma forms of ConceptNet concepts are aligned with the lemma forms of the words in WNA and the most common relations in the affective knowledge base are mapped into ConceptNet's set of relations, e.g., Hypernym into *IsA* and Holonym into *PartOf*. In particular, ConceptNet is first converted into a matrix by dividing each assertion into two parts: a concept and a feature, where a feature is simply the assertion with the first or the second concept left unspecified such as 'a wheel is part of' or 'is a kind of liquid'. The entries in the resulting matrix are positive or negative numbers, depending on the reliability of the assertions, and their magnitude increases logarithmically with the confidence score.

WNA, similarly, is represented as a matrix where rows are affective concepts and columns are features related to these. The result of aligning the matrix representations of ConceptNet and WNA is a new affective semantic network, in which commonsense concepts are linked to a hierarchy of affective domain labels. In such a semantic network, termed AffectNet, commonsense and affective knowledge are in fact combined, not just concomitant, i.e., everyday life concepts like 'have breakfast', 'meet people', or 'watch tv' are linked to affective domain labels like 'joy', 'anger', or 'surprise'. Such knowledge base results very useful when performing tasks such as emotion recognition or polarity detection from natural language text, as opinions and sentiments are often conveyed implicitly through context and domain dependent concepts, rather than through specific affect words.

IV. Multi-Dimensional Scaling

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 met 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. This kind of thinking is maybe the essence of human intelligence since in everyday life no two situations are ever the same and we have to continuously perform analogical reasoning for problem solving and decision making.

The human mind constructs intelligible meanings by continuously compressing over vital relations [25]. The compression principles aim to transform diffuse and distended conceptual structures to more focused versions so as to become more congenial for human understanding. In order to emulate such a process, principal component analysis (PCA) was previously applied on the matrix representation of AffectNet, a semantic network in which commonsense concepts were linked to semantic and affective features (Table 1). The result was AffectiveSpace.

PCA is most widely used as a data-aware method of dimensionality reduction [26]. PCA is closely related to the low-rank approximation method, singular value decomposition (SVD), in the sense that PCA works on a transformed version of the data matrix [27]. SVD seeks to decompose the AffectNet matrix $A \in \mathbb{R}^{\mathrm{n}\times\mathrm{d}}$ into three components,

$$A = USV^T, \tag{1}$$

where U and V are unitary matrices, and S is an rectangular diagonal matrix with nonnegative real numbers on the diagonal.

SVD has been proved to be optimal in preserving any unitarily invariant norm[1] $\|\cdot\|_M$ [27]:

$$\| A - A_k \|_M = \min_{\mathrm{rank(B)=k}} \| \mathrm{A} - \mathrm{B} \|_{\mathrm{M}}, \tag{2}$$

[1] A norm $\|\cdot\|_M$ is unitarily invariant if $\|UAV\|_M = \|A\|_M$ for all A and all unitary U, V.

TABLE I
A SNIPPET OF THE AFFECTNET MATRIX

AffectNet	*IsA-pet*	*KindOf-food*	*Arises-joy*	...
dog	0.981	0	0.789	...
cupcake	0	0.922	0.910	...
songbird	0.672	0	0.862	...
gift	0	0	0.899	...
sandwich	0	0.853	0.768	...
rotten fish	0	0.459	0	...
win lottery	0	0	0.991	...
bunny	0.611	0.892	0.594	...
police man	0	0	0	...
cat	0.913	0	0.699	...
rattlesnake	0.432	0.235	0	...
white tiger	0.230	0	0	...
...	...	...	...	...

where A_k, i.e., AffectiveSpace, is formed by only containing the top k singular values in S. Hence, in AffectiveSpace, commonsense concepts and emotions are represented by vectors of k 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,...,e_{k-1} of the vector space. 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 SVD, 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 someone happy` are found very close in direction in the vector space, while concepts like `feel guilty`, `be laid off`, and `shed tear` are found in a completely different direction (nearly opposite with respect to the centre of the space).

The problem with this kind of representation is that it is not scalable: when the number of concepts and of semantic features grows, the AffectNet matrix becomes too high-dimensional and too sparse for SVD to be computed [28]. Although there has been a body of research on seeking for fast approximations of the SVD, the approximate methods are at most ≈ 5 times faster than the standard one [27], making it not attractive for real-world big data applications.

It has been conjectured that there might be simple but powerful meta-algorithms underlying neuronal learning [29]. These meta-algorithms should be fast, scalable, effective, with few-to-no specific assumptions, and biologically plausible [28]. Optimizing all the $\approx 10^{15}$connections through the last few million years' evolution is very unlikely [28]. Alternatively, nature probably only optimizes the global connectivity (mainly the white matter), but leaves the other details to randomness [28].

In order to cope with the ever-growing number of concepts and semantic features, thus, we replace SVD with random projection (RP) [15], a data-oblivious method, to map the original high-dimensional data-set into a much lower-dimensional subspace by using a Gaussian $\mathrm{N}(0,1)$ matrix, while preserving the pair-wise distances with high probability. This theoretically solid and empirically verified statement follows Johnson and Lindenstrauss's (JL) Lemma [28]. The JL Lemma states that with high probability, for all pairs of points $\mathrm{x}, \mathrm{y} \in \mathrm{X}$ simultaneously,

$$\sqrt{\frac{m}{d}} \parallel x - y \parallel_2 (1-\varepsilon) \leq \parallel \Phi x - \Phi y \parallel_2 \leq \tag{3}$$

$$\leq \sqrt{\frac{m}{d}} \parallel x - y \parallel_2 (1+\varepsilon), \tag{4}$$

where X is a set of vectors in Euclidean space, d is the original dimension of this Euclidean space, m is the dimension of the space we wish to reduce the data points to, ε is a tolerance parameter measuring to what extent is the maximum allowed distortion rate of the metric space, and Φ is a random matrix.

Structured random projection for making matrix multiplication much faster was introduced in [30]. Achlioptas [31] proposed *sparse random projection* to replace the Gaussian matrix with i.i.d. entries in

$$\phi_{ji} = \sqrt{s} \begin{cases} 1 & \text{with prob. } \frac{1}{2s} \\ 0 & \text{with prob.} 1 - \frac{1}{s} \\ -1 & \text{with prob.} \frac{1}{2s} \end{cases}, \tag{5}$$

where one can achieve a $\times 3$ speedup by setting $s = 3$, since only $\frac{1}{3}$ of the data need to be processed. However, since our input matrix is already too sparse, we avoid using sparse random projection.

When the number of features is much larger than the number of training samples ($d \gg n$), subsampled randomized Hadamard transform (SRHT) is preferred, as it behaves very much like Gaussian random matrices but accelerates the process from $\mathcal{O}(n\,d)$ to $\mathcal{O}(n\,log\,d)$ time [32]. Following [33] [32], for $d = 2^p$ where p is any positive integer, a SRHT can be defined as:

$$\Phi = \sqrt{\frac{d}{m}} \mathrm{RHD} \tag{6}$$

where

• m is the number we want to subsample from d features randomly.

• R is a random $m \times d$ matrix. The rows of R are m uniform samples (without replacement) from the standard basis of $\mathbb{R}^d$.

• H$\in \mathbb{R}^{d\times d}$ is a normalized Walsh-Hadamard matrix, which is defined recursively: $H_d = \begin{bmatrix} H_{d/2} & H_{d/2} \\ H_{d/2} & H_{d/2} \end{bmatrix}$ with $H_2 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}$.

• D is a $d \times d$ diagonal matrix and the diagonal elements are i.i.d. Rademacher random variables.

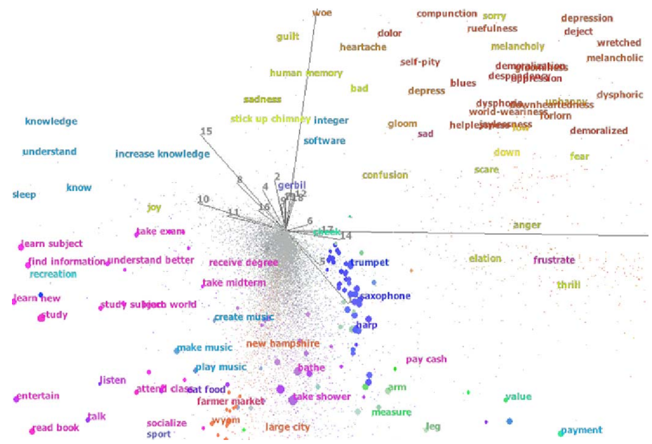

Fig. 1. AffectiveSpace.

Our subsequent analysis only relies on the distances and angles between pairs of vectors (i.e. the Euclidean geometry information), and it is sufficient to set the projected space to be logarithmic in the size of the data [34] and apply SRHT. The result is a new vector space model, AffectiveSpace (Fig. 1), which preserves the semantic and affective relatedness of commonsense concepts while being highly scalable.

V. Affective Analogical Reasoning

To reason on the disposition of concepts in AffectiveSpace, we use the Hourglass of Emotions (Fig. 2), an affective categorization model developed starting from Plutchik's studies on human emotions [35]. In the model, sentiments are reorganized around four independent dimensions whose different levels of activation make up the total emotional state of the mind. The Hourglass of Emotions, in fact, is based on the idea that the mind is made of different independent resources and that emotional states result from turning some set of these resources on and turning another set of them off [36].

In the model, affective states are not classified, as often happens in the field of emotion analysis, into basic emotional categories, but rather into four concomitant but independent dimensions, characterized by six levels of activation, which determine the intensity of the expressed/perceived emotion as a $float \in$ [-1,+1]. Such levels are also labeled as a set of 24 basic emotions (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 in a discrete form.

Such basic emotions are used as seeds in AffectiveSpace for classification by means of extreme learning machine (ELM) [37]. ELM is an emerging learning technique that provides efficient unified solutions to generalized feed-forward networks and, hence, have strong potential as a viable alternative technique for large-scale computing and machine learning in many different application fields, including text analysis [**?**], speech processing [39] and multimodality [**?**].

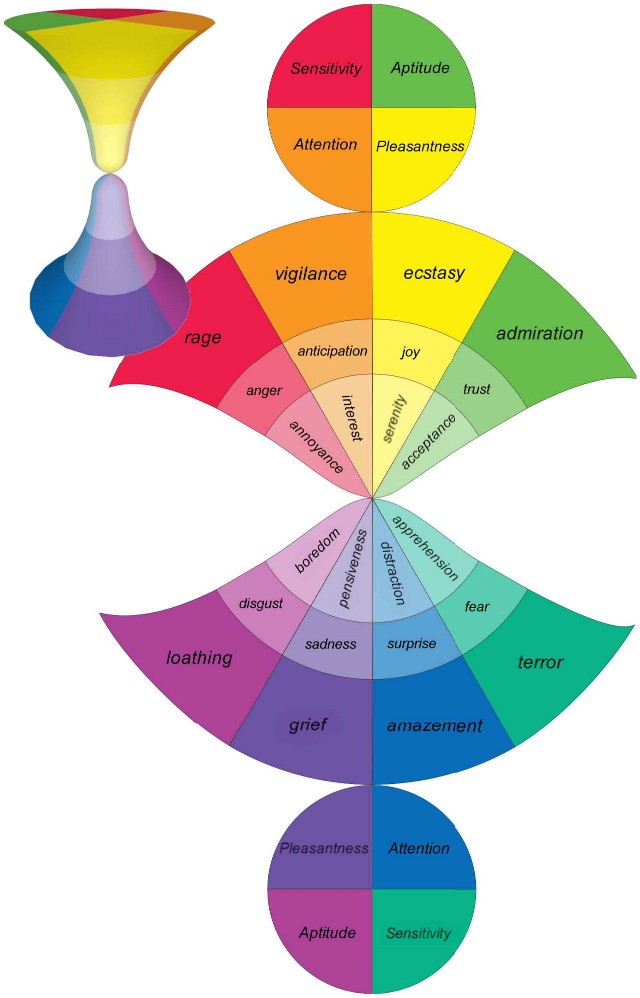

Fig. 2. The Hourglass of Emotions.

ELM learning theory [16] shows that the hidden neurons of generalized feedforward networks do not need to be tuned and these hidden nodes can be randomly generated. All the hidden node parameters are independent from the target functions or the training datasets. ELM theories conjecture that this randomness may be true to biological learning in animal brains [41]. In this work, we merge the advantages of ELM, e.g., fast learning speed, ease of implementation, and minimal human intervention, with the computational efficiency of random projections (RP) in order to deal with Big Data.

Let $\mathbf{x} \in \mathbb{R}^d$ denote an input vector. The function, $f(\mathbf{x})$, of an output neuron in an ELM that adopts L 'hidden' units is written as:

$$f(\mathbf{x}) = \sum_{j=1}^{L} w_j \cdot a(\mathbf{r}_j \cdot \mathbf{x} + b_j) \tag{7}$$

Thus, a set of random weights $\{\mathbf{r}_j \in \mathbb{R}^d;\ j = 1,..,L\}$ connects the input to the hidden layer; the j-th hidden neuron embeds a random bias term, b_j, and a nonlinear activation function, $a(.)$.

A vector of weighted links, $\mathbf{w} \in \mathbb{R}^L$, connects the hidden layer to the output neuron. The vector quantity $\mathbf{w} = [w_1, ..., w_L]$ embeds the degrees of freedom in the ELM learning process, which can be formalized after introducing the following notations:

- $\mathbf{X}$ is the N x $(d+1)$ matrix that originates from the training set. $\mathbf{X}$ stems from a set of N labeled pairs $(\mathbf{x}_i, y_i)$, where $\mathbf{x}_i$ is the i-th input vector and $y_i \in \mathbb{R}$ is the associate expected 'target' value.
- $\mathbf{R}$ is the $(d+1)$ x L matrix with the random weights.

Here, by using a common trick, both the input vector, $\mathbf{x}$, and the random weights, $\mathbf{r}_j$, are extended to $\mathbf{x} := [x_1, ..., x_d, 1]$ and $\mathbf{r}_j \in \mathbb{R}^{d+1}$ to include the bias term. Accordingly, the ELM learning process requires one to solve the following linear system

$$\mathbf{y} = \mathbf{H}\mathbf{w} \tag{8}$$

where $\mathbf{H}$ is the hidden layer output matrix obtained by applying the activation function, $a()$, to every element of the matrix:

$$\mathbf{XR} \tag{9}$$

The above expression clarifies that in the ELM scheme the hidden layer performs a mapping of the original d-dimensional space into a L-dimensional space through the random matrix $\mathbf{R}$, which is set independently from the distribution of the training data. In principle, the feature mapping phase may either involve a reduction in dimensionality ($L < d$) or, conversely, remap the input space into in an expanded space ($L > d$). The quantity L is crucial because it determines the generalization ability of the ELM. At the same time, it affects the eventual computational complexity of both the learning machine and the trained model. These aspects become critical in hardware implementations of the ELM model, where resource occupation is of paramount importance.

The present work exploits the fruitful properties of random projections to reduce dimensionality of data. The RP approach, in fact, can be employed to tune the basic quantity L. The ability of RP to preserve, approximately, the distances between the N data vectors in a subspace of lower dimension is a valuable property for machine learning applications in general. Indeed, this property is the conceptual basis of the novel approach that connects the ELM feature mapping scheme to the RP paradigm.

A new ELM model can be derived, if one set as hypotheses that 1) L should be smaller than d and 2) the mapping implemented by the weights $\mathbf{r}_j$ satisfies the JL lemma. Under these assumptions, the ELM mapping scheme always implements the dimensionality reduction process. In practice, one takes advantage of the properties of RP to obtain an ELM model that shrinks the size L of the hidden layer and reduces the computational overhead accordingly. The crucial point is that the JL lemma guarantees that the original geometry of the data is only slightly perturbed by the dimensionality reduction process; indeed, the degradation grows gradually as L decreases (given d and N).

In principle, the literature provides several criteria for the construction of a random matrix that satisfies the JL lemma. The present work focuses on matrices where the entries are independent realizations of $+1$ or -1 Bernoulli random variables; hence, matrix $\mathbf{R}$ is generated as follows:

$$\mathbf{R}_{i,j} = \begin{cases} 1/\sqrt{L}, & \text{with probability} 1/2 \\ -1/\sqrt{L}, & \text{with probability} 1/2 \end{cases} \tag{10}$$

VI. Evaluation

The proposed architecture investigates if an emulation of the biological neural system, represented by a randomness-based reasoning architecture, could outperform the state-of-the-art sentic computing framework based on PCA and k-medoids [42]. In particular, the proposed framework uses random projections to both represent and classify Big Data in a multi-dimensional vector space.

The performance of the proposed ELM are tabulated in Table II where they are compared with the state-of-the-art k-medoids approach, k-nearest neighbor (k-NN), and a random classifier in the context of emotion recognition. In particular, the percentage of entries where $b^* = b$ ('strict accuracy') is considered. However, since the used dataset can include noise and entries may incorporate a certain degree of subjectiveness, this criterion was relaxed by considering the accuracy of entries which have $|b^* - b| \leq 1$ ('relaxed accuracy'). As it can be seen from Table II, the proposed ELM approach outperforms the state-of-the-art k-medoids model, as well as the k-NN model and the random classifier.

	Strict accuracy	Relaxed accuracy
Random	14.3%	40.1%
k-NN	41.9%	72.3%
k-medoids	43.2%	74.1%
ELM	46.9%	84.3%

TABLE II
PERFORMANCE COMPARISON

In order to test the performance of the proposed approach in a more practical environment, the ELM was also embedded into an opinion mining engine for the inference of the 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 ConceptNet, WordNet and other linguistic resources. 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, differently from 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 presents". 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 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. In case any of the detected concepts is found more than once in the vector space (that is, any of the concepts has multiple senses), all the SBoC concepts are exploited for a context-dependent coarse sense disambiguation. In particular, to represent the expected semantic value of the clause as a whole, the vectors corresponding to all concepts in the clause (in their ambiguous form) can be averaged together. The resulting vector does not represent a single meaning but the 'ad-hoc category' of meanings that are similar to the various possible meanings of concepts in the clause [43]. Then, to assign the correct sense to the ambiguous concept, the sense of each concept that has the highest dot product (and thus the strongest similarity) with the clause vector has to be sought.

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 ELM, 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 (7).

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: 'top heap'>
SBoC#2:
<Concept: 'ok'>
<Concept: 'speaker'>
<Concept: !'good'++>
<Concept: 'see'>
SBoC#3:
<Concept: 'touchscreen'>
<Concept: 'put cloud nine'++>
SBoC#4:
<Concept: 'camera'>
<Concept: 'look good'−−>

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

TABLE III
STRUCTURED OUTPUT EXAMPLE OF OPINION MINING ENGINE

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 III). In order to evaluate the resulting opinion mining engine, a patient opinion database is used, and results obtained using k-NN and k-medoids are compared with those obtained using the ELM. The resource is a dataset obtained from PatientOpinion[2], 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 2,000 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 ELM generally outperforms k-medoids and k-NN (Table IV).

[2]http://patientopinion.org.uk

	k-NN	k-medoids	ELM
clinical service	70.1%	72.9%	81.8%
communication	69.8%	75.3%	78.1%
food	79.4%	79.6%	83.1%
parking	71.0%	72.5%	76.7%
staff	76.1%	76.1%	82.1%
timeliness	72.3%	73.0%	79.6%

TABLE IV
F-MEASURE VALUES RELATIVE TO PATIENTOPINION EVALUATION

VII. CONCLUSIONS AND FUTURE WORK

With the advent of Big Data, information extraction from the huge amount of available unstructured information derived from blog, wikis, and social networks is a very arduous task. While existing approaches to information extraction mainly work at a syntactic-level, computational techniques and tools were hereby employed to analyze text natural language at a semantic-level. In particular, we developed a randomness-based opinion mining engine that, first, deconstructs natural language text into concepts, then, encodes such concepts as coordinates of a multi-dimensional vector space, and finally infers the semantic and affective information associated with them.

The integration of random features for multi-dimensional scaling and random neurons for classification, in particular, has embedded a bio-inspired way of reasoning to carry out cognitive tasks such as emotion recognition and polarity detection. Such an ensemble model better grasps the non-linearities of the vector space of affective commonsense knowledge and, hence, improves the performance of the opinion mining engine. Randomness, moreover, allows for a more efficient analysis of Big Data in terms of both space and time consumption.

Since this study has shown promising results, further research is now planned to understand how the emulation of bio-inspired reasoning processes can aid big data analytics. Besides allowing for fast and frugal reasoning, moreover, randomness will be exploited to enable machines to represent knowledge and perform big data analytics in many different ways so that, whenever they reach a dead end, they can switch among different points of view and find one that may work.

REFERENCES

[1] L. Lewis, "Randomness and nondeterminsm," in *International Congress of Mathematicians*, (Zurich), 1994.
[2] A. Kolmogorov and V. Uspenskii, "Algorithms and randomness," *Theor. Veroyatnost. i Primenen.*, vol. 3, no. 32, pp. 389–412, 1987.
[3] E. Cambria, H. Wang, and B. White, "Guest editorial: Big social data analysis," *Knowledge-Based Systems*, vol. 69, pp. 1–2, 2014.
[4] Y. Zhai, Y.-S. Ong, and I. Tsang, "The emerging "big dimensionality"," *IEEE Computational Intelligence Magazine*, vol. 9, no. 3, pp. 14–26, 2014.
[5] E. Cambria, A. Hussain, C. Havasi, and C. Eckl, "Sentic computing: Exploitation of common sense for the development of emotion-sensitive systems," in *Development of Multimodal Interfaces: Active Listening and Synchrony* (A. Esposito, N. Campbell, C. Vogel, A. Hussain, and A. Nijholt, eds.), Lecture Notes in Computer Science, pp. 148–156, Berlin: Springer, 2010.
[6] E. Cambria, A. Hussain, T. Durrani, C. Havasi, C. Eckl, and J. Munro, "Sentic computing for patient centered applications," in *IEEE ICSP*, pp. 1279–1282, 2010.

[7] E. Cambria, T. Benson, C. Eckl, and A. Hussain, "Sentic PROMs: Application of sentic computing to the development of a novel unified framework for measuring health-care quality," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10533–10543, 2012.

[8] E. Cambria, "An introduction to concept-level sentiment analysis," in *Advances in Soft Computing and Its Applications* (F. Castro, A. Gelbukh, and M. González, eds.), vol. 8266 of *Lecture Notes in Computer Science*, (Berlin), pp. 478–483, Springer-Verlag, 2013.

[9] F. Xing, E. Cambria, and R. Welsch, "Natural language based financial forecasting: A survey," *Artificial Intelligence Review*, vol. 50, no. 1, pp. 49–73, 2018.

[10] A. Picasso, S. Merello, Y. Ma, L. Oneto, and E. Cambria, "Technical analysis and sentiment embeddings for market trend prediction," *Expert Systems with Applications 135*, vol. 135, pp. 60–70, 2019.

[11] E. Cambria, D. Olsher, and K. Kwok, "Sentic activation: A two-level affective common sense reasoning framework," in *AAAI*, pp. 186–192, 2012.

[12] R. Speer and C. Havasi, "ConceptNet 5: A large semantic network for relational knowledge," in *Theory and Applications of Natural Language Processing* (E. Hovy, M. Johnson, and G. Hirst, eds.), ch. 6, Springer, 2012.

[13] E. Cambria, A. Hussain, C. Havasi, and C. Eckl, "Common sense computing: From the society of mind to digital intuition and beyond," in *Biometric ID Management and Multimodal Communication* (J. Fierrez, J. Ortega, A. Esposito, A. Drygajlo, and M. Faundez-Zanuy, eds.), vol. 5707 of *Lecture Notes in Computer Science*, pp. 252–259, Berlin Heidelberg: Springer, 2009.

[14] C. Strapparava and A. Valitutti, "WordNet-Affect: An affective extension of WordNet," in *LREC*, (Lisbon), pp. 1083–1086, 2004.

[15] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: applications to image and text data," in *ACM SIGKDD*, pp. 245–250, 2001.

[16] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[17] X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, "News impact on stock price return via sentiment analysis," *Knowledge-Based Systems*, vol. 69, pp. 14–23, 2014.

[18] S. Rill, D. Reinel, J. Scheidt, and R. Zicari, "Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis," *Knowledge-Based Systems*, vol. 69, pp. 14–23, 2014.

[19] S. Jung and A. Segev, "Analyzing future communities in growing citation networks," *Knowledge-Based Systems*, vol. 69, pp. 34–44, 2014.

[20] A. Montejo-Raez, M. Diaz-Galiano, F. Martinez-Santiago, and A. Urena-Lopez, "Crowd explicit sentiment analysis," *Knowledge-Based Systems*, vol. 69, pp. 134–139, 2014.

[21] D. Bell, T. Koulouri, S. Lauria, R. Macredie, and J. Sutton, "Microblogging as a mechanism for human-robot interaction," *Knowledge-Based Systems*, vol. 69, pp. 64–77, 2014.

[22] C. Fellbaum, *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press, 1998.

[23] D. Lenat and R. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Boston: Addison-Wesley, 1989.

[24] E. Mueller, *Commonsense Reasoning*. Morgan Kaufmann, 2006.

[25] G. Fauconnier and M. Turner, *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books, 2003.

[26] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2005.

[27] A. K. Menon and C. Elkan, "Fast algorithms for approximating the singular value decomposition," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 5, no. 2, p. 13, 2011.

[28] D. Balduzzi, "Randomized co-training: from cortical neurons to machine learning and back again," *arXiv preprint arXiv:1310.6536*, 2013.

[29] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks," *Communications of the ACM*, vol. 54, no. 10, pp. 95–103, 2011.

[30] T. Sarlos, "Improved approximation algorithms for large matrices via random projections," in *FOCS*, pp. 143–152, 2006.

[31] D. Achlioptas, "Database-friendly random projections: Johnson-lindenstrauss with binary coins," *Journal of computer and System Sciences*, vol. 66, no. 4, pp. 671–687, 2003.

[32] Y. Lu, P. Dhillon, D. P. Foster, and L. Ungar, "Faster ridge regression via the subsampled randomized hadamard transform," in *Advances in Neural Information Processing Systems*, pp. 369–377, 2013.

[33] J. A. Tropp, "Improved analysis of the subsampled randomized hadamard transform," *Advances in Adaptive Data Analysis*, vol. 3, no. 01n02, pp. 115–126, 2011.

[34] N. Ailon and B. Chazelle, "Faster dimension reduction," *Communications of the ACM*, vol. 53, no. 2, pp. 97–104, 2010.

[35] R. Plutchik, "The nature of emotions," *American Scientist*, vol. 89, no. 4, pp. 344–350, 2001.

[36] M. Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York: Simon & Schuster, 2006.

[37] E. Cambria and G.-B. Huang *et al.*, "Extreme learning machines," *IEEE Intelligent Systems*, vol. 28, no. 6, pp. 30–59, 2013.

[38] Y. Li, S. Wang, Q. Pan, H. Peng, T. Yang, and E. Cambria, "Learning binary codes with neural collaborative filtering for efficient recommendation systems," *Knowledge-Based Systems*, vol. 172, pp. 64–75, 2019.

[39] E. Principi, S. Squartini, E. Cambria, and F. Piazza, "Acoustic template-matching for automatic emergency state detection: An ELM based algorithm," *Neurocomputing*, vol. 149, pp. 426–434, 2015.

[40] I. Chaturvedi, R. Satapathy, S. Cavallari, and E. Cambria, "Fuzzy commonsense reasoning for multimodal sentiment analysis," *Pattern Recognition Letters*, vol. 125, no. 264-270, 2019.

[41] G.-B. Huang, "An insight into extreme learning machines: Random neurons, random features and kernels," *Cognitive Computation*, vol. 6, no. 3, pp. 376–390, 2014.

[42] E. Cambria, T. Mazzocco, A. Hussain, and C. Eckl, "Sentic medoids: Organizing affective common sense knowledge in a multi-dimensional vector space," in *Advances in Neural Networks* (D. Liu, H. Zhang, M. Polycarpou, C. Alippi, and H. He, eds.), vol. 6677 of *Lecture Notes in Computer Science*, (Berlin), pp. 601–610, Springer-Verlag, 2011.

[43] C. Havasi, R. Speer, and J. Pustejovsky, "Coarse word-sense disambiguation using common sense," in *AAAI CSK*, (Arlington), pp. 46–51, 2010.