ASJP
Algerian Scientific Journal Platform

RIST

Information Processing at the Digital Age Journal

Organization of Knowledge and Advanced Technologies (OCTA)

https://multiconference-octa.loria.fr/

Chapter 2: Optimized Sentiments analysis Approach, Based On Aspects, Attention and Subjectivity notions For Textual Business Intelligence

Hammou FADILI[a]

[a]Hammou Fadili, (CEDRIC/CNAM Paris, Pôle Recherche & Prospective/FMSH), 54, bd Raspail 75006 Paris, France

Abstract

This article presents the results obtained in applying an innovative and optimized approach to textual semantic analysis in the service of decision-making. Significant improvements have been made in the existing procedures of sentiment and recommendation analysis, and in opinions mining, to enable better-motivated decisions and benefit from big data. These improvements concerned, especially, the support of the notions of aspects, attention and subjectivity to lighten the treatments, well adapted in the context of big data. The results obtained show the interesting contribution of the approach to the specific field of business intelligence (BI) relative to user behaviors analysis.

Keywords: Machine learning, Modeling and prediction, Natural Language Processing, Neural models, Semantics, Sentiments analysis, Text Mining

1. Introduction

This article presents a research on applying sentiment and recommendation analysis, opinion mining and user behavior understanding to Business Intelligence (BI). The work and its results are part of a general research strategy on economic intelligence and their exploitation in the context of Big Data.

In this context, data comes mainly from the reviews and opinions of users/customers, reported via social medias, forums, blogs, sales sites, etc., on all kinds of topics, such as events, products, attitudes, etc., to express their sentiments and experiences. The formalization of sentiment and recommendation analysis, and

the exploration of users' opinions and experiences, have become a major issue in the BI domain. Indeed, several studies and statistics confirm that more than 80% of users shopping on the Internet consult the comments and opinions of former users before making their own purchases [49]. Recommendations influence our opinions about products, services, etc., and therefore influence our purchases.

Moreover, despite advances in this area, several recent studies show that only 29% of companies and organizations use the data in their decisions [48]. The main reason for this is that existing solutions are not yet mature enough while the cost of implementing them is still too high. These technologies can also be used to enable companies to measure and better protect their reputation.: Individuals who use the Internet freely to express sentiments and opinions on everything can damage the reputation and activities of organizations. The latter must take these phenomena into account to protect themselves better. E-reputation has also become a new emerging field of research. Businesses and institutions therefore, have a great interest in understanding feedback from their customers. They must invest heavily in automatic systems analysis, to be helped to adapt to users' requirements and improve their profits; in the new open world of business intelligence that is the Web.

The phenomena of business data analysis for intelligence purposes in the context of Big Data challenges the traditional methods and approaches of modeling, design, implementation and evaluation of information and decision systems. Thus, the integration of new technologies, such as semantic textual analysis and artificial intelligence [50], is an unavoidable necessity. Whereas most existing solutions deal mainly with structured data, they must evolve to survive in the context of big and unstructured data. They encounter several problems when it comes to integrating and/or federating unstructured heterogeneous data sources as natural language texts: the exact meaning of words and thus texts in their contexts is very difficult to detect and to exploit. These problems are even more accentuated, in the case of Web texts often written with short texts (SMS), abbreviations, acronyms, styles & very particular elements of speech such as irony, humor, and so on.

To answer these questions, we have engaged in a research work as part of a global approach whose objective is to set up innovative methods to deal with the specific problems still posed by the semantic analysis of textual data in the service of decision-making. The goal is to make improvements to existing procedures used in sentiment and recommendation analysis and in opinion mining to enable better motivated decisions and benefit from big data. This will allow organizations to know how their products and services are perceived and to adapt them to the demands of their users/customers.

The issues raised during this work may relate to several areas of BI research but, with a view to simplification, we concentrated our research and development efforts on sentiment analysis and opinion mining. We were interested in the semantic analysis of unstructured and noisy data for the sentiment analysis.

This article is structured as follows. The first part will be devoted to a review of the state of the art, followed by a second part on the issues raised. In the third part, we present our contribution, before concluding the article.

2. State of the art

This part is devoted to a review of the state of the art in sentiment analysis in the context of Big Data, according to the following major points: semantic sentiment analysis, domains of use of sentiment analysis

and opinion mining, the different levels of analysis and, finally, the different approaches to calculating polarities.

A. Sentiment analysis and semantic text mining

A sentiment is a tendency to feel an emotion about an object or about a person [1] [2]. It also includes people's opinions, evaluations and feelings about entities, events, etc. [6]. In the context of Big Data, sentiments are often expressed in writing through product reviews, websites, blogs, discussion forums, and so on. In the context of the Web, these writings often convey messages and opinions difficult to detect and extract, because they are frequently expressed in languages and styles specific to the Web: SMS, acronyms, abbreviations, irony, humor, etc.

Sentiment analysis consists in attributing a relative polarity to subjective elements (words and sentences that express opinions, sentiments, emotions, ...) to decide on the orientation of a document (Turney, 2002). The interest in sentiment analysis and opinion mining, in a context where taking charge of several important elements such as speech acts, metaphor, frozen idiomatic expressions ..., naturally situates sentiments analysis in the general context of semantic text analysis of Big Data in the service of business intelligence. Big Data here is defined as consisting mainly of massive, unstructured & heterogeneous textual data

Given the nature of the data, mainly textual, the techniques used to detect the messages conveyed and the opinions, must exploit Natural Language Processing (NLP) and linguistics to identify the polarity (positive, negative, or neutral) of a word, a sentence, a text or even a corpus.

B. Domains of use

Sentiments analysis is based, for the most part, on text mining. It can target the study, the detection and the extraction of sentiments, opinions, emotions and subjectivities in the text (Pang et al., 2004) in specific areas, such as advertising, marketing, production, business, politics, psychology, etc.

C. Classification of sentiment analysis technologies

Sentiment analysis can be used in several areas and for different needs: studies of the polarities and subjectivities of words (SentiWordNet and equivalents), expressions (phrases), documents (the most common case), corpora, etc. This involves different approaches and strategies characterized by different levels of analysis.

1) Word level analyses

Word level analyses deal with the polarities of words. They represent the bases of the analyses of the other levels: sentence, expression and document. Word level analyses are of two types: those based on lexicons of sentiments and those based on corpora.

2) Lexicon-based approaches for sentiment analysis

These approaches are considered unsupervised as they exploit dictionaries and lexicons where the words are characterized by polarity scores (cf. [44]). For example, the positive score of the word “excellent” is 0.8, its neutral score is 0.3 and its negative score is 0 (see [45]). Several methods are then used to calculate the polarity of several words: opinions, sentences, paragraphs and documents.

3) Corpora-based approaches for sentiment analysis

This kind of methods is considered supervised as it exploits a pre-processed corpus, i.e., a corpus that has been already annotated in order to train and learn a polarity classifier. In general, in this case, we cannot be limited to the "Positive", "Neutral" and "Negative" polarities, but we can enrich and refine the classes of polarities with others according to needs.

4) Sentence-level analyses

The phrase refers to a small group of words, forming a conceptual unit, usually a component of a clause. Sentiment analysis of the level of the sentence is as important as sentiment analysis of the level of the word. It allows the analysis of the higher levels: expression, document, etc. It is an important element in the process of analyzing sentiments. Several studies have been conducted in this area. They exploit systems based either on machine learning [46], on statistical models [47] or on dictionaries of polarity [45].

5) Expression level analyses

The term also called "complete sentence" refers to a set of words that is complete, usually containing a subject and a predicate, conveying a statement, question, exclamation…; and consists of a main clause and sometimes one or more subordinate clauses. An expression can express a thought, a sentiment, an opinion, etc.

Several studies have also been conducted in this area (see [34]) and technologies for the classification of complete sentences were developed based on machine learning. The comments are first pretreated and annotated with grammatical annotations, among others. Their polarity is then calculated by combining partial results (words and sentences), using algorithms such as [29] based on statistical methods combined with a log-likelihood computation to calculate the scores.

6) Document level analyses

This kind of analysis looks at the sentiment of the whole document, for example, of specific piece of information, an opinion, a comment, a forum, a blog, etc. Several works have been done for the sentiment analysis of the document. They generally exploit machine learning-based approaches for classifying and inferring the polarity at the document level. In [48], the authors exploited unsupervised methods to classify documents using several steps:

- They extracted adjectives and adverbs using grammatical annotation systems,
- Then they extracted the sentences and their polarities ([31])
- Finally, they calculated the average to deduce the polarity of each document.

The results obtained were of the order of 70%.

D. Polarity measurements

From a technological point of view, the Big Data context imposes the use of machine learning for calculating polarity. Several studies have shown that: training a system on large amounts of data greatly increases its performance. Big Data is therefore a considerable asset for machine learning-based systems.

In this section, we will present some methods for classification and calculating scores and polarities; which fall into two categories:

- Unsupervised calculation methods, from lexicons of sentiments
- Supervised calculations methods, from annotated corpora

1) Unsupervised calculations from lexicons of sentiments

Lexical based methods of sentiments calculation are unsupervised because they do not require annotated corpora. They are based on the automatic extraction of discriminant characteristics (features) from text and the exploitation of sentiments lexicons from resources such as SentiWordNet [66] and SentiStrength [71] for classification.

In [49] for example, the authors calculate the sum of a synSet scores for different Tags of words, to which they associate a polarity (negative, neutral, positive). Note that for Tags that are not in SentiWordNet, they associate the score 0. They also refined the analysis by choosing the right score according to the grammatical category (noun, adjective and verb). They explain this by the following example: the word "short" admits 11 adjectives, 3 nouns, 7 adverbs and 2 verbs for the different meanings. For the adjective Tag, the term "short" is -1, for the nounTag the term is 0, and so on.

The methods in this category are based on additional features that can refine the analysis. We can target a point of view of analysis, for example a product, an organization, a service, a theme, etc. In [61], the authors explain their process as follows: they targeted a product and then extracted some features from user comments. They then identified the comments of each characteristic (by grouping or clustering), then designated all the comments of a positive, neutral or negative characteristic. Finally, they refined their experiences by combining the results obtained for each characteristic of a product.

Our approach belongs to and improves this last category. Instead of limiting the calculations directly to combinations of the scores of certain characteristics, we extend and enrich all the characteristics that we extract automatically and confront with a sentiment lexicon to submit to a recurrent neural network (RNN) with long-term memory (LSTM), on which improvements have been made, for the calculation of the global polarity (see our approach below).

2) Supervised calculations methods, from annotated corpora

The classification methods based on annotated corpora are supervised, they need an annotated corpus. They are used to train and learn automatic polarity measurement systems. The research in this area is slim, because of the difficulties involved in manual annotation processes, especially in the case of Big Data. Among research works conducted in this category, we find, for example, [47], where several news articles were classified into 7 categories ranging from 1 (very negative) to 7 (very positive) (cf. [56]).

3) « Emoticons » based calculations methods

Another way to calculate polarity scores is to use emoticons, which are icons that express an emotion, a state of mind, a sentiment, feeling, and so on. They are often used in social networks such as Facebook, Twitter. Several works have exploited emoticons for classification. Contents containing positive emoticons have been classified as positive, those containing negative Emoticons are classified negative and those containing no emoticons are classified as neutral (see [59]). The use of emoticons is limited and does not represent a solution in itself, because, on the one hand, not all documents include emoticons and, on the other hand, several studies have shown that there is often a mismatch between the emoticons and the associated sentiments.

3. Problems and motivation

From the point of view of content, sentiment analysis has become an important discipline in Big Data mining. Although there are several proposals and approaches in the literature, there is no solution that can support all aspects of sentiments analysis. Their numerous limitations are due to several considerations, as reflected the following findings, among others:

- The quality of sentiment analysis depends on the degree of support for the semantics of the texts that convey them;
- The presence of sentiments and opinions is conditioned by subjectivity;
- Sentiments depend on the domain, on the analytical point of view, etc.
- Sentiments analysis requires substantial data and technological resources, in addition to sophisticated processing systems.

We consider that these aspects, poorly or not at all modeled and supported in the current systems, are important in terms of optimization and efficiency. They constitute the basis of the problems to be solved in our work, which will be presented through the two prisms of subjectivity and semantics for sentiments analysis.

From the technological point of view, we believe that these aspects remain challenges for sentiment analysis, but also that they will be addressed by possible improvements in the exploitation of advanced artificial intelligence technologies to extract the knowledge inherent in the text.

Indeed, overall, there are several technologies used in the field of sentiment analysis. According to the literature, deep learning technology (Deep Learning) is now, by far, the most popular and exploited technology. It provides the best results compared to other technologies, especially in the context of Big Data.

Most of the existing technologies exploit the words in text directly, regardless of their meanings or their order. The most commonly used technology in this field "Word Embeddings" or "Word2Vec" and its various implementations (SKIP-GRAM & CBOW) [33], which does not ‘learn’ the meaning of words nor their order in their vector’s representations. It is obvious that the loss of order implies ‘automatically' the loss of semantics: moreover, we know several sentences, where the meaning changes when we change the order of words. Implementing technological solutions that take into account the meaning and order of words can make significant improvements. This is a first major technological problem we considered in current work.

Another problem encountered by the deep learning architectures exploited in the field of sentiment analysis is the fixed size of networks. A configured and trained classical neural network keeps its parameters throughout its entire life cycle. This poses a problem in the case of sentences, paragraphs, texts, etc. composed of variable numbers of words and has had a decisive impact in the choice of our solution. We adopted approaches based on an improved variant of recurrent neural networks (RNN) that are independent of the number of words to be processed and therefore the length of sentences and documents. We used N-GRAMS and sliding windows of fixed sizes to work around this problem, despite the loss of information occasioned by the fact that the meaning of a word is related to the sentence containing it and not necessarily to a window of words that surround it.

Additionally, deep learning means multilayer architectures, which in some cases can be composed of a large number of layers. It can be argued that since:

- The training of the network is done by the optimization of the parameters minimizing the prediction error, by the backpropagation of the error;
- The optimization is done by the gradient descent method, which is used to modify the coefficients (parameters of the system);
- In the case of a recurrent network, the learning is calculated by multiplying the coefficients of a layer by a specific value, λ, calculated from the loss, as many times as the rank of the layer in the network;

Then:

- If λ is less than 1, the learning of the first layers is multiplied by a number close to zero, so they do not learn anymore. We speak in this case of the Vanishing of the Gradient;
- If λ is greater than 1, the learning of the first layers is multiplied by a high number, provoking the explosion of the coefficients (no learning either). In this case, we speak of Explosion of the Gradient.
- Addressing the problem of the vanishing or the explosion of the gradient is another important element we consider.

4. Our contribution and experimentations

A. Context of the work

This section briefly recalls the general context of the work; in terms of issues and activities, and their evolution.

Initially, this work was done in the context of several projects on the analysis and exploitation of structured data and more specifically on business intelligence (BI) from databases. The evolutions of information systems and uses, generating significant amounts of heterogeneous data and nonexistent or complex data models, forced us to extend our processes to all kinds of data (not just structured) by integrating the semantic analysis of unstructured and more specifically textual data.

B. Our approach and experiments

The scientific and experimental approaches we adopted are of two types: those related to contents and their semantics, and those considering the technological and processing problems.

From the technological point of view, we have made the choice of using deep architectures because they have been tested and successfully used in the case of natural language processing, and also because they are well adapted to sentiments analysis. During our work, several configurations were tested, some of which generated a lot of difficulties and/or were abandoned because of certain layouts of symbolic and numerical/statistical layers in the learning models. The location of symbolic models in the general process was a problem; because they are difficult to set up, except for very limited cases. They require a lot of resources and knowledge, and therefore a lot of work.

To circumvent these difficulties, we integrated unsupervised models upstream of the process. These models are based on mathematical calculations in optimized, well-adapted semantic spaces. The advantage is that one does not need previously processed information; the data to be analyzed and the mathematical models are used directly to deduce the learning models and their representations.

This approach has the advantage of combining the two kind of learning (supervised and unsupervised), at different levels, and with specific layout. Its originality lies in the fact that, on the one hand, it incorporates important notions into the model of the sentiments analysis: polarity, subjectivity, etc. while, on the other hand, it uses these notions to generate a rich semantic data model on which we applied the learning algorithms.

As previously mentioned, deep learning has been successfully exploited in many areas including natural language processing. One of the best uses is the generation of dense semantic vector spaces (Mikolov 2013). More recently, deep learning networks were then enhanced by creating Recurrent Neural Networks (RNNs) to support sequential data, where each state is calculated from its previous state and the new entry. These recurrent networks can propagate information in both directions: to the input layers and to the output layers. They are an implementation of neural networks, close to the functioning of the human brain, where information can spread in all directions while exploiting memory via recurrent connections propagating the information of a subsequent learning (information stored).

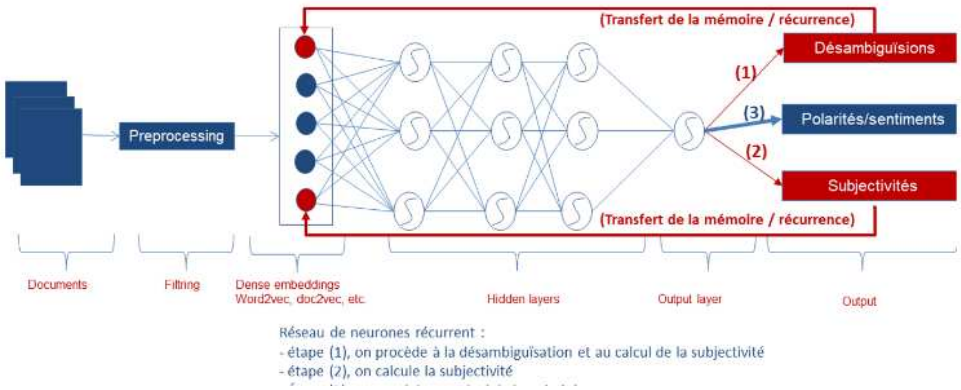

Fig. 1. Adaptation of a deep architectures for sentiments analysis

The depth of RNNs can be high because of their sequential nature, generally depending on the number of words to be processed. This can provoke:

- Vanishing of the Gradient in the first layers and stopping learning from a certain depth
- Explosion of the Gradient in the first layers and stopping learning from a certain depth

The Long Short-Term Memory (LSTM) architecture of RNNs was designed to address these issues, by optimizing control gates for the propagation of information in the network.

For the needs of Big Data, we adopted and improved the LSTM architecture initially based on three gates. We have additionally integrated the concepts of perspective, followed by the notion of attention. The new LSTM model now allows to control:

- What to use from the input
- What to use from the hidden state
- What to send to the output

and this depending on the chosen point of view or perspective and by paying attention to relevant elements.

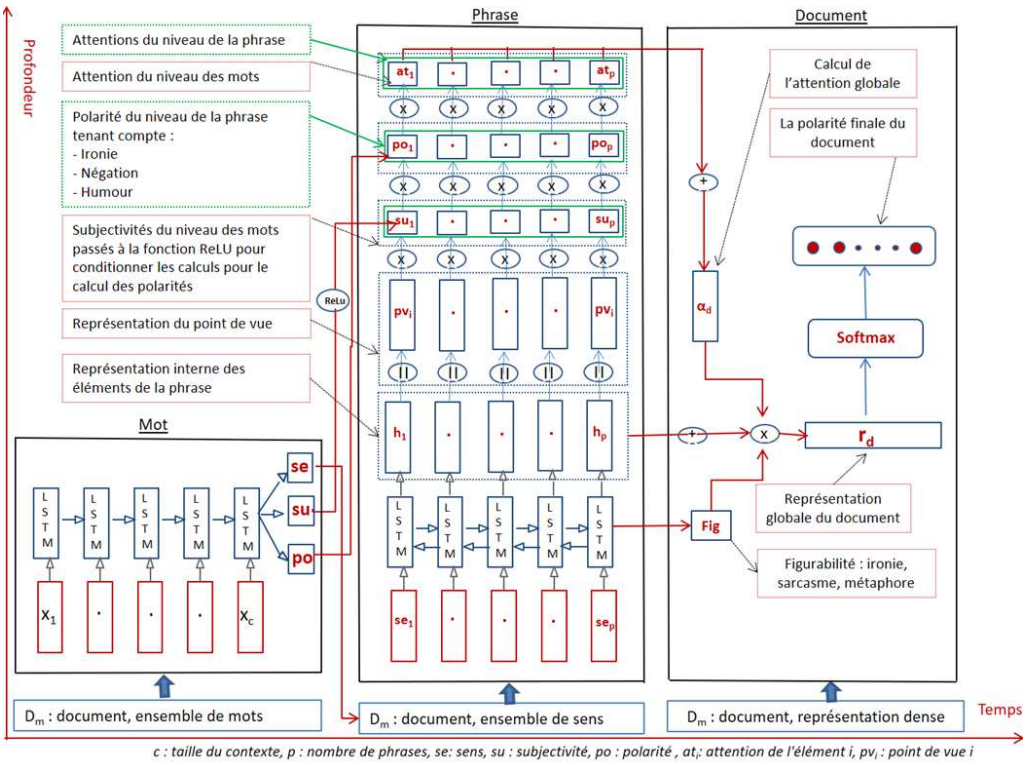

Fig. 2. Graphic version of our algorithm

These gates allow to cancel certain information that is useless for prediction and to reinforce other elements. It can be demonstrated mathematically that this architecture, in addition to optimizing the calculations in Big Data, makes it possible to solve the problems related to the vanishing and explosion of the gradient described above.

From a technological perspective, we have adapted the global semantic context model for the specific treatments of subjectivity, polarity, irony, metaphor, expressions (locutions). The notion of semantic context we are talking about here concerns all the parameters that can influence the meaning of words in the context of BI. This is a difficult notion to define, because there are no studies today defining in an exhaustive manner this notion in general and in the field of BI. It has been deduced that the context in the domain of BI can also be viewed as an aggregation of parameters of the context of the natural language (H. FADILI 2017), increased by new parameters specific to BI.

It should be noted that, in this section, only the new parameters that are to be combined with those of the natural language will be presented to form the overall context for BI. We classified them as following:

- Subjectivity: parameters to represent the subjectivity or not, of a word, a sentence, a document, etc.
- Polarity: parameters to represent the polarity, of a word, a sentence, a document, etc.
- Discourse analysis context: parameters to represent aspects of discourse analysis: irony, metaphor, expression, etc.

The detection of subjectivity consists in determining if a unit of language (word, phrase, document) expresses a personal attitude, an opinion, etc. and, if so, what is its polarity? Polarity has been conditioned by subjectivity to lighten the processing. At the level of the approach, the innovation consists of associating subjectivity to the sense and not to the words, to circumvent the issue of ambiguity negatively impacting the subjectivity and polarity calculations. We have distinguished two phases of analysis, after the disambiguation phase:

- Calculation of subjectivity, followed by
- Polarity calculation.

Indeed, a subjective word can have different polarities depending on the context. Its meaning can be modified by irony, humor, etc. Disambiguation alone is not enough. Dictionaries, mapping tables and ontologies can be used to model and represent certain elements of the context. We can, for example, use mappings to recognize the use of words in a domain, to find the complete forms of acronyms and initials or even to extract expressions for specific treatments.

As with natural language, sentiment analysis also needs the most complete representation of a word, a text (comment, recommendation, etc.), or a corpus, to capture all the discriminating information needed to deduce its exact polarity. Figure 1 models a representation of these elements, integrating several important notions for the analysis of unstructured data in the service of BI.

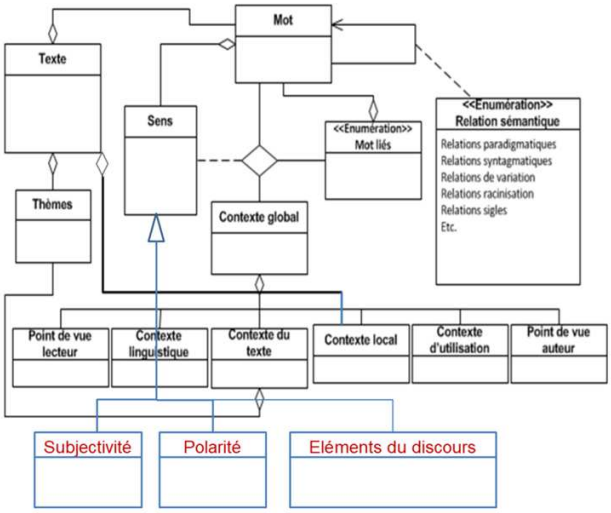

Fig. 3. Semantic model augmented by subjectivity, polarity and elements of discourse

C. Test scenarios

Several modules and programs have been developed that implement the elements of the overall process, mainly:

- Automatic extraction of the various semantic features of words and texts for sentiments analysis.
- Implementation of a deep learning system based on RNN and LSTM model with the attention and domain mechanisms described previously.
- We have also designed a development environment that centralizes access to all modules:
- Integration of all the elements in a single workflow implementing all the modules of the process.

As required in 'honest' machine learning systems, we divided the generated learning data into three parts:

A first part, representing 20% of the dataset, was used for validation, to optimize the hyper-parameters of the system: the learning step, the type of the activation function and the number of layers.

The rest of the dataset was divided into two parts:

- 60% for training, to estimate the best coefficients (wi) of the neural network function, minimizing the error between the real outputs and the desired outputs.
- 20% for tests, to evaluate the performance of the system.

During the learning phases, the system is autonomous. It generates the characteristics (features) of the text for the training, so that the trained model will be able to deduce the correct sense of each word of the text, as

well as the subjectivity and the polarity of the concerned elements (words, sentences, documents, corpora, etc.), when applied to unseen data.

D. An overview of the results

The results obtained during training and testing of the system are shown in Figures 2 and 3.

1) Evolution of accuracy during training & tests

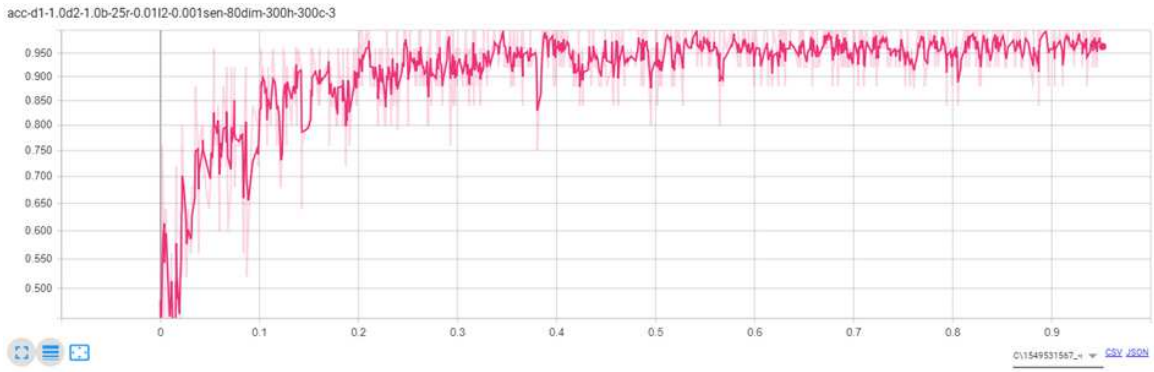

Fig. 4. An overview of the results: accuracy

2) Evolution of the loss during training & tests

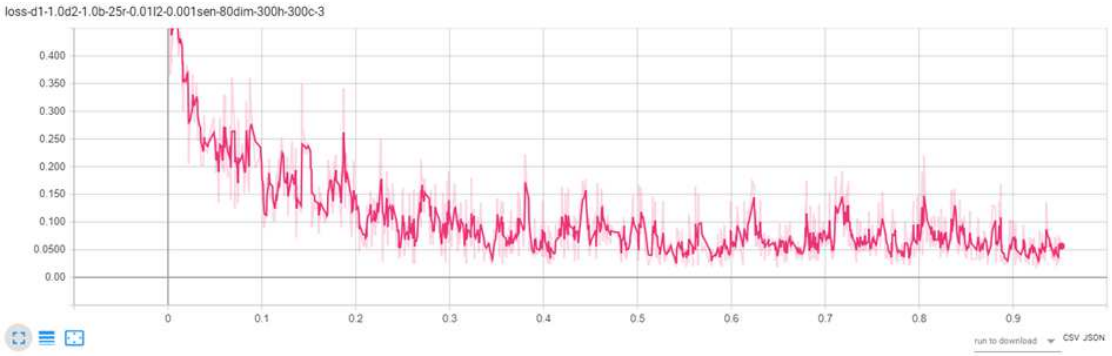

Fig. 5. An overview of the results: loss

The results of the tests show the good performance of the approach, from the evolutions of the accuracies and the errors. The system succeeds thanks to the learning on a part of the instances to deduce in the end the polarity of the concerned elements (corpus, document, paragraph or sentence) with better rates of precision and error.

5. Conclusion

Sentiment analysis has become a very important discipline in the exploitation of Big Data. Although there are several proposals and approaches in the literature, there is at present no solution that can handle all the requirements of the sentiment analysis process, especially, the support of the notions of aspects, attention and subjectivity to lighten the treatments of large masses of data. It is in this context that we have proposed a solution that targets the following aspects:

- Improved disambiguation at the beginning of the process;
- Better use of global context to deduce semantics and then subjectivity and polarity;
- Potential for mapping ontologies and other knowledge bases to solve problems such as acronyms, SMS, phrases (expressions), etc.
- Implementation of an optimized LSTM version of a deep neural network, with the attention and aspects notions.

At the current state of progress of the project, only a part of the problems described above has been addressed and solved; the remaining parts constitute our planned future work.

References

[1] A Neelakantan, J Shankar, A Passos, A McCallum. Efficient non-parametric estimation of multiple embeddings per word in vector space. Conference on Empirical Methods in Natural Language Processing, 2014
[2] Cui Tao, Dezhao Song, Deepak Sharma, Christopher G. Chute, Semantator: Semantic annotator for converting biomedical text to linked data. Journal of Biomedical Informatics, Volume 46, Issue 5, Pages 882-893 (October 2013). DOI: 10.1016/j.jbi.2013.07.003
[3] Das, T. K., & Kumar, P. M. (2013). Big data analytics: A framework for unstructured data analysis. International Journal of Engineering and Technology, 5(1), 153-156.
[4] Boury-Brisset, A.-C. (2013), Managing Semantic Big Data for Intelligence., in Kathryn Blackmond Laskey; Ian Emmons & Paulo Cesar G. da Costa, ed., 'STIDS' , CEUR-WS.org, , pp. 41-47 .
[5] Delia Rusu , Blaž Fortuna , Dunja Mladenić. Automatically Annotating Text with Linked Open Data (2011).Venue: In 4th Linked Data on the Web Workshop (LDOW 2011), 20th World Wide Web Conference.
[6] Archit Gupta, Krishnamurthy Viswanathan, Anupam Joshi, Tim Finin, and Ponnurangam Kumaraguru. Integrating Linked Open Data with Unstructured Text for Intelligence Gathering Tasks. Proceedings of the Eighth International Workshop on Information Integration on the Web, March 28, 2011.
[7] Isabelle Augenstein. Lodifier: Generating Linked Data from Unstructured Tex". ESWC 2012
[8] Marin Dimitrov. From Big Data to Smart Data. Semantic Days May 2013
[9] Khalili, A.; Auer, S. & Ngonga Ngomo, A.-C. (2014), conTEXT -- Lightweight Text Analytics using Linked Data, in 'Extended Semantic Web Conference (ESWC 2014)'.
[10] E. Khan, “Addressing Big Data Problems using Semantics and Natural Language Understanding,” 12th Wseas International Conference on Telecommunications and Informatics (Tele-Info ‘13), Baltimore, September 17-19, 2013.
[11] E. Khan, "Processing Big Data with Natural Semantics and Natural Language Understanding using Brain-Like Approach”, submitted to Journal– acceptance expected by Dec. 2013 Jan 2014.
[12] James R. Curran, Stephen Clark, and Johan Bos (2007): Linguistically Motivated Large-Scale NLP with C&C and Boxer. Proceedings of the ACL 2007 Demonstrations Session (ACL-07 demo), pp.33-36.
[13] Hans Kamp (1981). A Theory of Truth and Semantic Representation. In P. Portner & B. H. Partee (eds.), Formal Semantics - the Essential Readings. Blackwell. 189-222.
[14] Minelli, Michael & Chambers, Michele & Dhiraj, Ambiga 2013. Big Data, Big Analytics: Emerging Business Intelligence and Analytics Trends for Today's Businesses.
[15] Chan, Joseph O. "An Architecture for Big Data Analytics." Communications of the IIMA 13.2 (2013): 1-13. ProQuest Central. Web. 6 May 2014.
[16] H. Fadili. Towards a new approach of an automatic and contextual detection of meaning in text, Based on lexico-semantic relations and the concept of the context., IEEE-AICCSA , May 2013.
[17] George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.
[18] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

[19] Hiemstra, P.H., Pebesma, E.J., Twenhofel, C.J.W. and G.B.M. Heuvelink, 2008. Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network. Computers & Geosciences, accepted for publication.
[20] Christian Bizer, Tom Heath, Kingsley Idehen, Tim B. Lee. Linked data on the web (LDOW2008), In Proceedings of the 17th international conference on World Wide Web (2008), pp. 1265-1266.
[21] Jianqing Fan, Fang Han, Han Liu. Challenges of Big Data analysis National Science Review, Vol. 1, No. 2. (1 June 2014), pp. 293-314.
[22] http://wiki.dbpedia.org/
[23] Publication MEDES 2016 : Towards an Automatic Analyze and Standardization of Unstructured Data in the context of Big and Linked Data. H. FADILI.
[24] Publication TICAM'2016 : Le Machine Learning : numérique non supervisé et symbolique peu supervisé, une chance pour l'analyse sémantique automatique des langues peu dotées. H. FADILI.
[25] Frijda, N. H., Mesquita, B., Sonnemans, J., & Van Goozen, S. (1991). The duration of affective phenomena or emotions, sentiments and passions.
[26] Shand, A. F. (1920). The foundations of character: Being a study of the tendencies of the emotions and sentiments. Macmillan and Company, limited.
[27] Zhou, X., Tao, X., Yong, J., & Yang, Z. (2013, June). Sentiment analysis on tweets for social events. In Computer Supported Cooperative Work in Design (CSCWD), 2013 IEEE 17th International Conference on (pp. 557-562). IEEE.
[28] Park, C., & Lee, T. M. (2009). Information direction, website reputation and eWOM effect: A moderating role of product type. Journal of Business research, 62(1), 61-67.
[29] Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. Expert Systems with Applications, 41(16), 7653-7670.
[30] Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013, January). Mining online user-generated content: using sentiment analysis technique to study hotel service quality. In System Sciences (HICSS), 2013 46th Hawaii International Conference on (pp. 3119-3128). IEEE.
[31] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. Icwsm, 10(1), 178-185.
[32] Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012, July). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In Proceedings of the ACL 2012 System Demonstrations (pp. 115-120). Association for Computational Linguistics.
[33] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), 267-307.
[34] Denecke, K. (2008, April). Using sentiwordnet for multilingual sentiment analysis. In Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on (pp. 507-512). IEEE.
[35] Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In Proceedings of the 31st International Conference on Machine Learning (ICML-14) (pp. 1188-1196).
[36] Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2), 15-21.
[37] Agarwal, A., Biadsy, F., & Mckeown, K. R. (2009, March). Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (pp. 24-32). Association for Computational Linguistics.
[38] Besag, J. (1986). On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society. Series B (Methodological), 259-302.
[39] Paltoglou, G., & Thelwall, M. (2012). Twitter, MySpace, Digg: Unsupervised sentiment analysis in social media. ACM Transactions on Intelligent Systems and Technology (TIST), 3(4), 66.
[40] Singh, V. K., Piryani, R., Uddin, A., & Waila, P. (2013, January). Sentiment analysis of textual reviews; Evaluating machine learning, unsupervised and SentiWordNet approaches. In Knowledge and Smart Technology (KST), 2013 5th International Conference on (pp. 122-127). IEEE.
[41] Rao, D., & Ravichandran, D. (2009, March). Semi-supervised polarity lexicon induction. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (pp. 675-682). Association for Computational Linguistics.
[42] Esuli, A., & Sebastiani, F. (2007). SentiWordNet: a high-coverage lexical resource for opinion mining. Evaluation, 1-26.
[43] Hung, C., & Lin, H. K. (2013). Using objective words in SentiWordNet to improve sentiment classification for word of mouth. IEEE Intelligent Systems, 1.
[44] Boudia, M. A., Hamou, R. M., & Amine, A. (2016). A New Approach Based on the Detection of Opinion by SentiWordNet for Automatic Text Summaries by Extraction. International Journal of Information Retrieval Research (IJIRR), 6(3), 19-36.
[45] Forrester. (2016). Think You Want To Be "Data-Driven"? Insight Is The New Data. [online] Available at: https://go.forrester.com/blogs/16-03-09-think_you_want_to_be_data_driven_insight_is_the_new_data/.
[46] Amiri, H., & Chua, T. S. (2012, July). Sentiment Classification Using the Meaning of Words. In Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence.
[47] Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 168-177). ACM.

[48] Zhang, W., & Skiena, S. (2009, September). Improving movie gross prediction through news analysis. In Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 01 (pp. 301-304). IEEE Computer Society.

[49] Zhao, J., Dong, L., Wu, J., & Xu, K. (2012, August). Moodlens: an emoticon-based sentiment analysis system for chinese tweets. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1528-1531). ACM.

[50] Hammou Fadili. Semantic Mining Approach Based On Learning of An Enhanced Semantic Model For Textual Business Intelligence Information Systems and Economic Intelligence (SIIE), IEEE conference, Feb 2020, Tunis, Tunisia.

[51] Hammou Fadili. Deep learning of latent textual structures for the normalization of Arabic writing International Society for Knowledge Organization (ISKO), IEEE conference, Feb 2020, Tunis, Tunisia.