Deep Learning Model for Sentiment Analysis in Multi-lingual Corpus.

Lisa Medrouk, Anna Pappa

To cite this version:

Lisa Medrouk, Anna Pappa. Deep Learning Model for Sentiment Analysis in Multi-lingual Corpus.. Neural Information Processing. ICONIP 2017, Nov 2017, Guangzhou, China. pp.205-212. hal-01744518

HAL Id: hal-01744518
https://hal.science/hal-01744518

Submitted on 29 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

[1]

DRAFT

[1] to cite this article : "Medrouk, Lisa and Pappa, Anna", "Deep Learning Model for Sentiment Analysis inMulti-lingual Corpus", eds (Liu, Derong and Xie, Shengli and Li, Yuanqing and Zhao, Dongbin and El-Alfy, El-Sayed M.), in Neural Information Processing, 2017,Springer International Publishing", pp. "205–212", isbn="978-3-319-70087-8"

Deep Learning Model for Sentiment Analysis in Multi-lingual Corpus

Lisa Medrouk and Anna Pappa

LIASD, Université Paris 8, France
{lm,ap}@ai.univ-paris8.fr
http://www.ai.univ-paris8.fr/

Abstract. While most text classification studies focus on monolingual documents, in this article, we propose an empirical study of poly-languages text sentiment classification model, based on Convolutional Networks *ConvNets*. The novel approach consists on feeding the deep neural network with one input text source composed by reviews all written in different languages, without any code-switching indication, or language translation. We construct a multi-lingual opinion corpus combining three languages: English French and Greek all from *Restaurants Reviews*. Despite the limited contextual information due to relatively compact text content, no prior knowledge is used. The neural networks exploit n-gram level information, and the experimental results achieve high accuracy for sentiment polarity prediction, both positive and negative, which lead us to deduce that ConvNets features extraction is language independent.

Keywords: Deep Learning, Opinion Mining, Sentiment Analysis, ConvNets

1 Introduction

In the NLP research question of 'how a machine could learn and understand human language', deep learning has a huge potential to revolutionize AI techniques and help us get further. Every language is complex, with emotion, dialects, tone. Due to the richness and complexity of human language, the problem of sentiment analysis is non trivial. Natural language sentences have complicated structures, considering the expressiveness and ambiguities. However, deep neural networks can address that without the need to produce complex engineered features [20]. In this research we focus to that 'how a machine could automatically predicting the orientation of subjective content in a multi-lingual environment', by using a deep learning model, without any prior knowledge or treatment, code-switching or language translation. Given the accelerated growth of online social networks, there is a tremendous amount of documents provided in different languages. In this article, we attempt to report a series of experiments tackling multi-language sentiment analysis using one important class of deep learning models: ConvNets [27].
Computer programs seem to understand our basic utterances, but they meet

DRAFT

measurable difficulty when it comes to understanding nuances. Aspects of language such as sarcasm or humour, are expressed in a different way in English or in French. We find these aspects in the opinions expressed in reviews. The opinions may explicitly show pleasure or displeasure, via a bag of matching words, or they implicitly convey approval or disapproval by using a simple negation or more complex ironic or even sharper cynical phrasal idioms. These features are different from one language to another. Neural networks layers may learn and predict in a multi-lingual environment using the same model, as if it worked for each language separately.

Children who experience two languages from birth typically become native speakers of both [10]. Studies show that as in monolingual development, successful acquisition of different languages for bilingual children correlates with quality and quantity of speech that they hear in each language [15]. Given this premises we decided to adopt a naïve approach, consisting on a relatively balanced exposure of three mixed languages as an input to a ConvNet, in order to study the polylingual learning process acquisition of a ConvNet for a specific NLP task.

Deep networks have already been successfully used for mono-lingual sentiment analysis, mainly English datasets, [25] predict sentiment distributions using recursive autoencoders with Recursive Neural Tensor Network [26], [24] explored an application of deep recurrent neural networks to the task of sentence-level opinion expression extraction, [6] perform sentiment analysis for short texts using a ConvNet from character to sentence level information named (CharSCNN) using two convolutional layers, [16] have used Deep Convolutional Neural Network for short text Twitter sentiment analysis. For other languages, [17] compare CNN, and LSTM in sentiment analysis of Russian tweets,[22] explored four different architectures DNN, DBN and a combined Auto Encoder with DBN for text sentiment analysis in Arabic. [7] compare supervised methods for sentiment analysis in a multilingual environment without deep network.

In this study we focus on sentiment polarity classification at a document level applied to one domain "Restaurant reviews" across three heterogeneous languages: an Anglo-Saxon, a Roman and Hellenic language. Sentiment analysis of Restaurant reviews is challenging since a single review may convey multiple sentiments related to the different restaurant's aspects. Previous studies for sentiment analysis mainly uses four feature categories for sentiment analysis: Syntactic [4], semantic [1], link-based[2], stylistic features[3]. The use of ConvNets allow us to tackle the problem without focusing on a special feature and let the model learn its appropriate features.

DRAFT

2 Overall Architecture

ConvNets are (*feed-forward*) networks suitable for detecting neighborhood correlations, making them particularly interesting for NLP tasks. ConvNets have been successfully used for sentence classification [11], product feature mining[28], semantic modeling of sentence[9] and other NLP tasks. The *one dimension* con-

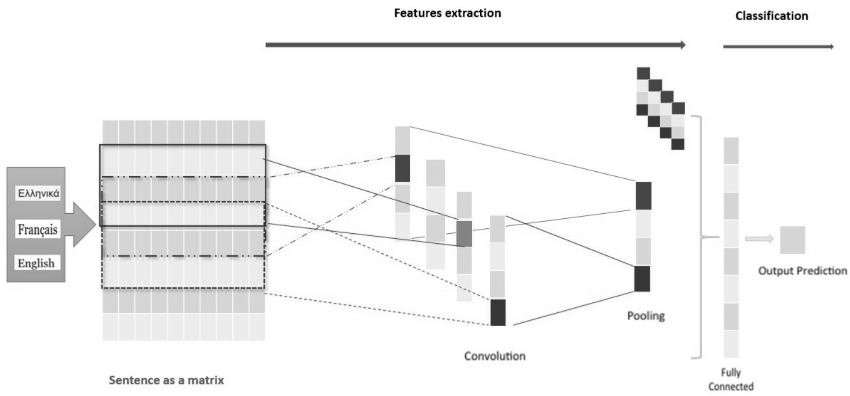

Fig. 1. The proposed architecture showing the mixed languages corpus as the input, the convolutional layer followed by the pooling layer, the dense layer and finally the output prediction.

volution involves a filter with a specific window size, sliding over a window of words in order to extract different features. This operation named *convolution operation* is applied to every n-gram of the input text. N-gram features capture shallow structure of sentences, identifying local relations between words. For a filter represented by a weight vector $W \in \mathbb{R}^{h \times d}$; a convolution on a sentence of n consecutive words expressed by the sentence matrix $X \in \mathbb{R}^{s \times d}$ can be expressed as follows:

$$c_i = f(W \cdot X_i :_{i+n-1} + b_f) \tag{1}$$

where $(\cdot)$ is the dot product between the sub-matrix sentence from i to j and the filter, b being the biais. f a non linear function and c the feature map for the filter $c \in \mathbb{R}^{s-n+1}$.
Our ConvNets architecture is close to the deep learning models presented in [8][11] for sentence classification tasks and is mainly inspired by [11] who established that using one convolutional layer with little hyperparameters tuning performs equally well than using multiple layers. Our model is therefore composed by a single convolutional layer, followed by a non-linearity in order to allow the network to learn non-linear decision boundaries. We choose the activation function *ReLU Rectified Linear Unit* defined as $F(x) = max(x, 0)$. A pooling layer is then added by simply returning the *max* value from each previous maps capturing features with the highest value and reducing dimensionality and number of parameters. For regularization we used a dropout [14] applied to a fully connected layer to avoid hidden units co-adaptation and finally a sigmod classification layer defined b the activation function $f(x) = (1 + e^{-x})^{-1} f : R \to [0, 1]$.

DRAFT

3 Datasets Description and Experimentation

For testing the proposed experiences, we constitute three labeled corpora of restaurant reviews, all extracted in their original language to avoid any kind of noise due to a translation system. We obtain 40.000 French reviews, 20.000 English and 2600 in Greek. Due to different data sets sizes we conducted our experiments with the following divisions, in order to evaluate separately the predictive classification :

- English and French with 40.000 reviews (10.000 French and positives, 10.000 French and negatives, 10.000 English and positives and 10.000 English and negatives)
- French, English and Greek with 7200 reviews, 2400 for each language equally divided between positives and negatives reviews (1.200 of each).
- Fench with 40.000 reviews. 20.000 positives and 20.000 Negatives
- English with 20.000 reviews. 10.000 positives and 10.000 Negatives
- Greek with 2600 reviews. 1300 positives and 1300 Negatives

For our study, we choose to work with raw text avoiding any pre-trained embedding as we decided to equally treat the three languages. Pre-trained embeddings might advantage the English language considering the massive datas available for training, GloVe[5] for example helps obtaining high level embedding for English corpus with no equivalence for French and Greek. The input sentences are simply tokenized and converted into a matrix $X \in \mathbb{R}^{s \times d}$, rows are d-dimensional word vectors for each token and s denote the sentence length.
The model is trained using stochastic gradient descent using Adam optimizer[13]. To find effective hyperparameters, we vary one hyperparameter at the time and kept the other ones unchanged while testing the model French-English. Every test has been run for 8 epochs. As a result we set on the following choices for all the other experiments:

Table 1. The hyperparameters selection. The experiment range for each hyperparameter and the final choices.

Hyperparameters	**Experiment Range**	**Choice**
Word vectors dimension	25-100	34
CNN number of filters	30-100	64
CNN Linear filters	3-7	5
Dropout rate	0-1	0.4
Pooling	2-4	4
Optimisation	Adam, RMSprop	Adam

DRAFT

4 Results and Evaluation

For the evaluation process, we first tested the ConvNet model for each mono language separately, then we applied the model for two mixed languages (French & English) and finally we replicated the operation for a mix of three languages (French, English & Greek) at the same time as the input process. We applied the same hyperparameters obtained on the testing phase as described in section 3 for all the evaluations.
Table 2 highlights all the evaluation results with the different input types, showing higher accuracy score with the mono language French 40K experiment, which is also the one with the largest datasets for one language input source. The most interesting results are the comparison of the bi-lingual French/English results with the mono French 20K and English 20K, as we observe that bi-lingual and mono language inputs perform almost equally 92% − 93%), showing ConvNets ability to process multiple languages at the same time without any pretreatement or language indication.
Tri-lingual performance results being at 88% is also good considering the very small datasets samples available, the comparison of this result with the small Greek dataset experiment also shows the ConvNets ability to perform equally on mono lingual and tri-lingual inputs.

We also noticed that the mono and multi-language models have in some cases perform the same qualification mistakes. For example, the following sentence :
"food was absolutely delicious. however, i felt the customer service was lacking and with further training in this area could be rectified. i dined with 14 other friends sitting at 4 separate tables, at the end of the evening our bill arrived as requested per table. they were all incorrect with items added that we didn't order. in no way...",
labelled as Positive, has been predicted as Negative with both English monolingual experiment and tri-lingual experiment. The prediction is a hard task in sentences with multiple conflicting sentiments needed more pretreatment for the different aspects of an entity as showed by [23], [12]. The same remark can be expressed for the following French sentence :
"repas tres honereux et avec peu de saveur, tres deçu du repas, cadre sympathique service accueillant beaucoup mieux a lyon et pour moins cher",
labeled Negative and predicted Positive in both models (mono French and trilingual model).
These results show that ConvNets extract features indifferently while treating mono of multi-language datasets. They also show that quantity affects the model's performance.
For baseline purposes we applied the same experiment using *Support Vector Machines* as SVM, known to be computationally efficient, robust and accurate. Results showed in Table 3 highlight high accuracy for mono and combined languages tests. ConvNets seems to be more sensible to data quantity in comparison to SVMs.

DRAFT

Table 2. Precision, Recall and F1-score for different input types *mono, bi and tri* mix languages. Showing higher accuracy score with the mono language French 40K experiment and high accuracy for bilingual inputs.

Languages	Nbr datas	Classes	Precision	Recall	F1-score	Support
English		Negative	0.95	0.90	0.92	6057
French	40000	Positive	0.90	0.95	0.93	5943
		Total	**0.93**	**0.92**	**0.92**	**12000**
		Negative	0.94	0.94	0.94	2974
French	20000	Positive	0.94	0.94	0.94	3028
		Total	**0.93**	**0.92**	**0.92**	**6000**
		Negative	0.92	0.91	0.92	2973
English	20000	Positive	0.92	0.92	0.92	3027
		Total	**0.92**	**0.92**	**0.92**	**6000**
		Negative	0.94	0.95	0.95	6738
French	45000	Positive	0.95	0.94	0.95	6762
		Total	**0.95**	**0.95**	**0.95**	**6000**
English		Negative	0.88	0.89	0.89	2972
French	7200	Positive	0.89	0.88	0.88	1077
Greek		Total	**0.88**	**0.88**	**0.88**	**2160**
		Negative	0.82	0.88	0.85	359
Greek	2600	Positive	0.87	0.80	0.84	361
		Total	**0.84**	**0.84**	**0.84**	**720**

DRAFT

Table 3. SVM accuracy results.

Languages	Nbr datas	Accuracy
English	20000	91%
French and English	40000	93%
French, English and Greek	7200	90%

5 Conclusions and Future Work

In this paper, a deep learning model is proposed for multi-language sentiment prediction of polarity. The novel approach consists on testing the learning process for three different mixed languages, simultaneously, using one text corpus input, composed by restaurants reviews, for the same model. The comparison of the results in mono and multi-language model, showed a quasi similar performance using the ConvNet Model. The same neural network is applied to mono and multi-language input. The obtained accuracy results allow us to conclude that ConvNets extract features indifferently while treating mono or multi-language data without any pretreatment or language code-switching or indication. Even though, the experiment is applied to one specific NLP task, we tend to deduce, that ConvNets perform, like in human language acquisition, in terms of learning process, with the need of quality and quantity data, able to handle multi-language text corpus without any prior knowledge or preprocessing (like segmentation) or any specific language switch-coding indication which is in our sens, the real value of this work.
In a future work we intend to explore hybrid architectural neural models and extend this work by using extra data in other NLP domains.

References

1. Turney, P.: Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics , pp. 417–424, Stroudsburg (2002)
2. Efron, M.: Cultural orientations: Classifying subjective documents by cocitation analysis. In: Proceedings of the AAAI Fall Symposium Series on Style and Meaning in Language, Art, Music, and Design, pp. 41-48 (2004)
3. Wiebe, J.,Bruce, T., Bell, R., Martin, M.: Learning subjective language IN: Comput. Linguist, pp. 277-308 ,(2004)
4. Pang, B.,Lee, L, Vaithyanathan, S.:Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10, Association for Computational Linguistics, Stroudsburg, PA, USA, 79–86 (2002)
5. Pennington, J., Socher, R., Manning, D., : GloVe: Global Vectors for Word Representation. In: Empirical Methods in Natural Language Processing (EMNLP), pp.1532–1543 (2014)
6. Dos Santos, N., Gatti,M.: Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. In: Proceedings of COLING, the 25th International Conference on Computational Linguistics: Technical Papers,pages 69-78, Dublin (2014)
7. Vilares, D., Alonso, M., Gomez-Rodriguez, C. Supervised Sentiment Analysis in Multilingual Environments. In: Information Processing & Management, (http://dx.doi.org/10.1016/j.ipm.2017.01.004) (2017)
8. Collobert, R., Weston, J., Bottou, L., Karlen, M.,Kavukcuoglu,K., Kuksa, P.: Natural Language Processing (Almost) from Scratch. In: Journal of Machine Learning Research 12 pp. 2493-2537 (2011)

DRAFT

9. Kalchbrenner, N., Grefenstette, E., Blunsom,P.: A Convolutional Neural Network for Modelling Sentences. In: ACL - Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pp.655-665 Baltimore, April (2014)
10. Garcia-Sierra, A., Rivera-Gaxiola, M., Conboy, B., Romo, H., Klarman, L., Ortiz, S., Kuhl, P.: Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. Journal of Phonetics, (2011)
11. Kim, Y.: Convolutional Neural Networks for Sentence Classification.In: Empirical Methods in Natural Language Processing, pp. 1746-1751 August (2014)
12. Ruder, S., Ghaffari, P., Breslin, J. : Deep Learning for Multilingual Aspect-based Sentiment Analysis. IN:INSIGHT-1 at SemEval-2016 Task 5 (2016)
13. Kingma, D., Ba,J. : Adam: A Method for Stochastic Optimization. IN:3rd International Conference for Learning Representations, (2015)
14. Hinton,G., Srivastava,N. , Krizhevsky,A. Sutskever,I.Salakhutdinov,R.: Improving neural networks by preventing co-adaptation of feature detectors. IN: CoRR, (2012)
15. Byers-Heinlein, K., Lew-Williams, C.: Bilingualism in the Early Years What the Science Says: LEARNing Landscapes, pp. 95-112 Autumn (2013)
16. Convolutional Neural Networks.In: Proceedings of tSeveryn, A., Moschitti, Al.: Twitter Sentiment Analysis with Deep he 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 959–962 (2015)
17. Arkhipenko, K., Kozlov, I., Trofimovich, J., Skorniakov, K., Gomzin, A., Turdakov, D.:Comparison of neural network architectures for sentiment analysis of russian tweets. In:Computational Linguistics and Intellectual Technologies,Proceedings of the International Conference Dialogue 2016
18. Chollet, F.: Keras. In : GitHub, `https://github.com/fchollet/keras`(2015)
19. Bing, L. :Sentiment Analysis and Opinion Mining. In: Morgan and Claypool (2012)
20. Denecke, K. :Using SentiWordNet for multilingual sentiment analysis. In: 2008 IEEE 24th International Conference on Data Engineering Workshop (2008)
21. Krizhevsky, A., Sutskever, I., Hinton, G.:ImageNet Classification with Deep Convolutional Neural Networks. In: Advances in Neural Information Processing Systems 25, pp. 1097–1105 (2012)
22. Sallab, A., Baly, R., El Hajj, W., Shaban, K. Deep Learning Models for Sentiment Analysis in Arabic. IN: Arabic NLP workshop, ACL-IJCNLP, The 53rd Annual Meeting of the Association for Computational Linguistics and The 7th International Joint Conference of the Asian Federation of Natural Language Processing, Beijing, China, (2015)
23. Wang, B., Liu, M. Deep Learning for Aspect-Based Sentiment Analysis. IN: DeepLF (2015)
24. Irsoy, O., Cardie,C.:Opinion Mining with Deep Recurrent Neural Networks. IN: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, pp. 720-728 (2014)
25. Socher, R., Pennington,J., Huang,E. H., Ng,A. Y. , Manning,C.:Semi-supervised recursive autoencoders for predicting sentiment distributions. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 151-161 (2011)
26. Socher, R., Perelygin,A., Wu, A., Chuang, J., Manning,C., NG,A., Potts, C., Manning,C.:Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In:Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. pp. 1631–1642 (2013)

DRAFT

27. LeCun,Y., Bottou,L. , Bengio, Y., Haffner,P.:Gradient-based learning applied to document recognition. In: Proceedings of the IEEE, 86(11), pp. 2278-2324, November. (1998)
28. Xu, L., Liu, K., Lai,S., Zhao, J.: Product feature mining: Semantic clues versus syntactic constituents.In:Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pp336346, Baltimore, Maryland, USA, June (2014)
29. Collobert, R., Weston, J.: A unified architecture for natural language processing: Deep neural networks with multitask learning. In: Proceedings of the 25th International Conference on Machine Learning, ICML , pp 160-167, New Yory (2008)

DRAFT