5th International Conference on Advanced Research Methods and Analytics (CARMA2023)
Universidad de Sevilla, Sevilla, 2023
DOI: http://dx.doi.org/10.4995/CARMA2023.2023.16474

Measuring Social Mood on Economy during Covid times: effects of retraining Supervised Deep Neural Networks

Elena Catanese[1], Mauro Bruno[1], Luca Stefanelli[2], Francesco Pugliese[1]
[1]Istat, Rome Italy [2]University of Bergamo, Italy

Abstract

Supervised Machine learning approaches are popular techniques used for sentiment analysis tasks. However, such techniques have strong limitations due to their sensitivity to the quantity and quality of the training datasets and may fail when training data are biased or insufficient. In the present study we address the impact of Covid on a deep learning classifier based on long-short term memory neural network (LSTM). This classifier is used to compute a daily sentiment index on Italian tweets with economic content, for the first five months of 2020 (more than 11 million of tweets are classified). We show how retraining the model with a set of annotated tweets containing reference to Covid increase the accuracy of the classifier. The accuracy is measured by analyzing the dynamics of the index. We will show that during pandemic the retrained index decreases coherently with most Italian economic indicators.In addition, we analyze how the training and tuning procedures (one-step, two-steps with fine-tuning) affect the daily dynamics of the index.

Keywords: *Sentiment Analysis, Artificial Neural Networks, Deep learning, Twitter data, Word Embedding Models*

1. Introduction

In this paper, we describe a new methodology to calculate a daily sentiment index that aims to depict the population mood about the economy in Italy, such as Social Mood on Economy Index (SMEI) (Catanese, et al., 2022). So far, Italian National Institute of Statistics (Istat) adopted an unsupervised procedure for estimating the social mood on economy by exploiting a lexicon-based approach (since 2018). Our initial choice fell on lexicons, due to the lack of Italian labelled datasets of tweets for sentiment analysis. Within the last few years, the number of labelled Italian datasets has increased, so we started to investigate cutting-edge supervised deep learning algorithms for sentiment analysis purposes. This work relies on a binary sentiment classifier (positive vs negative content) built by means of a Bidirectional Long short-term memory (LSTM) neural network. The training phase consists of a two-steps procedure: a) an unsupervised word embedding training model built on a unlabeled training set of Italian tweets extracted from the SMEI; b) a supervised training aimed at generating the model that computes the sentiment score.

Recently the outbreak of Covid changed structurally the content of twitter conversations. For this reason, we decided to retrain the original and previous neural network so to include a Covid related annotated dataset and to analyze accuracy improvements. In addition, we analyzed the impact on the daily index of the training procedure by splitting it into two steps. As a result, we analyzed four scenarios: original model; retrained models (both one step;two).

In this work we want to address the following question: how is the algorithm able to intercept Covid “drop” in social mood during the lockdown period.

The paper is structured as follows: in section 2 we describe LSTM methods, the different scenarios and the dataset used in our simulations. In section 3 the dynamics of the daily social mood index, within the period January 2020 - May 2020, according to the different scenarios are commented. Conclusions are drawn in Section 4.

2. Methods

2.1 Recurrent Model for Text Classification

Our classification model is a long short-term memory (LSTMs) (Hochreiter, & Schmidhuber, 1997) which are a type of Recurrent Neural Network (RNN) (Medsker, & Jain, 2001) able to process long sequences of data. In general terms, a LSTM memory cell is composed 4 gates: an input gate, an output gate, a forget gate and a self-recurrent neuro. The input and output gates control the interactions between neighboring memory cells and the memory cell itself. Whether the input signal can alter the state of the memory cell is controlled by the input gate. On the other hand, the output gate can control the state of the memory cell on whether it can alter the state of other memory cell. In addition, the forget gate can choose to remember or forget its previous state. LSTMs layers are composed of a number of cells, usually activated by means of hyperbolic tangents, which regulate the flow of information through the system of gates, in the form of sigmoid functions. RNNs are suitable for Natural Language Processing (NLP) applications thanks to their capability to connect previous pieces of information to the present tasks. In this work, the proposed model is a neural network built upon a two-layer Bidirectional LSTM (Huang, et. al., 2015). Bidirectional LSTMs (BiLSTM) are an extension of LSTMs which can improve the performance depending on the given task. BiLSTMs are basically two independent LSTMs trained on the input sequence. The first LSTM processes the sequence data forwards, whereas the second LSTM processes the sequence data backward with two separate hidden layers. BiLSTM connects the two hidden layers to an output layer. This structure provides the network both forward and backward information at each step.

The key feature of BiLSTM is the ability to capture the "context" of each word within the text in a very powerful way since one LSTM works on the left context and the other LSTM "understands" the right context. For instance, the word "bank" can assume different meanings according to its context, the left context can provide one meaning such as in "river bank". But the right context matters too, because it provides another meaning, for instance in "Bank of America", the BiLSTM can capture these different gradients of meaning by means of the double LSTM.

2.2. Training Process and Computation of daily index

The input layer of the neural network is an embedding layer, i.e., an embedding space resulting from a word-embedding model. The embedding model has been built on a corpus of SMEI tweets using the fastText algorithm (Bojanowski, et al., 2017). The output of this model is a vector space, where each word has a semantic vectorial representation. The underlying idea is that encoded words "closer" in the vector space are expected to be similar in meaning. The dimension of this vector space is set to 300. Then, it is possible to map the input layers into a two-dimensional matrix: one dimension represents the word within the corpus and the other is its vectorial embeddings representation. This matrix is the input of the first LSTM layer and the subsequent output is the input of the second LSTM stage. The use of two stacked LSTM layers allows the model to capture the semantic relationships between words and sentences (Graves, et al., 2013). The first layer has 128 cells while the second 32. Both use a hyperbolic tangent (Tanh) activation function and a dropout rate of 0.5 for regularization. For dimensionality reduction, a 1-dimensional Max Pooling layer is then adopted to convert inputs (with various lengths) into a fixed-length vector. Finally, the output layer is a dense layer, i.e., a single fully-connected layer, which is a binary classifier. It uses a Sigmoid function, which is the predicted sentiment classification of each tweet: if the resulting quantity is higher than 0.5, then the tweet is classified as positive, otherwise negative.

The training process of the classifier (which allows the sentiment scoring of the output layers) can be carried out in a unique step or can be split into two phases, where the second step consists of a fine-tuning with the scope to specialize the classifier in a specific domain. To fine-tune the model, the dataset used for the one-step procedure needs to be split.

The fine tuning of the original model was carried out by using Italian economic tweets used for SMEI, while the retrained model for Covid uses the Feel-it dataset (Bianchi, et. al., 2021).

The original two-step model is pre-trained on a dataset composed by labelled Italian tweets coming from a variety of domains. The "pre-training" set is a merge of two datasets widely used for sentiment analysis, Sentipolc (Barbieri, et al., 2016) and Happy Parents (Mencarini, et al., 2019). The tweets within the training data include political and generic tweets, whereas the test data include tweets extracted with a socio-political topic via hashtags and keywords related to #labuonascuola. The Happy Parents is a dataset of Italian tweets related to parenthood. The merged dataset is composed of 6501 labelled tweets, the 39.44% are positive and the remaining are negative tweets. Then, the model is fine-tuned on a balanced set of 900 labelled tweets (year 2016) concerning economic topics used for internal uses in Istat. In the new retrained model, Istat dataset is always used in the first-step model.The fine-tuning of the model is performed by using an additional dataset: Feel-it consisting of 2037 tweets being the positive 35.73%. These tweets were retrieved by monitoring trending topics each day between 20th August to 12th October 2020, using the Twitter API. Feel-it dataset contains 662 COVID-19-related tweets.

Both datasets have been split into a training set and a validation set according to a proportion of 80/20. The model classification accuracy has been evaluated using the F1-Score. It is worth it to notice that these datasets are not recent and date back to 2016 in the best case. The trained model is then used to predict the sentiment of a set of 11,979,986 tweets in Italian referred to the period January - May 2020 extracted from Twitter by Istat by using a set of keywords related to economy as a filter. The predicted sentiment of each tweet is then used to build the daily index, which is computed as:

$$I = \frac{N_p - N_n}{N_p + N_n}$$

Where N_p is the share of tweets classified as positive each day while N_n is the share of tweets classified as negative. The same set of unlabelled tweets, together with other SMEI tweets othe months of April and May 2021 (for a total of 15.115.421 tweets), were used for theconstruction of the embedding Fastetxt space used as input layer.

3. Results

The original model achieves 0.80 as F1-Score on the validation dataset of the first step, 0.79 of the second step, while in the first step a 0.80. The retrained model achieves its highest accuracy on the validation of the fine-tuning set 0.86, 0.83 in the one-step and slightly less on first-step 0.79. When measuring the accuracy, with respect to the validation dataset, all models show high F1-Score values.

Table 1. F1-Score in the four scenarios.

Variable	With Feel-It	Without Feel-It
One-Step	0.83	0.8
Fine-Tuning (First step)	0.79	0.8
Fine-Tuning (Second Step)	**0.86**	0.79

The indexes created using the predicted sentiment in the different scenarios are illustrated in Figure 1, 2. The original model records a breakdown since the beginning of the Covid-19 pandemic, when a strong lockdown was imposed to the country. The index shows a downwards level-shift within the period between the 7th of March and the 21st of April, i.e., a full lockdown period in the country. In this time period it seems that one-step vs two-step procedure has the only effect of an upward translation. This is due to the fact in the fine tuning a balanced dataset (50% positive, i.e. mean 0) is utilized while in the one step procedure the dataset is unbalanced (40% positive, i.e. -0.2 index on average). The index, as shown in Figure 1, has some outliers, that need a deeper analysis. As expected, the maximum of the time-series is observed on the 1st of January. We analyzed the second maximum of the sentiment index on the 6th of March and the minimum on 11th of April. While the minimum value has a consistent meaning, the positive peak seems to be a false positive.

In the minimum value, s*pesa* (expense) is again among the most common words, a further confirmation that such term is correlated with negativity. The debate is focused around *mes* (the Italian word for the European Stability Mechanism), which appears to be negatively characterized in the twitter debate, as we observe other words such as *governo* (government), euro, *debito* (debt), *tasse* (taxes) in the conversations about MES.

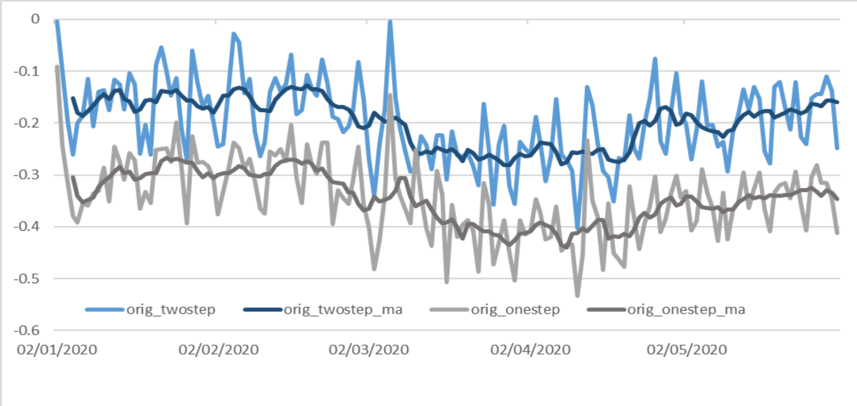

Figure 1: Original Model: daily index and weekly moving average (ma), in the two training processes (one step; two-step)

Concerning the positive peaks, we observe that when Coronavirus appears with the words *Italia*, *famiglia/e* (family), the tweets tend to be positively classified. The positivity linked to these words seems to confirm the intuition that the classifier assigns a sentiment according to the co-occurring words. Words as *famiglia/e* or *Italy* may have an intrinsic positive meaning, probably due the labelled data-set Happy Parents (more likely for family) or Sentipolc (for Italy). For this reason, we retrained the model.

In figure 2 we first observe that in this case the fine-tuning and the one-step procedure produce different results, and that the downward induced by Covid and Lock-down is more evident in the one-step trained model and is almost double than the original model or the two-step model. With respect to the original model both the re-trained models begin their descent since the 21st of February which is actually when the fear about the Italian spread of Covid began. With respect to the most positive values, we observe that in both cases, we have that second positive (or less negative) value is recorded on 13th April 2020 day on which Italy acceded to MES fund and deficit increase for economic restart was announced by the Government. It must be stresses that also the current SMEI index records such a positive value. While the minimum value in the two-step procedure is again the 11th of April in the one-step it is recorded on the 29th of March, day on which a new record of Covid deaths was witnessed. Finally, it must be stressed that the Feel-it dataset is the most negatively unbalanced dataset thus the two-step procedure lowers the average mean value of the index. Some useful additional insights can be obtained by analyzing the day over day variation of the weekly moving average (which is not a trend), as shown in Figure 3 and Figure 4.

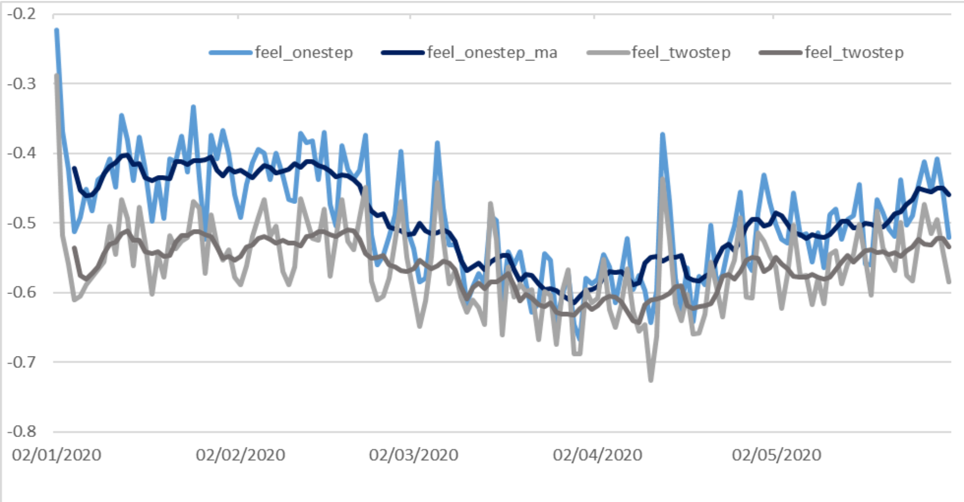

Figure 2: Retrained Model with Feel-it: daily index and weekly moving average, in the two training processes (one step; two-step)

In this case we observe that the retrained model with Feel-it shows since February 21st the same variations while they differ in the first part. In this sense the fine-tuning is able to specialize the model in Covid times as well as in the one-step procedure. When utilizing the whole dataset in the one-step procedure probably the rest of non-Covid related tweets (almost 1400) have the capability of changing the dinamics until late February.

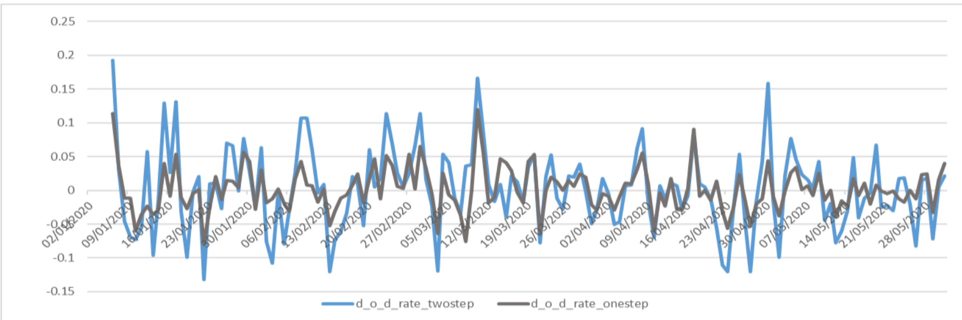

Figure 3- Day over day variation for original models.

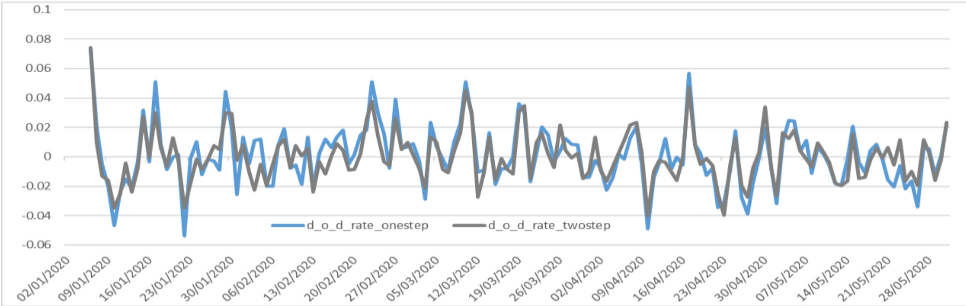

Figure 4- Day over day variation for Feel-it retrained models(lower)

If we analyse the original model, we observe that the one-step procedure smoothens the dynamics of the index as the day over day variations are always between the ones of the two-step.In the feel-it we observe an opposite behaviour.

4. Conclusions

The original index was already showing a breakdown during the Covid pandemic most likely induced by the fact that the input of the LSTM model is a word embedding model where Covid tweets have semantic relationships within SMEI tweets. However, we observed some misclassifications due to the training process. For this reason, we re-trained the model with a recent labelled dataset that contains lexical reference to the pandemic, e.g. Covid-19 terms, lockdown. Even if the size of Covid tweet is very modest we observe more coherently that the breakdown begins since late February 2020, and that the index decrease sharpens with respect to the original model as expected. It is not clear the specialization level provided by a two-step training. As a rule, it is advisable to increase the overall size of the annotated dataset in the first training step, so that fine-tuning could be performed to dynamically retrain the model on a relatively small dataset. In the present study we show that in case of Covid better results are obtained by a one-step procedure.

References

Barbieri, F., V. Basile, D. Croce, M. Nissim, N. Novielli, and V. Patti. 2016. "Overview of the Evalita 2016 SENTIment POLarity Classification Task", *Proceedings of Third Italian Conference on Computational Linguistics*.

Bianchi, F., Nozza, D., & Hovy, D. 2021. "Feel-it: Emotion and sentiment classification for the italian language", *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Association for Computational Linguistics.

Bojanowski, P., E. Grave, A. Joulin, T. Mikolov. 2017. “Enriching Word Vectors with Subword Information”, *Transactions of the Association for Computational Linguistics*, Volume 5: 135-146.

Catanese, E., M. Bruno, M. Scannapieco, and L. Valentino. 2022. “Natural language processing in official statistics: The social mood on economy index experience”, *Statistical Journal of the IAOS*, Volume 38, Issue 4: 1-9.

Graves, A., N. Jaitly & A. R. Mohamed. 2013. “Hybrid speech recognition with deep bidirectional LSTM”, *Proceedings of the 2013 IEEE workshop on automatic speech recognition and understanding*, pp. 273-278.

Hochreiter, S., & Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Huang, Z., Xu, W., & Yu, K. 2015. Bidirectional LSTM-CRF models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

Medsker, L. R., & Jain, L. C. 2001. Recurrent neural networks. *Design and Applications*, *5*, 64-67.

Mencarini, L., Hernández-Farías, D. I., Lai, M., Patti, V., Sulis, E., & Vignoli, D. 2019. “Happy parents’ tweets”. *Demographic Research*, Volume 40, Article 25: 693-724.