



Sentiment Analysis of Facebook Post Comments Using Advanced BiLSTM and BERT Models

Md. Takveer Ahmed

Post Graduate, Department of Computer Science and Engineering, Faculty of Science and Information Technology, Daffodil International University, Dhaka, Bangladesh
Email: 222-25-160@diu.edu.bd

ABSTRACT

The fast expansion of social media has transformed the ways people communicate their thoughts, feelings, preferences and attitudes about public issues, products, services and social events. Facebook is a popular social network site that has a significant number of comments from users, which provide valuable information for public mood evaluation. However, sentiment detection in Facebook comments is difficult due to casual language, spelling mistakes, sarcasm, code-switching, emoticons, acronyms and context related expressions. These language distinctions are often missed by traditional sentiment analysis methods. Here, we propose a comparative deep learning architecture for sentiment analysis on Facebook post comments based on upgraded Bidirectional Long Short-Term Memory (BiLSTM) network and Bidirectional Encoder Representations from Transformers (BERT). The proposed approach covers data collection, preprocessing of data, sentiment annotation, model development, training and comparative analysis. BiLSTM model aims to learn sequential dependency in both directions. BERT use bidirectional contextualized representations to capture the contextual meaning of the words. Model performance should be evaluated by accuracy, precision, recall, macro-F1 score and confusion matrices. The paper also addresses issues of class imbalance, interpretability of the model, computational cost, privacy and ethical data handling. The study constructs an empirical framework to determine their relative efficacy under controlled experimental conditions, rather than assuming one architecture is universally better. The proposed study contributes to social media analytics by identifying methodological aspects for developing more reliable and context-sensitive sentiment classification algorithms.

Keyword: Sentiment analysis, Facebook comments, BiLSTM, BERT, deep learning, natural language processing, social media analytics, opinion mining.

1. Introduction

Social media platforms have come to be important instruments for communication, public discourse, entertainment, marketing and information exchange. Users publish, like and comment to share their views on politics, companies, schools, public services, social movements and personal experiences. The chats provide an enormous amount of unstructured textual data from which collective thoughts and emotional patterns can be mined. Sentiment analysis, often known as opinion mining, is a Natural Language Processing (NLP) problem that allows to automatically identify the polarity or subjective orientation of text utterances. Comments on Facebook are usually immediate responses to posts, adverts, news stories, organizational announcements, and public debates, and they can be quite useful for sentiment research. These comments can give researchers insight into audience reactions, businesses consumer input, and lawmakers public sentiment on specific issues.

But sifting through thousands of comments manually is time consuming, expensive, and hard to scale. Automated sentiment categorization presents a feasible technique to mine useful information from social media interactions. However, there are a number of linguistic and computational problems in sentiment analysis of Facebook comments. They also often use casual spelling, grammar errors, acronyms, emoticons, hashtags, repetitive punctuation, and multilingualism. A typical mixed attitude answer would be: “Good service but bad delivery!” Similarly, “Excellent!” While “another delay!” could sound like encouragement, it could also be sarcasm expressing discontent. These examples show that sentiment categorization is more than simply identifying positive and negative languages. Text classification is mostly based on traditional machine learning methods such as logistic regression, naive bayes and support vector machines. These approaches can obtain competitive results when correct features are employed but their success is strongly dependent on the quality of feature engineering and representation.

Some of these restrictions are overcome by algorithms based on deep learning that can learn useful representations from text input. Bidirectional Long Short-Term Memory (BiLSTM) networks are recurrent neural models which can learn sequential dependencies. BiLSTM can interpret the text in both forward and backward directions and so it can hold the information of previous and succeeding words. This is useful in circumstances when the sentiment depends on the order of words and the words around. One of the most important advances in contextual language modeling was BERT, developed by Devlin et al. (2019). BERT utilizes the Transformer architecture and self-attention mechanisms to construct context-aware word embeddings. The representations are different from normal word embeddings in the sense that they can change depending to the context of the word. This is quite useful, particularly when dealing with ambiguous statements, negations and complex conversations. In this research, we offer a comparative technique for sentiment analysis of Facebook comments using state-of-the-art BiLSTM and BERT models. We are interested in three forms of sentiment: positive, negative and neutral. The main objective is to explore how to create and assess such structures in an effort to determine the optimal technique for a particular data collection and computational environment.

1.1. Research Objective

Sentiment analysis has been improved a lot but there are still numerous difficulties with Facebook comments that need to be addressed. Pre-processing is complicated by informal language and diverse linguistic patterns. Models are biased toward the majority classes due to the uneven distribution of sentiment. Further comparison of BiLSTM with BERT without considering data splits, preprocessing techniques, training settings and assessment measures may lead to wrong findings.

The study aims to give a replicable comparison technique to examine the effectiveness, limitations and practical application of complex BiLSTM and BERT models to Facebook comment sentiment classification.

1.2 Objectives of the study State the objectives of the study

1. Create a system that can automatically gather and structure and tag Facebook comments for sentiment analysis.
2. Design and build a complicated BiLSTM model to classify the comments into positive, negative and neutral.
3. Fine-tune the pre-trained BERT model for the same classification job.

4. Evaluate models on accuracy, precision, recall, macro-F1 score and per class performances.
5. Consider the influence of linguistic variety, class imbalance and computational constraints on the performance of the model.
6. Understand the ethical and practical consequences of automated social media comment analysis.

1.3 Research Questions

The study asks:

RQ1: How successfully do BiLSTM and BERT characterize sentiment on Facebook comments? RQ2: Which model performs best in the positive, negative and neutral classes? –

RQ3: How does informal language, denial, sarcasm and code flipping affect sentiment classification?

RQ4: What are the trade-offs between prediction performance and computational cost for practical deployment?

2. Literature Review

2.1 Sentiment analysis and opinion mining

Sentiment analysis is the problem of detecting the views, feelings, and attitudes of evaluation conveyed in textual data. Liu (2012) states describes sentiment analysis as the task of extracting subjective information and identifying the polarity of sentiments. The sentiment analysis can be done at the document, sentence or aspect level depending on the study objective. Document level classification gives sentiment for the entire text whereas sentence level analysis gives polarity to individual sentences. Aspect based sentiment analysis takes it one step further by linking opinions to specific characteristics or features of a product, service or issue. People might comment on Facebook about the quality of the product positively but adversely about delivery. The sentiment classification is typically done in positive, negative and neutral classes. The positive emotion is approval or satisfaction. The negative sentiment is disapproval or discontent. The neutral sentiment is no clear positive or negative tendency. However, precise annotation criteria are required to distinguish between neutral and weakly polarized claims. For a survey of the field of computational subjectivity and sentiment we refer the reader to Pang and Lee (2008). Results emphasize the role of representation, classification methods and context interpretation. These limits are especially important when considering social media apps, as short comments typically do not contain any grammatical or contextual information.

2.2 Classical Machine Learning Techniques

The earlier works on sentiment classification primarily employed bag-of-words representation, TF-IDF, sentiment lexicons and hand-crafted linguistic features. The representations were paired with classifiers such as naïve Bayes, support vector machines and logistic regression. Traditional models are still useful because they are often efficient, interpretable in some contexts and acceptable as baseline systems. But the bag of words representations tend to lose all the information about order of the words. They may also have trouble with contextual ambiguity, sarcasm and long range linkages between sentences. For example, “The product is good”. “The product is not good”. They are using similar phrases, but actually they are different in attitude. If a representation does not convey negation well, both comments may be misclassified. Deep learning architectures are providing alternate ways to gain sequential and contextual knowledge.

2.3 Recurrent Neural Networks using LSTM

RNNs are a type of neural networks that are capable of analyzing sequences of data. RNNs have a hidden state that stores information from earlier sections of the sequence. However, conventional recurrent networks may suffer from vanishing or exploding gradients, which prevent them from learning long-term dependencies. This difficulty of learning long-range dependencies was addressed by Hochreiter and Schmidhuber (1997) proposing Long Short-Term Memory (LSTM) networks. LSTM uses memory cells and gating mechanisms to control information flow. The main gates are the input gate, the forget gate and the output gate. The forget gate decides what information to discard, the input gate decides what fresh information to add to the memory cell and the output gate decides what information to give to the hidden representation. These techniques enable LSTM to learn useful sequential patterns, while loosening some of the limitations of usual recurrent networks. Order of words is typically important for emotional meaning and so LSTMs have been used for sentiment categorization. But a regular LSTM can only read information in one direction, which means it can only access words that appear later in the sequence.

2.4 Sentiment Classification using Bidirectional LSTMs

A bidirectional LSTM is a combination of two recurrent networks that move in two directions: left to right and right to left. They mix their hidden representations to encode the information in both directions. Such structure can help uncover sentiment expressions in Facebook comments that are seen differently dependent on the words around. The overall sentiment is mostly due to the duality of "The camera is pricey but good. Bidirectional processing utilizes contextual information in both directions of a word. An advanced architecture for BiLSTM may include pretrained word embeddings, subword aware tokenization, dropout, attention mechanisms and a dense classification layer. Although attention can be used to provide weight to informative tokens, it should not be seen as a full explanation of the model's predictions. BiLSTM also has its restrictions. In highly parallelisable designs training may be slower than sequential computing. The performance depends on the quality of the tokenization, the amount and variety of the training dataset and the ability of the embedding to capture informal or multilingual expressions.

2.5 Transformer Architecture - BERT

The Transformer architecture was introduced by Vaswani et al. (2017) to characterize the interactions of the elements in a sequence by attention processes instead of recurrence. By self-attention, a token can inject into the context of permissible attention information from other tokens. BERT is a language representation model that uses the bidirectional Transformer, developed by Devlin et al. (2019). BERT was pre-trained in the original setup with masked language modeling and next sentence prediction. BERT learns contextual representations fine-tuned for downstream tasks. For sentiment analysis, a pretrained BERT encoder can be used with a classification head on top and finetuned on labeled comments. Fine-tuning is the process of training the parameters of the model to find the patterns associated with the sentiment categories in the target data set. BERT can be helpful when the attitude is dependent on context, or when negation, word relations and subtle semantic alterations are involved. But the computational costs could be significant, especially for huge data sets or long input sequences. Comments that are dominated by Bengali, Banglish or other language variants could likewise provide a problem to the standard English BERT models, unless an appropriate multilingual or language-specific model is utilized.

2.6 Research Gap

Earlier works show the importance of recurrent neural networks and transformer based models for text classification. However, general model design statements can't be used to decide Facebook comments' architecture. The performance may depend on empirically selected parameter settings in the dataset, the suitability of the annotations, the distribution of the various classes, the language used and the pre-processing and computation power. A reasonable comparison will be conducted on a comparable training, validation, testing partition of a dataset between BiLSTM and BERT, and must be made on the same data set. It should also report class statistics, rather than accuracy. Certain situations such as sarcasm, mixed emotions, negation and code switching should be taken special consideration in the study. We intend to fill this methodological gap in this paper by developing a reproducible paradigm of an experiment for comparative assessment. It provides an approach to evaluating the design of your formal design, not just a catchy phrase that another design has already proven more successful in practice.

3. Research Methodology

3.1 Research Method

In order to score two deep learning models for sentiment classification, researchers were carried out using a quantitative experimental approach. BiLSTM and BERT are the primary experimental conditions while the model architecture is the independent variable. Dependent variables are categorization measures, processing needs and error patterns. Two models need to be trained on the same labeled data set and evaluated on the same data set. Another baseline would be to just use a standard machine learning classifier, such as logistic regression with TF-IDF features. This would allow to see if deep learning models offer any pragmatic benefits over simpler models.”

3.2 Data Gathering

The data is to be either Facebook comments on a very specific series of public posts, which are relevant to the research objective, or public postings on Facebook. Sample topics include product reviews, University news, public service discussions or comments to articles/issues in the media or community. Data collection will be conducted in compliance with legal rules, platform agreements, permits and ethical research guidelines. Researchers should not presume that information is free of restrictions on its collection and publication just because it is public. If direct collecting is not allowed, instead collect the data from a fully licenced dataset or a sanctioned research export. The data must include the following: Sentiment label, a unique anonymous identification for the record and content in the comments. Any other metadata that would be required for the focus of the study or legally and morally allowable should be collected. Several thousand comments might be a good starting point for the target data set, but this sample size should be based on the class distribution to be expected, the difficulty (or cost) of annotating the comments, and the required level of statistical precision. There need to be sufficient samples of each sentiment type, and it should not be regarded as a representative sample of all users of Facebook.

3.3 Annotation of Sentiments

All comments should fall into one of the 3 categories:

Positive expressions include those of approval, satisfaction, praise and a positive evaluation; Negative expressions include those of disgust, wrath, disappointment, and a negative evaluation; and Neutral expressions include a statement of fact or a description about a fact, but no explicit judgment of whether it is positive or negative.

Comments containing multiple sentiments should be labeled with a documented rule (e.g. dominating overall sentiment) or eliminated from the three-class task if their polarity cannot be reliably determined.

A portion of the data, ideally the entire data set if resources permit, should be labeled separately by at least two skilled annotators. Cohen's kappa can be used to measure agreement between two annotators, but a suitable multi-rater statistic can be utilized when more than two annotators are involved. Disputes should be resolved through a documented adjudication process. The team needs to pay special attention to sarcastic comments, rhetorical questions, emoticons, and unclear utterances. The annotation rules should define how to handle such cases prior to the training of the model.

3.4 Data pre-processing

Preprocessing should keep information pertinent to sentiment and remove noise that is irrelevant. The recommended approach consists of the following:

1. Remove duplicate records and invalid/commentless records.
2. Whitespace and Unicode normalisation handling.
3. Determination of language or language mix of multilingual comments.
4. Make general regularisations of some spellings where this does not affect meaning.
5. Managing URLs, user Mentions, hashtags and Emoticons based on rules established.
6. Ensure that phrases such as "not", "never" and their language-specific counterparts are noted.
7. Tokenizing comments with the tokenizer of the relevant model.

Don't scrub aggressively. Eliminating all punctuation and repeated characters may lead to loss of sentiment signals. For instance 'Excellent!!' and 'Excellent' may be learned in the same way, but at some varying level of emotional intensity. The preprocessing technique must be also be architecture aware. BiLSTM uses a uniform sequence representation and padding scheme, but BERT uses its own subword tokenizer, special tokens and attention masks. Therefore, the raw data and the partitioning of this data must be uniform and the partitioning for different models must be performed separately. As such, the raw data and divisions of such data must be commensurate and the divisions of such data for different models must be independently completed.

3.5 Splitting the dataset

The data set should be divided into three sets: training, validation and test. A possible partitioning is 70% training, 15% validation and 15% testing. The number of sentiment classes within each partition should be approximately the same after stratified split. A grouped split may be preferable when comments come from numerous posts. In such an approach, the comments of the same post are grouped in a single partition, reducing the chance of a model to learn post-specific language that appears in the test set. Before you split them, you should detect any near-duplicate comments or duplicated templates. The test set must not be touched during model selection or hyperparameter tuning. If multiple experiments are done, the same test partition should be used for the final comparison.

3.6 Advanced architecture of BiLSTM model

The architecture of the proposed BiLSTM consists of an embedding layer, a bidirectional recurrent layer, an optional attention mechanism, a dropout layer and a dense output layer. First, each remark is converted into a sequence of tokens. Tokens are mapped to dense vectors using pre-trained embeddings or an embedding layer constructed on the training

corpus. Pretrained embeddings can help improve performance by accurately reflecting the language and vocabulary of the dataset. The second step is to process the sequence using the forward and reverse LSTM networks. Their secret states are connected such that they can gather information in both directions. Then, an attention method can be used to assign different weights to the created token representations, so that the classifier can pay attention to the most informative parts of the comments. Third, we employ dropout to reduce overfitting. The last representation is fed to a dense layer, which outputs a softmax distribution estimating the probability of each sentiment category.

The model outputs for three classes:

$$[P(y=c|x), \text{quad } c \in \{\text{positive, negative, neutral}\}.]$$

The expected category is the one with the highest estimated probability.

The parameters of the model should be set by validation studies. Possible options are embedding dimensions of 100-300, hidden dimensions of 64-256 to each direction, dropout rates of 0.2-0.5 and batch sizes of 32-128. These are experimental starting points not necessarily perfect arrangements.

4. Evaluation Metrics

Accuracy is the proportion of the total comments that are correctly classified.

$$[\text{Accuracy}] = \frac{TP + TN}{TP + TN + FP + FN}$$

For multiclass classification, calculate class-specific true positives, false positives, false negatives and true negatives from the confusion matrix.

Precision is the fraction of the class that is predicted to be in that class.

$$[\text{Precision}] = \frac{TP}{TP + FP}$$

Recall is the ratio of the number of correct identifications in a class.

$$[\text{Recall}] = \frac{TP}{TP + FN}$$

The F1-score is the harmonic mean of precision and recall:

$$[F1 = 2 \times \frac{[\text{Precision}] \times [\text{Recall}]}{[\text{Precision}] + [\text{Recall}]}]$$

Precision + Recall

Because there may be a varying number of positive, negative and neutral comments within Facebook datasets, macro-F1 is the primary metric to evaluate it. It calculates F1 scores separately for each of the three classes and then averages the scores equally among the three classes. Weighted-F1 values can also be provided to take the prevalence of classes into account.

For those who like to find complaints other options are available for a model with high accuracy but low recall for negative comments. Foreseeing this, this study should include all metrics and explore error analysis matrices to understand the design of errors made by each model.

5. Proposed Experimental Analysis

5.1 Comparative study

The two designs need to be evaluated with the same held-out set. The reporting system for this experiment is outlined below: For now the results are on hold, while the models are trained and tested. Methods: BERT accuracy and Advanced BiLSTM. For sale. . . . For measuring instruments.

Precision Macro To be measured.

Recalling Macros. To be measured. To be gaged.

Macro-F1: To be measured| To be measured F1 for positive class In F1| To be evaluated
 Positive class| To be determined.
 The minimum grade of F1| To be announced. For the measure.
 Neutral F1 | For measuring. TBD.
 The number of days for training will be announced. To be estimated.
 Latency for inference To be evaluated. To be measured
 Peak memory utilization | To be measured |

The final evaluation needs to determine the overall best-performing model considering class-specific weaknesses and computational expense. When the difference in macro-F1 is minor, repeated runs or bootstrap confidence intervals can be used to assess if the observed difference is consistent. It can be used when the assumptions of an appropriate statistical test are satisfied.

5.2 Anticipated BiLSTM Performance

BiLSTM can be a good compromise between complexity and accuracy of the model, especially if the dataset is of medium size and the word representations are appropriate for the language. It can model sequences, which allows it to recognize local and long-range dependency. Its bidirectional processing combines information from both sides of each token. Performance can suffer from vocabulary coverage, distinctive phrasing, spelling differences, and lack of training examples. An embedding trained on a small data set will not contain new slang, foreign terms. Paying attention and better embedding can help to alleviate some of these problems, but do not provide guaranties of accurate categorization.

5.3 How BERT Should Behave

BERT is built to learn contextual representations through pretraining on a broad scale. This might allow you to recognize attitude-laden phrases that depend on the context and to better understand negation, contrast and confusing claims. Fine-tuning further permits the adaption of pretrained representations to the target dataset. BERT might need greater computational resources and hyperparameter adjustment. Domain mismatch with pre-training data and Facebook comments can limit its usefulness. Even with contextual representations, irony and cultural references will still probably be impossible to comprehend. These are theories based on architectural elements, not results of the proposed experiment. Performance is evaluated by a controlled evaluation.

5.4 An Error Analysis

A proper comparison would be to compare faulty estimates to not only aggregate data.

Standard errors:

- Negation: "The service is bad."
- Sarcastically, "great, another problem!"
- Mixed opinions "Design is gorgeous but quality is bad.
- Code switching: a statement that switches between English and Bangla or other language.
- Emoji dependence: remarks where the majority of the sentiment is represented by emojis.
- Context-dependent: comments such as "Exactly!" that depend on the original post.
- Ambiguous neutrality corresponds to the circumstance when annotators have conflicting interpretations of factual statements.

Researchers should study errors of each class and uncover repeated patterns. As an example, the original post that is utilized as contextual knowledge can be deemed to be ethical and consistent with model design (where appropriate).

6. Discussion

The suggested comparison is useful in computer research as well as in the practical analytics of social media. BiLSTM and BERT are two different ways of representing language: BiLSTM learns sequential patterns using token embeddings, while BERT builds contextual representations via Transformer self-attention and large-scale pre-training. The relative benefits of each model will likely be dictated by the properties of the dataset. For tiny domain specific datasets, a well-regularized BiLSTM with proper pre-trained embeddings can achieve a competitive performance with reduced deployment costs. In datasets with higher language diversity, BERT may benefit from the contextual knowledge learned during pre-training. These possibilities should be truly considered, not treated as universal norms. Also, the comparison shows that precision alone is not enough. Suppose we have a dataset with many positive comments and few negative ones. It is also possible that the model classifies most of the comments as good yet still scores well. But it might not realize negative input. This issue is also evident from the Macro-F1, class-specific recall and confusion matrices. Another important issue is the link between comment mood and post context. A short answer may not indicate a fair opinion. For example, “I agree” cannot be fully understood without knowing what the user agrees to. If the aim of the research is to classify sentiment expressed at the time of the meeting, then the first post or a particular conversational environment may need to be addressed. If you only have comment text, you should be aware of this constraint. Language variation is particularly significant for research in Bangladesh and other multilingual settings. Users can write in Bangla, English, Banglish and regional phrases in one remark on Facebook. Recurrent and Transformer models can struggle with transliteration, non-standard spelling and culturally unique idioms. A multilingual model is not always correct for all language varieties, it should be tested with typical labeled data. Results from such an exercise could assist influence decisions on model deployment. If BERT works better and more stably in macro-F1, the additional computational cost of using BERT may be worthwhile for applications where higher classification accuracy is needed. If a BiLSTM can perform equally well with much fewer resources, it could be a better option for low-cost or resource-constrained applications. Sentiment analysis can help firms identify common problems, track public reactions and summarize feedback from their audience. However, machine generated categories are not to be treated as definitive measures of popular mood. Social media users may not be representative of the general population, and exposure to comments may be altered by platform algorithms, moderation and participation habits.

7. Ethical Considerations and Limitations

Sentiment analysis on social media data should be performed considering the ethical issues. They should collect only data that they are permitted to use, should minimize the collection of personal data, should anonymize records where appropriate and should not publish identifying statements without good reason and adequate precautions. Appropriate consideration should be given to compliance with relevant institutional review processes and platform policies. May lead to Bias in Sentiment Classification. A model trained on one society might misinterpret dialect, slang, political speech, comedy, or culturally-specific language used in another. If criticism was misclassified as neutral or positive, an analysis could reach a biased conclusion. When disagreement is met with neutrality it is taken as hostility, and may indeed fan the flames of negative feeling. It's a terribly flawed paradigm. First, this work does not include any experimental dataset or the results of trained models. And so the performance of the comparison is unproven. Second, the sentiment labels of three classes are too simple to describe complex emotional expressions, and cannot contain all subtle viewpoints. Third, the generalizability of linguistic and domain differences may limit the conclusions. (4) Sarcasm and comments that depend on context are hard to classify. Finally, the computational

requirements could prevent us from deploying ever more complex models in real time applications. Future research directions include aspect-based sentiment analysis, multi-lingual and language specific models, domain adaptability, semi-supervised learning, and explainability approaches. The temporal fluctuations of emotion as well as the effect of post-context on the classification performance should be investigated in future studies.

8. Conclusion

We provide a methodological framework for sentiment analysis of Facebook post comments using advanced BiLSTM and BERT models. The study explores the challenges of informal language, contextual ambiguity, class imbalance, code flipping and understanding of brief social media comments. It provides a defined experimental design with data collection, annotation, pre-processing, model construction, controlled testing, and thorough evaluation. BiLSTM provides a sequential modeling technique that captures information in both directions, while BERT provides contextual representations via Transformer-based pretraining. Both designs have potential advantages and their success depends on the data set quality, linguistic features, model set up, and computational resources.

A suitable comparison must have accuracy, macro precision, macro recall, macro-F1, class specific data, confusion matrices, and computational metrics. It should also provide error analysis to expose issues that aggregate measurements may mask. Without actual studies, it is technically inaccurate to say that one model achieves a certain accuracy or clearly outperforms the other. The main contribution of this work is a research technique for evaluating algorithms using deep learning for Facebook sentiment analysis that is reproducible. Empirical evidence to decide the optimal model for a given social media analytics assignment would be provided by using the framework on a properly licensed and well-annotated dataset.

References

1. Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019) 'BERT: Pre-training of deep bidirectional transformers for language understanding', Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, pp. 4171–4186. doi: 10.18653/v1/N19-1423.
2. Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', Neural Computation, 9(8), pp. 1735–1780. doi: 10.1162/neco.1997.9.8.1735.
3. Liu, B. (2012) 'Sentiment analysis and opinion mining', Synthesis Lectures on Human Language Technologies, 5(1), pp. 1–167. doi: 10.2200/S00416ED1V01Y201204HLT016.
4. Manning, C.D., Raghavan, P. and Schütze, H. (2008) Introduction to Information Retrieval. Cambridge: Cambridge University Press.
5. Pang, B. and Lee, L. (2008) 'Opinion mining and sentiment analysis', Foundations and Trends in Information Retrieval, 2(1–2), pp. 1–135. doi: 10.1561/15000000011.
6. Pennington, J., Socher, R. and Manning, C.D. (2014) 'GloVe: Global vectors for word representation', Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543. doi: 10.3115/v1/D14-1162.
7. Vaswani, A. et al. (2017) 'Attention is all you need', Advances in Neural Information Processing Systems, 30.
8. Zhang, L., Wang, S. and Liu, B. (2018) 'Deep learning for sentiment analysis: A survey', Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253. doi: 10.1002/widm.1253.