Inferring the actitude toward self-driving cars from tweets using an NN-based supervised classification algorithm

1 Description

Internet can be used as one important source of information for machine learning algorithms. In particular, twitter has become a valuable source of information for companies and social agents. Supervised classification methods are used to identify and classify the reaction to products and events [1, 2, 3, 4, 5].

2 Objectives

The goal of the project is to infer the reactions of users to self-driving cars from their tweets. The problem can be transformed into classifying the tweets in different classes according to the sentiment they express about self-driving cars. A database of tweets will be used for the analysis [1]. An NN-based supervised classifier should be created to predict the sentiment contained in each tweet.

The project should apply Neural Networks for: I) Find suitable feature representations for this problem that are very usable for other ML classifiers, OR, II) Implement NN-based classifiers for this problem, OR III) The combination of I and II (e.g., using an RBM to find the features and a Multi-layer Perceptron to classify the problem using the extracted features). In case II), the students are free to decide which feature representation is more appropriate for the data. In case I), they can use any classifier with the NN-based features.

The student should: 1) Design any preprocessing of the web pages in the dataset (see suggestions below); 2) Define and learn the classifier using the training data. 3) Design the validation method to evaluate the accuracy of the proposed classification approach. 4) Answer to the following questions in the report:

- What class of problems can be solved with the NN? (e.g., supervised vs unsupervised problems)
- What is the network architecture? (e.g., type and number of layers, parameters, connectivity, etc.).
- What is the rationale behind the conception of the NN?
- How is inference implemented? (e.g., How is the information extracted from the network?). Type of prediction or type of inference process.
- What are the learning methods used to learn the network ? Algorithms used for learning the network.

As in other projects, a report should describe the characteristics of the design, implementation, and results. A Jupyter notebook should include calls to the implemented function that illustrate the way it works.

---

[1] The "Twitter sentiment analysis: Self-driving cars" dataset can be downloaded from `https://www.crowdflower.com/data-for-everyone/`.

3 Suggestions

- Formalize the task as a classification problem where the classes of the problems are: very positive, slightly positive, neutral, slightly negative, or very negative.

- Use the `pattern` Python package `https://github.com/clips/pattern` to extract primary features from the tweets.

- See example `https://github.com/clips/pattern/blob/master/examples/ 03-en/07-sentiment.py` to see tools for sentiment analysis in `pattern`.

- Use NNs to extract more elaborated features from the original features.

- Implementations can use any other Python library.

- If classes are not well balanced you may use performance measures different to the accuracy.

References

[1] Nicholas Beauchamp. Predicting and interpolating state-level polls using twitter textual data. *American Journal of Political Science*, 61(2):490–503, 2017.

[2] Aron Culotta. Training a text classifier with a single word using twitter lists and domain adaptation. *Social Network Analysis and Mining*, 6(1):8, 2016.

[3] Jimmy Lin and Alek Kolcz. Large-scale machine learning at twitter. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, pages 793–804. ACM, 2012.

[4] Marco Pennacchiotti and Ana-Maria Popescu. A machine learning approach to twitter user classification. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, volume 11, pages 281–288, 2011.

[5] Agus Sulistya, Abhishek Sharma, and David Lo. Spiteful, one-off, and kind: Predicting customer feedback behavior on twitter. In *International Conference on Social Informatics*, pages 368–381. Springer, 2016.