A Comparison of Machine Learning Models to Predict the Outcome of Swiss Federal Votes Using the Text of the Official Voter Pamphlets

Daniel Müller [1,2] | (1) Stellus.ch, Liestal, Switzerland, (2) Alumni, EPFL Extension School, Lausanne, Switzerland

Aim

Comparing the accuracy of being able to predict the outcome of Swiss Federal votes based on four types of machine learning models. The models are based on the text of the official voter pamphlets in german containing the bill and the arguments in favor of or contrary to them, published by the Federal Chancellery (1) (pictures of pamphlets in Figure 1).

Figure 1: Examples of voter pamphlets

Methods

Since 1978 an official voter pamphlet has been sent to every eligible Swiss voter a few weeks prior to every popular vote. From these pamphlets 357 records and a bag of words (BoW) were extracted in early 2017 and matched with the official outcome of the votes. The data set was then split by a ratio of 85% to 15% (303 to 54 records). A logistic regression model (LRM) was trained and tuned (hyper-parameters in Table 1). The accuracy of the LRM was tested on the 14 upcoming votes since May 2017 until November 2018. Details of the LRM are available at www.stellus.ch (2).

In late 2018 three artificial neural networks (ANNs) were developed based on the same records as for the LRM. The data was thoroughly cleaned, and non-lexical words were removed. The ANNs are based on architectures used in sentiment classification (3,4). Keras, a python deep learning library (5) containing a tokenizer able to count words by frequency was fitted on the training set, transformed and zero-padded to sequences of the same length in the validation and test set. The input layer consists of randomly initialized word embeddings, followed by either a single layer of a 1D convolution and max-pooling (sCNN), three layers of 1D convolutions (tCNN) or a long short-term memory layer with 100 units (LSTM model). All ANNs contain one or two dropout-layers (architectures in Figure 2). The records were split as for the LRM and used to train, validate and tune the models (hyperparameters in Table 2). The accuracy has been tested on the latest 14 popular votes. To compare to the majority class, a dummy classifier (DC) was applied on the data sets. The results have also been compared to opinion polls retrieved by gfs, Bern (6).

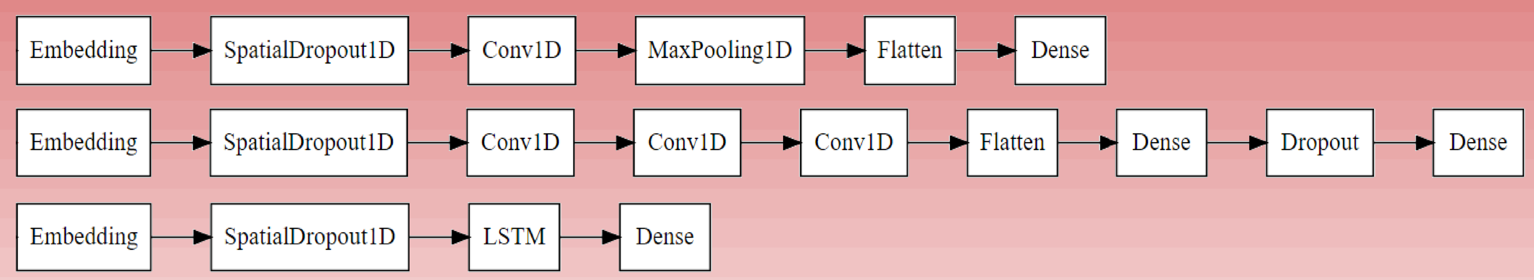

Figure 2: Architectures of the single convolution neural network (sCNN), three convolutions neural network (tCNN) and long short-term memory model (LSTM)

Table 1: Hyperparameters of the logistic regression model (LRM)

Model	Size of bag of words	L2-penalty	Learning rate	Iterations
Logistic regression (LRM)	50,628	100	5e-6	1000

Table 2: Hyperparameters of the artificial neural networks (ANNs)

Model	Trainable parameters	Optimizer	Dropout	Vocabulary size	Embedding vector length
Single convolution (sCNN)	56,873	rmsprop	0.0	300	30
Three convolutions (tCNN)	2,542,905	adam	0.1	1000	150
Long short-term memory (LSTM)	1,020,501	adam	0.8	4500	200

Results

The sCNN is the most accurate model on the validation and test set followed by the LRM, tCNN and LSTM. (Accuracies in Figure 3, including F1- scores in Table 3).

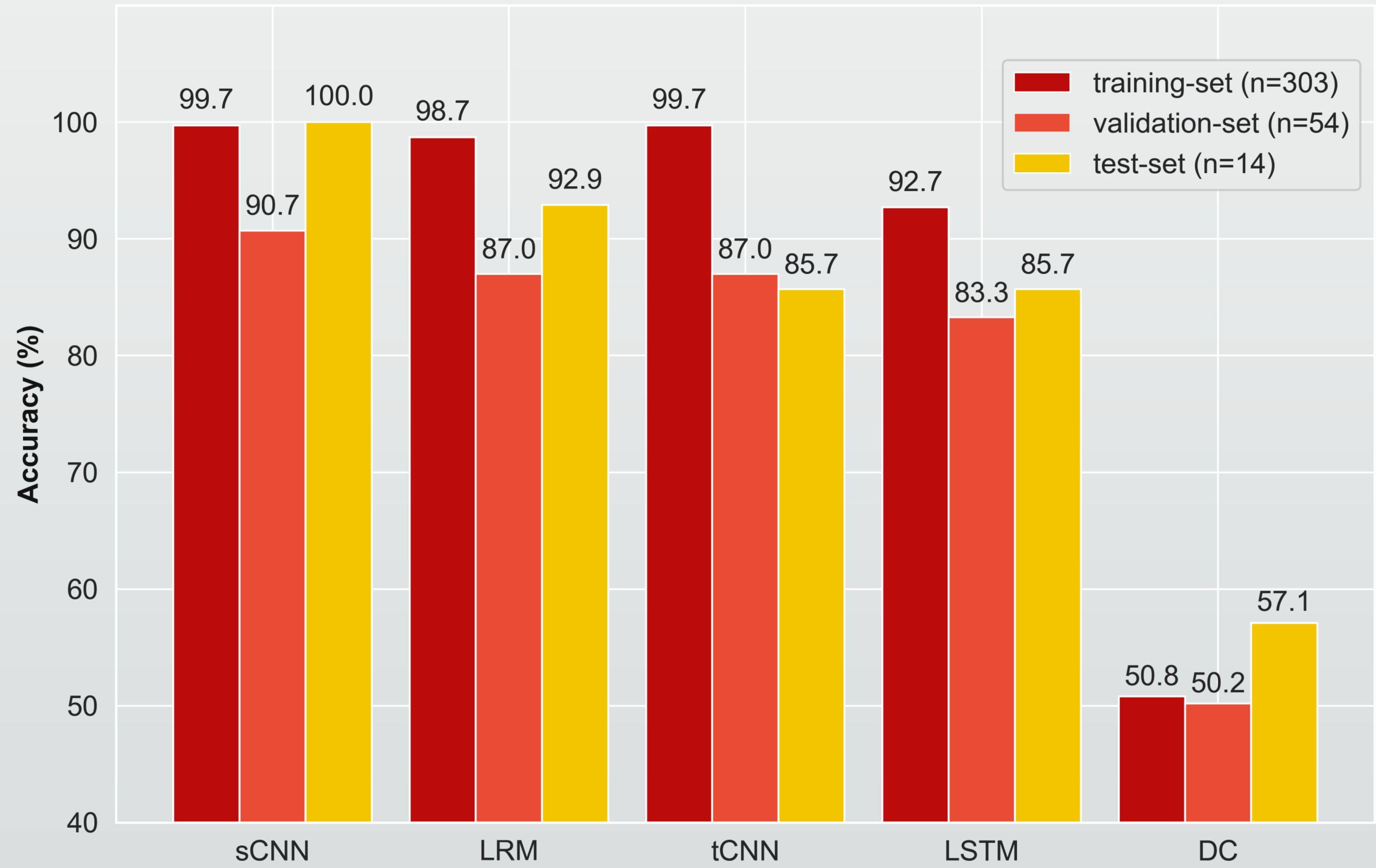

Figure 3: Accuracies of the models

Table 3: Accuracies and F1- scores of the models

Model	Training-set (n=303)	Validation-set (n=54)		Test-set (n=14)	
	Accuracy (%)	Accuracy (%)	F1-Score	Accuracy (%)	F1-Score
Logistic regression (LRM)	98.7	87.0	0.87	92.9	0.92
Single convolution (sCNN)	99.7	90.7	0.91	100	1.0
Three convolution (tCNN)	99.7	87.0	0.87	85.7	0.86
Long short-term memory (LSTM)	92.7	83.3	0.85	85.7	0.87
Dummy Classifier (DC)	50.8	50.2	NaN	57.1	NaN

Discussion

A classic LRM or modern ANNs based on word embeddings, convolutions and long short-term memory layers yield high accuracies in the prediction of Swiss Federal votes. This is notable considering the heterogeneity of the subjects of the votes over the last 40 years. It's remarkable that the best model (sCNN) is based solely on the 300 most frequent words extracted from the training set.

The analysis is based on high dimension, low sample size data. Linear models such as the LRM are quite easy to tune in this setting. To prevent ANNs from overfitting the dimensionality of the word vectors (vocabulary size, embedding vector) must be reduced or considerable dropout needs to be applied (7). The sCNN performs better than the deeper networks in this context.

The machine learning models have been compared to opinion polls among eligible voters enquiring how they would vote two weeks before the vote (6). All four models predict the outcome of the 14 latest votes more accurately (mean accuracy 91%) than the polls (79% accuracy, a comparison is shown at www.stellus.ch (2)).

Conclusions

The LRM and ANNs yield high accuracies in the prediction of Swiss Federal votes. The sCNN performs best followed by the LRM. ANNs need strong measures to prevent overfitting.

All models predict the outcome of the last 14 Swiss Federal votes more accurately than opinion polls among eligible voters two weeks before the vote.

Try it for yourself

The LRM is publicly available as a web service at www.stellus.ch (2).

References

1. Eigenössische Bundeskanzlei. [Online] https://www.bk.admin.ch/bk/de/home/dokumentation/abstimmungsbuechlein.html.

2. Müller, Daniel. stellus.ch. [Online] https://www.stellus.ch/home/hintergrund/#worum-es-geht.

3. Ebermann, Thomas. liip.ch. [Online] https://www.liip.ch/en/blog/sentiment-detection-with-keras-word-embeddings-and-lstm-deep-learning-networks.

4. Szymkowiak, Théo. Using Convolutional Neural Net for Sentiment Analysis. [Online] https://github.com/Theo-/sentiment-analysis-keras-conv/blob/master/train_keras.py.

5. Chollet, François. Keras. [Online] https://keras.io.

6. gfs.bern. [Online] https://www.gfsbern.ch/de-ch/.

7. Deep Neural Networks for High Dimension, Low Sample Size Data. Bo Liu, Ying Wei, Yu Zhang, Qiang Yang. [Hrsg.] Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17). Hong Kong University of Science and Technology, Hong Kong: s.n., 2017.