Exercise 2.1

2020-11-17

The goals for today

At the end of these exercises, you'll be able to:

1. Train, evaluate and interpret a logistic regression classifier
2. Fine-tune and evaluate BERT models
3. Use zero-shot learning

When finished, hand in the code + answers (preferably in a Jupyter Notebook) on Blackboard.

The manual for the exercises is available at:
https://jveerbeek.gitlab.io/dm-manual/

Today, we are going to try automatically reproducing the classification scheme used in the research by Wieringa et al. (2018) using different classification methods. We will focus on the first part of the classification scheme: the encoding strategy (affirmative, negotiated or oppositional).

1. Load the data `labeled_tweets.p`. The data are the tweets by U.S. representatives – labeled by us! The column `text_a` represents the tweet, `text_b` is the headline + description of the article, and `label` is the labeled encoding strategy (0 = affirmative, 1 = negotiated, 2 = oppositional).
2. Inspect the dataframe: how many examples for each label are in the data set?

1 Logistic Regression

1. Train a logistic regression classifier on BOW features and evaluate it. How does it perform, and how this performance compares to just random

selecting an integer between 0 and 2?

2. Try to interpret the model. Based on the words with the highest coefficients, what do you think the model has learned? (Note: because we have three labels now, using `lr.coef_` is a bit different than in the binary setting. The coefficients are now a list, where `lr.coef_[0]` are the weights for the first label (affirmative), `lr.coef_[1]` for the second (negotiated), and `lr.coef_[2]` for the third (oppositional)).
3. Use TF-IDF features instead of raw counts. Does it improve the performance of your model?

2 BERT: supervised

Use Google Colab for these exercises instead! You only have to install the transformers library there (!pip install transformers)

1. Fine-tune a DistilBERT model on our task and evaluate its performance: does it perform any better than using a logistic regression classifier?
2. Instead of using a DistilBERT which is trained of all kinds of texts, we could also make use of BERTweet, which is – the name says it all – specifically trained on tweets. To use BERTweet, we'll have to make a few adjustments to the code in the manual [also make sure you install the `emoji` package (!pip install emoji)].

   First of all, we have to load a different tokenizer:

```
  from transformers import AutoTokenizer
  tokenizer = AutoTokenizer.from_pretrained('vinai/bertweet-
    base', normalization=True)

```

   And of course, we also have to load a different BERT model:

```
from transformers import
    TFAutoModelForSequenceClassification

model = TFAutoModelForSequenceClassification.from_pretrained
    ('vinai/bertweet-base', num_labels=len(labels))

```

   Because of some curious bug in the transformers library, before we **fit** our model (model.fit), we have to add the following line to our code as well.

```
model.roberta.return_dict = False

```

3. Fine-tune the BERTweet model on our dataset and evaluate its performance. Does it run any better than DistilBERT?
4. Look at the confusion matrix of BERTtweets predictions: what can you conclude about the classification of tweets with the label *oppositional*?
5. During our Q&A yesterday, Tim de Winkel said the interpreting *retweets* as a communicative act takes some contextual knowledge: you have to know how retweets are generally used. Can you relate this to the performance of the models?

3 Zero shot classification

1. Think of the topics that were relevant during the U.S. elections in the last three months. Use these topics as candidate labels.
2. Classify the first ten tweets in the `labeled_tweets.p` dataset, and evaluate the predicted labels manually. Would this model fruitful to use to label the topics of these tweets?

   BONUS CHALLENGE: classify the tweets using the classification scheme of Wieranga et al. (2018) with the zero shot pipeline and evaluate its performance. How well does it perform?