

REPORT OUT: ASPECT BASED SENTIMENT ANALYSIS USING BART

28/03/2022

Giuseppe Marino

Target Audience: Sentiment Analysis of influencer-led product review comments on YouTube



CONTENTS

- Review of Questions to Answer/ Hypotheses / Approach
- Discuss technical challenges
- Dataset preparation
- Initial Findings
- Deeper Analysis
- Future works
- Implementation
- Hypothesis Results

SECTION I: QUESTIONS TO ANSWER

Sentiment Analysis of YouTube comments on products reviews led by an influencer:

- Discover perceptions based on online reviews (Sentiment Analysis) is nowadays essential for appropriate managerial marketing strategies: the fine-grained task of Aspect Based Sentiment Analysis, classifying targets and sentiments at once, can inquire comments into the most loved products reviews led by our client, a YouTube influencer.
- Deliver a Sentiment Analysis model with low-computing power and requiring a small training dataset
- Deliver marketing strategies led by the results of our Sentiment Analysis.

SECTION 2 : INITIAL HYPOTHESIS

- We expect that Sentiment analysis reaches higher metrics scores via deep learning-based models than lexicon-based and machine learning-based models.
- We expect that Transfer Learning knowledge from pretrained model on large corpora avoids common deep learning issues such as high-computing power detrimental over the training phase and compensate for insufficient training examples.
- We expect that thanks to the above properties, namely low computational power and small data resources, our model can be installed without the need for a central server to upload the new taken data and carry out further elaboration, thereby enabling an easily scalable detection.
- We expect that our Sentiment Analysis will lead to powerful marketing strategies for our YouTube influencer client.

SECTION 3: APPROACH

- **Training model.** For Aspect Based Sentiment Analysis, we have chosen zero-shot classification with a [BART model pretrained](#) on the Multi-genre Natural Language Inference (M-NLI) dataset. It is based on the [BART-large-MNLI](#) repository by [Hugging Face](#).
- **Implementation.** The model input is typical of a classification task: Youtube comments on influencer-led product reviews, and label candidates, such as target products and sentiments. The model transfers its knowledge on the M-NLI dataset by converting the classification task to an entailment task over the tokenization phase and back to classification after training.

SECTION 3: PRE-TOKENIZATION

- **Extract examples** in the dataset: 1) add column name, 2) filter out input examples equal to None, 3) concatenate questions and answers if both are present for better inference, as answers can be short. 4) convert dataset to list and remove special characters with the use of regex
- **Extract true labels** in the dataset (when labels are provided): it is essential to extract labels and filter examples according to associated labels as to have a minimum number of examples per label. However, note that the dataset does not need to be balanced as far as we used appropriate metrics such as F1 score.
- **Annotate new true labels** in the dataset (when labels are not provided): you can use text annotation tools such as [doccano](#) to annotate new labels in the dataset.
- Provide a base of labels amongst the true labels selected above for the **entailment alias multi-label classification** task.

SECTION 3: ZERO-SHOT CLASSIFICATION

- BART-large M-NLI pipeline calls on examples and label candidates
- Pipeline transforms each example-label pair in a premise-hypothesis pair
- Pipeline tokenizes the premise-hypothesis pair
- Pipeline infers entailment of the premise-hypothesis pair, alias **zero-shot classification**, which is considered a technique of fine tuning
- Logits corresponding to the output of the entailment inference are directly used for classification of label candidates

TECHNICAL CHALLENGES

- Hereafter, we show the model potentials through two public datasets: the Yahoo answers and emotion datasets.
- For reducing the computation time, the Yahoo answer dataset is reduced to 400 examples corresponding to 4 different topics evenly distributed (balanced dataset): Society & Culture, Science & Mathematics, Health and Education & Reference. For evaluating multi-label classification, we have also annotated these examples with four new emotion labels: Joy, Love, Sadness and Anger.
- For evaluating single-label classification on topics and emotion separately, we have reduced the emotion and topics datasets to 400 examples and selected the same labels as above.

Topics Dataset: questions/answers examples to be passed as premises in the entailment model

index	Yahoo questions/answers
0	<p>My friend feels pain in a particular spot in his upper spinal chord when he presses that spot.so wat is it?You can't press on the spinal cord.....it's inside the spinal column and totally protected by the bones in your spine.</p>
1	<p>Men, have you ever been the subject of sexual harrassment in the workplace?Yes, a lot of times actually...\n\nWomen do not view it the same way when the sides are flipped. They figure that all men "welcome" their advances. Normally, many do. But for some of us, we are in love with someone and it's just kind of annoying. It seems really desperate too...\n\nGood luck!</p>
2	<p>Is there such a thing as love at 1st sight? Any testimonies? Yes. I fell in love at first sight... 5 years after I met my DH...\n\nAs for "how do I know for sure it will last?" Well, love is a commitment. It lasts if you DECIDE it will last. Infatuation will come and go, but love last if you are committed to loving and being with someone the rest of you life.</p>

Topics Dataset: topics/emotions labels to be passed as hypothesis in the entailment model

The BART-large M-NLI model will infer the entailment between premises and hypothesis. The former are the Yahoo questions/answers examples and the second are either topics or emotions.

For evaluating multi-label classification, we have a set of two true labels, namely one topics and one emotion.

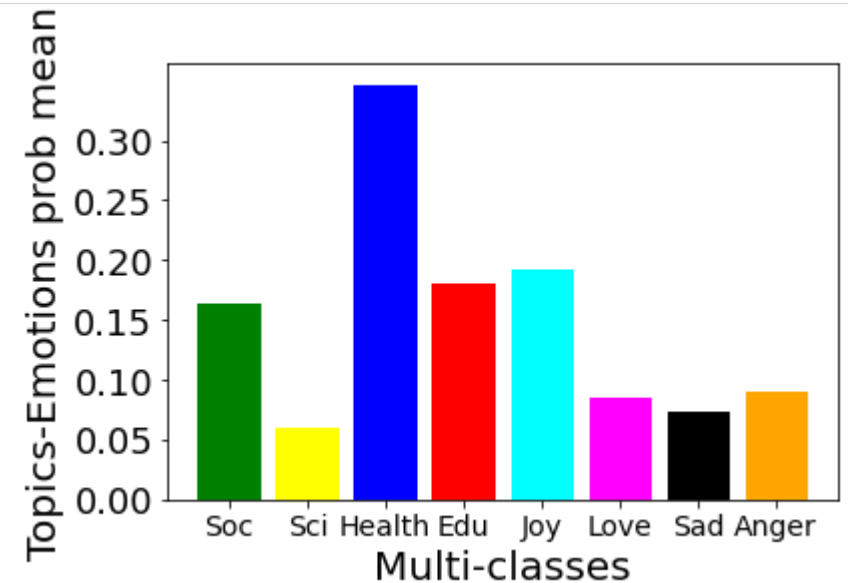
On the right, we have reported the set of true labels corresponding to the examples in the previous slide, where topics were labelled from the [literature](#), while emotions have been labelled by us using [doccano](#).

index	Topics labels	Emotions labels
0	Health	Sadness
1	Society Culture	Anger
2	Society Culture	Love

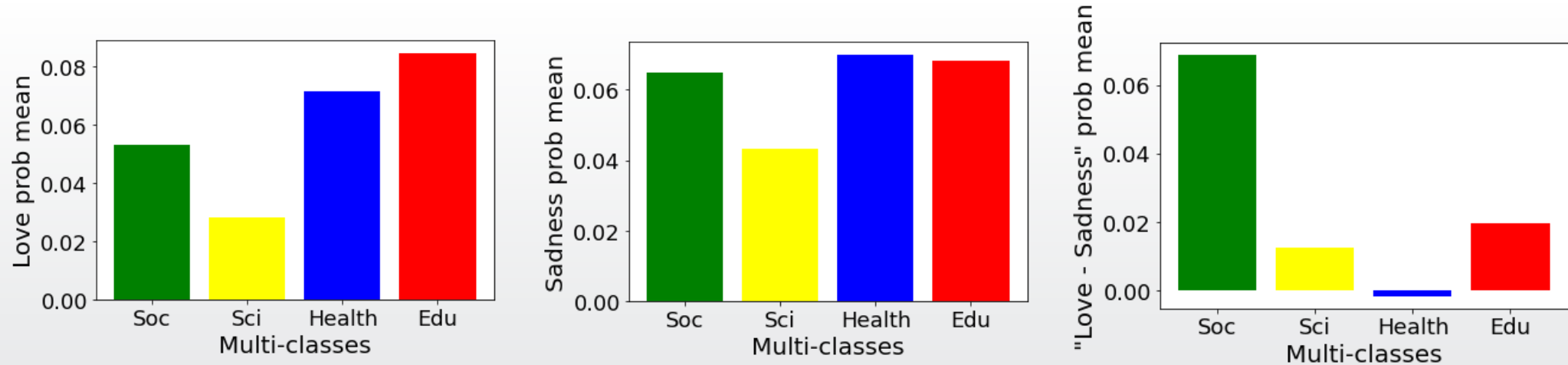
INITIAL FINDINGS : Multi-label Classification

- The Zero-shot Multi-label classifier on topics and emotions candidates is evaluated with a F1 score of 37 (as good as state-of-the-art*)
- The histogram on the right reports the probability mean of all examples per label.
- Interestingly, the parallel classification of emotions and topics provides us with an Aspect Based Sentiment Analysis tool as we can infer targets and associated sentiments simultaneously.
- We can note that the model well predicts the topics Health, above all others, and that positive emotions are more frequent than negative ones.

*Seo et al. IEEE Express 2020, Yin et al. arXiv:1909.00161. 2019



DEEPER ANALYSIS : Multi-label Classification



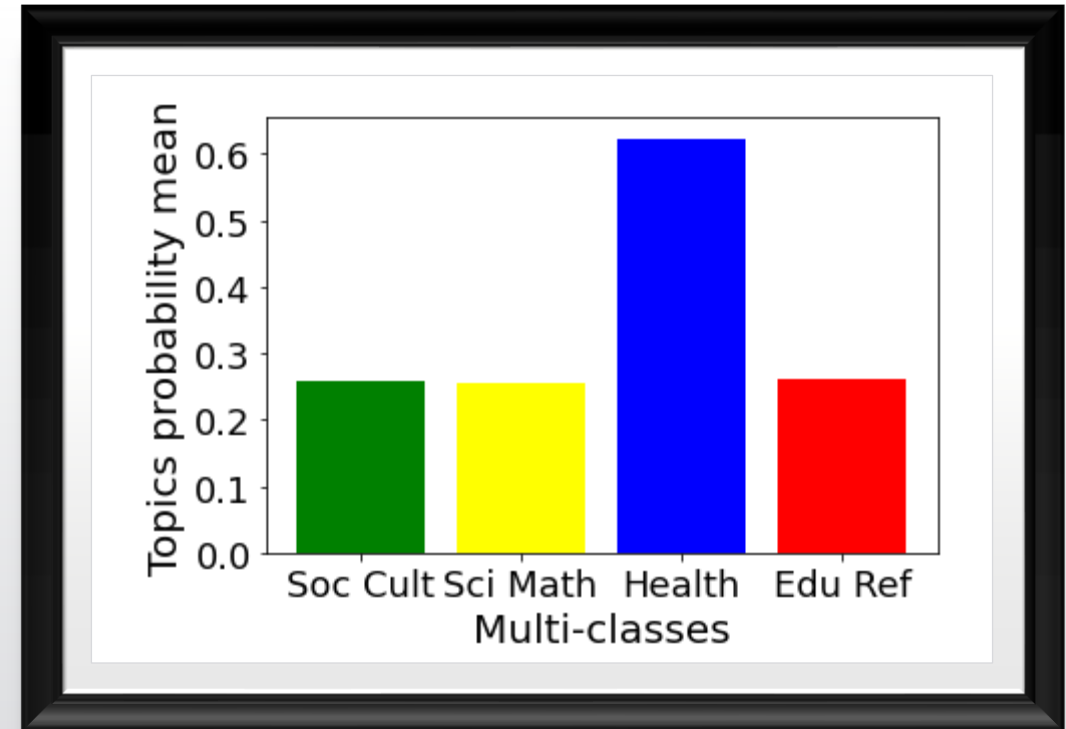
- The histograms above show a zoom of the Zero-shot classifier on two different emotion probabilities, namely love and sadness, for those examples that are positive on the 4 topics candidates.

- In particular, the third histogram reports the subtraction between love and sadness probabilities: we can see that Society topics are more positive than Health ones. This is because the former talk more about relationships, while the latter about diseases.

STATE-OF-THE-ART: Single-label Classification on Topics

- In order to benchmark our model to the state-of-the-art, we need to evaluate single-label classification on only topics. The Zero-shot classifier on the topics candidates predicts a F1 score of 40 (above state-of-the-art on topics F1=37.9 *).
- The probability mean for the true positive cases is reported in the histogram on the right. Each bar represents the true positive examples per candidate. This gives information about precision and recall and so it is directly comparable to the F1 score.
- We can see how the model well predicts the topics Health, above all others.

*Seo et al. IEEE Express 2020, Yin et al. arXiv:1909.00161. 2019



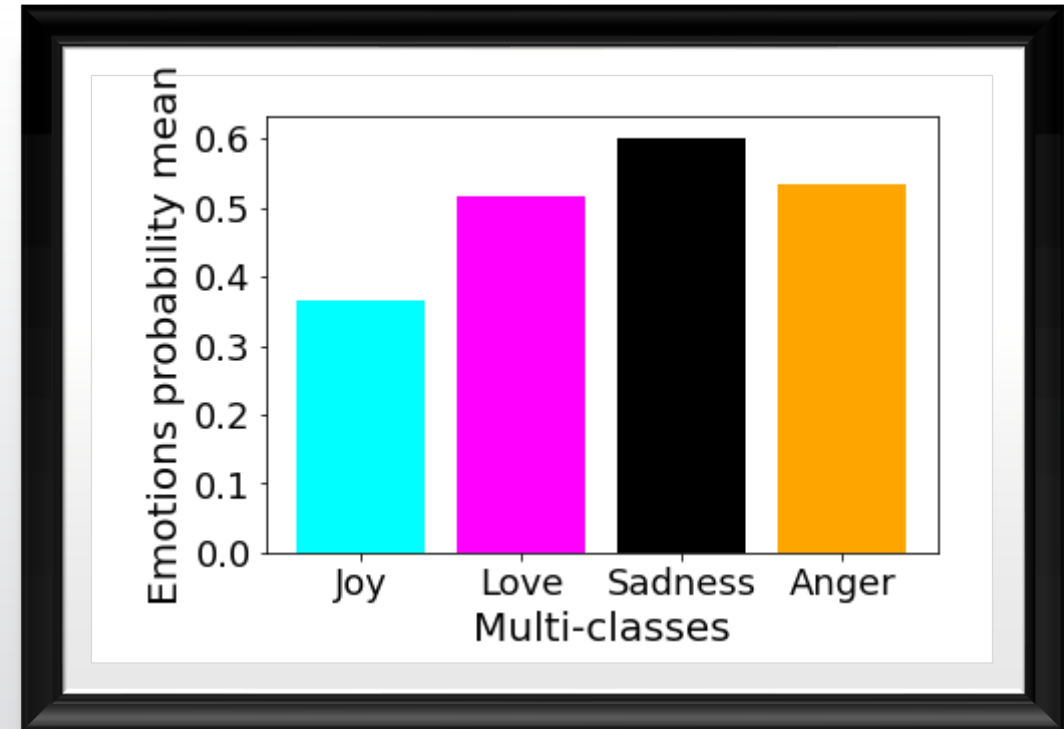
EMOTIONS DATASET : comment-label pair to be used as premise-hypothesis pair

class label	Comment
joy	RT @daveweigel: Hypocrisy and politics go hand...
joy	RT @DavidCornDC:This may be outrageous. It is...
joy	RT @DavidCornDC: If only we could catch Trump ...

STATE-OF-THE-ART: Single-label Classification on Emotions

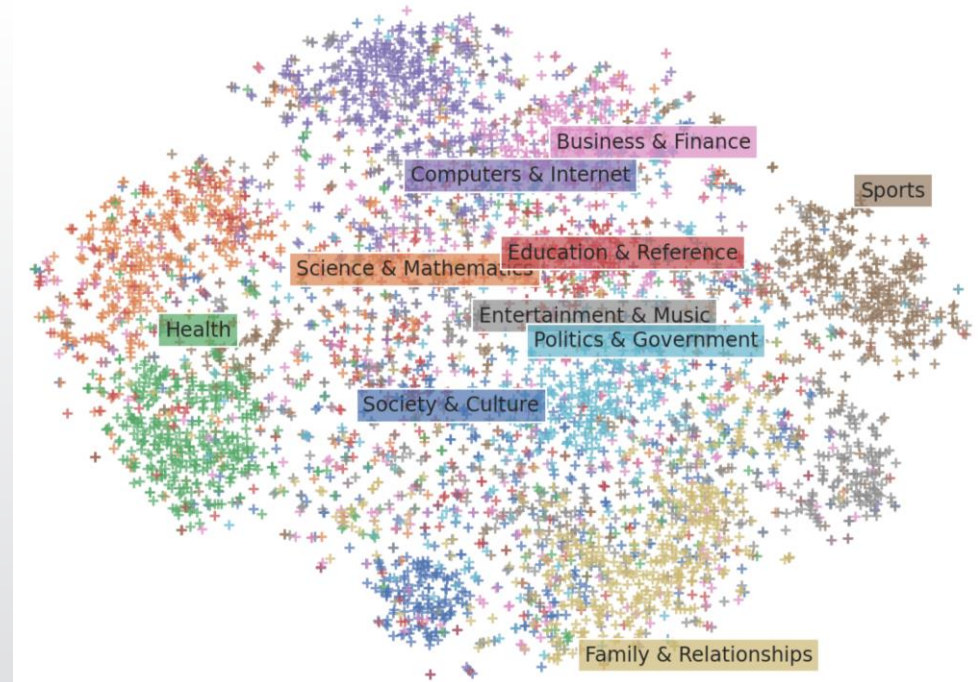
- The Zero-shot classifier on the emotion candidates predicts a F1 score of 45 (above state-of-the-art on emotions F1=22.3 *).
- The probability mean for the true positive cases is reported in the histogram on the right. Each bar represents the true positive examples per candidate. This gives information about precision and recall and so it is directly comparable to the F1 score.
- We can see how the model predicts best negative sentiments than positive ones.

*Seo et al. IEEE Express 2020, Yin et al. arXiv:1909.00161. 2019



FEATURE WORKS :VISUALIZE EMBEDDINGS

- For further assessing the model performance, we can visualize examples embeddings trained by the BART-large M-NLI model and reduced into 2D vectors via the t-SNE algorithm.
- As shown in the blog of [Joe Davison](#) (see graph on the right on Yahoo answers dataset), visualizing embeddings is informative of the distance between true labels and example embeddings.



SBERT to Wordvec projection embeddings of Yahoo dataset

IMPLEMENTATION – YouTube product review comments

- The Zero-shot classification, discussed in this presentation, has been implemented to serve with the local dataset of our YouTube influencer client and is capable to inquire current and future comments into successful product reviews.
- As the model only needs the inference phase, installation has been decentralized and carried out by using the client low-computational power resources.
- The products classified as positive had common interests and values with the videos created by the influencer, thereby confirming that the community trust resided on his content.
- Positive products have also been found associated to brand interactions: likes, comments and reposts, as this approach allowed the products to shine further.
- Our client influencer used this Sentiment Analysis to strengthen partnerships with brands and ask for higher campaign rewards.

HYPOTHESIS RESULTS :



- Thanks to the use of deep learning-based models, Aspect based Sentiment Analysis has been performed with higher metrics scores, F1 = 40 and 45 on the topics and emotion datasets, than lexicon-based and machine learning-based models*.



- Zero-shot classification with a BART pretrained model enables low computing power thanks to the direct classification inference and compensates for insufficient training examples thanks to the use of a model pretrained on large dataset such as BART-large and M-NLI, respectively.



- Our Sentiment Analysis of YouTube comments on product reviews led by an influencer, suggested that successful products should share content with the influencer videos and that brand interactions count.



- For faster time inference (about three times), we recommend a different architecture based on sentence-based BERT and Graph embeddings**.

*Seo et al. IEEE Express 2020, Yin et al. arXiv:1909.00161. 2019

**Chen, Q., et al. IEEE Int of Things Journal (2021).