

複数 BERT モデルで作るツールによって コメントの可読性向上を試みる

Using multiple BERT models for Improving the Readability of YouTube Comments

伊熊秀磨 ダビド・ラマモンジソア

Hidemaru Ikuma, David Ramamonjisoa

岩手県立大学ソフトウェア情報学部

faculty of Software and Information Science Iwate Prefectural University

Abstract: Nowadays, it is common to read the comments along with the video when watching a YouTube video. While reading the comments can provide new insights and empathy, some comments can be insulting or offensive. YouTube has its own way of displaying comments based on the top ratings or newest. In this study, we created a comment analysis tool that adds five analysis criteria to comments: harmful, spam, sentiment, sarcasm, and constructive. In this paper, we propose a tool and a framework to analyze and visualize comments on YouTube. Finally, we present some evaluation results.

1. はじめに

ユーザのコメントには、そのコンテンツに対する貴重な情報が含まれており、様々な利用方法が提案されている。また、それらの表示方法についても、ランキング化やフィルタリング、要約など様々な研究がされている[1]。本論では現在の YouTube のコメント表示方法に疑問を持ち、より良い改善方法があるのではないかと考えた。ユーザによる評価では、有害なコメントを判断することができず、多様なコメントを素早く理解することも難しい。そこで、BERT (Bidirectional Encoder Representations from Transformers) を用いた学習を行うことで、コメントのフィルタリングとソートを行い、より見やすいコメント欄を作ることを目的とした。

2. 手法

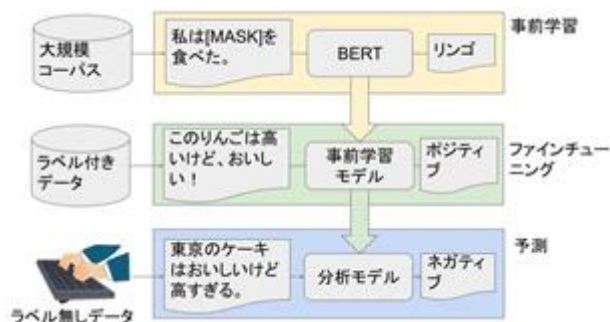


図 1 モデル作成の流れ

コメント分析には、Google が開発した BERT-Tiny 事前学習モデル[2]を使用した。有害[3]、スパム[4]、センチメント[5]、皮肉[6]、建設的[7]の 5 つのデータセットを用いて、図 1 のようにそれぞれファインチューニングを行ったモデルを分類モデルとしている。

ツールの仕組みは図 2 のようになっている。最初に、ツールで動画 URL の入力を行うと YouTube DataAPI からコメント情報を取得し、データベースに保存する。次に、分析モデルはデータベースのコメントを分析し、結果をデータベースに返す。最後に、ツールはデータベースの情報を読み込み表示する。ユーザは 5 つの分析基準を用いてフィルタとソートをしたコメントを読むことができる。



図 2 ツールの仕組み

3. データについて

有害[3]、スパム[4]、センチメント[5]、皮肉[6]、建

設的[7]の5つのデータセットは表1のように収集されたラベル付きコメントである。

表1 コメントデータセット

	Web サイト	件数
有害	Wikipedia	223,549
スパム	YouTube	1,956
センチメント	Twitter	1,600,000
皮肉	TheOnion、 HuffPost	28,619
建設的	SFU News	12,000

4. ファインチューニングの設定

コメント分析には BERT モデルを用いている。BERT モデルとは、Bidirectional Transformer encoder を Layer としたモデルである。事前学習には Masked Language model と、Next Sentence prediction が行われている。

BERT モデルは keras ライブラリを用いて次のように設定した。テキストデータの埋め込みに事前学習モデルの BERT-Tiny モデル[2]を使用した。隠れ層 128 ユニット、Layer2 層、Attention head2 のモデルである。SpatialDropout1D で 10%の単語を無視した後、それらを Dense にて sigmoid 関数で全結合している。loss 関数には binary_crossentropy、optimizer には adam、batch_size は 512 としている。

4.1 学習回数の模索

データセットを学習データ 9 割、テストデータ 1 割に分割し、最も良いスコアとなる学習回数を模索した。また模索の際、最後の結合時に、先頭 1 ユニットを用いる 1unit スタイル (1unit) と、全てのユニットを用いる sequence スタイル (SEQ) の 2 パターンを用意して模索した。有害、スパム、皮肉、建設的データセットは 100epoch まで行い、センチメントデータセットは 10epoch まで行った。

4.2 学習回数の決定

模索で得られた正答率のスコアを参考に、各モデルの学習回数と結合方法を表2のように決定した。

表2 学習回数と結合方法

	有害	スパム	センチメント	皮肉	建設的
学習	1 回	10 回	5 回	2 回	2 回
結合	1unit	SEQ	SEQ	SEQ	SEQ

それぞれのデータセットのモデルを評価した結果は表3のようになっている。適合率、再現率、F1 値は 2 クラスのマクロ平均を算出している。センチメントのスコアが他に比べて低い。これは、データセットが大規模であり、十分な学習を模索することが

できなかったことが原因である。

表3 各モデルのスコア

	有害	スパム	センチメント	皮肉	建設的
正答率	0.956	0.954	0.833	0.897	0.927
適合率	0.908	0.954	0.833	0.897	0.926
再現率	0.836	0.953	0.833	0.896	0.929
F1 値	0.868	0.953	0.833	0.896	0.927

5. 結果と考察

5.1 完成したツールの説明

5つのモデルを使用することで、図3のWebサイトが完成した。左から順番に投稿日時、投稿者、コメント、グットボタンの数、有害、センチメント、皮肉、スパム、建設的となっている。このうち、最後の5つはモデルによって予測した値を表示している。




🗓️	👤	🗨️	👍	👎	😊	🙄	🗑️	👉
2021-11-24 22:53:09		<p>[discussed] UPDATE: We managed to get Blender to run on the CMP card after posting this video. The results are not very good due to limited memory and PCIe bandwidth, here are render time results for the BMW blender demo. Lower is better.</p>	3152	0%	15%	99%	100%	100%
2021-11-25 16:42:34		<p>Reply [discussed] personally don't believe you mining is the future and can be damaging to the climate with its power consumption. Cpu mining is more eco friendly and some really good projects popping up. Been watching venus coin as it appears to be eco friendly and one of the most advanced crypto out there and fixing a lot of problems around crypto. Lead</p>	1	0%	38%	98%	100%	100%
2021-11-25 15:11:27		<p>Reply [discussed] @G88th No sorry these and the A1000 are mainly meant to be used in a VM. To be able to use the entire cores on one VM you have to buy a special license from Nvidia for V GPU. If not you can only use pass thru on hyper V and lower end VMware licenses you can only pass thru 4 core clusters. If you upgrade to full VMware Enterprise you can pass</p>	2	0%	15%	2%	100%	100%
2021-12-14 03:27:30		<p>top [hey if you smart guys want to explain how a graphics card can "mine" a "currency?" Because it's all a scam in my opinion. There's no physical way to obtain a Bitcoin. You just believe in this weird fake currency. I can't bring myself to invest in fake money. Lol I'll buy</p>	0	0%	22%	10%	100%	100%
2021-12-13 14:19:40		<p>top [So basically, its not a profitable miner. Until you resell it. So whos buying these used? I would love to know just how much this brings in. Combine that with the difficulty increase I wonder how long it will be profitable for. (once you reach the break even point. A great test would be to run this till its no longer making anything. sell it and see where you end up.</p>	0	0%	27%	0%	100%	100%
2021-12-13 14:19:40		<p>top [SO DISGUSTING THAT THINGS LIKE THIS EXIST. STEAL FROM THE RICH!!! IF YOU</p>						

図3 分析ツールの表示画面

discussion のボタンを押すことで、図4のように対話の様子を表示することができる。

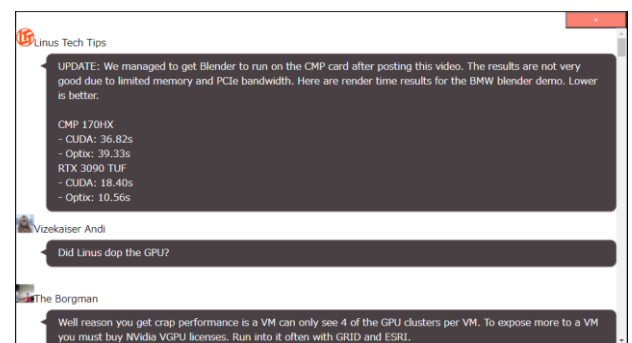


図4 discussion の表示

図5のように、ツールはソートや5つの予測値に対してフィルタを行うことができ、その分布やトップレベルのコメントとリプライのコメントの割合も表示する。フィルタは予測値の下限と上限を設定することで、その範囲内のコメントのみを表示する。分布は x 軸と y 軸に 5 つの分析基準を指定することで 2 次元座標にプロットされる。プロット点にマウスオーバーすることでコメントの一部が表示され、クリックすることで詳細の場所へジャンプする。

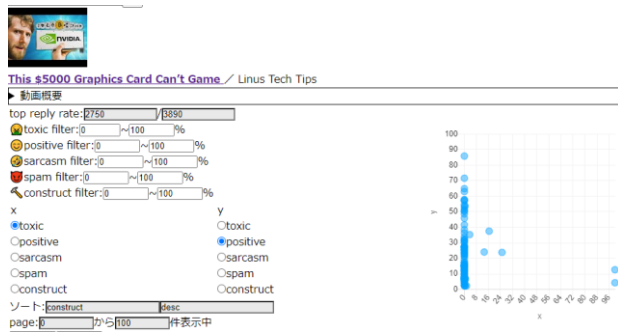


図 5 分析ツールの操作画面

5.2 ツールの使用感

ツールを本研究室の3年生に使用してもらった結果、以下のようなフィードバックを得た。

- コメント件数の多いコメントを取得した場合、コメントの表示が途中で止まってしまう。
これは YouTubeAPI に起因する問題であるため、API の取得上限解放を申請することで解決する。ただし、それでも取得が途中で止まってしまった場合、何らかの代替手段が必要である。
- グットボタンが多い順にできるか
できるが、この質問は操作性が分かりにくいという問題に起因している。より直感的な操作方法の検討や操作マニュアルの表示が必要である。
- ランダムにコメントを取得したい
ソートやフィルタの機能は実装していたが、ランダムの検討はしていなかった。
- 「F*CK」が入っていると必ず toxic が 99% になる

F*CK は肯定的であれ否定的であれ少し汚い表現であるため、ツールが良い働きをしている。

また、spam の予測に関してはデータセットが少ないということもあり、spam 以外のコメントも spam であると判定してしまう傾向にある。

6. おわりに

本稿では、複数 BERT モデルで作るツールによってコメントの可読性向上を試みた。YouTube のコメント表示方法と異なる手法を用いることで、新しい視点で新しい知識やコミュニティと出会うことができる。しかし、表 3 のスコアで示した通り予測モデルの精度が悪いため、フィルタとしての役割は果たせていない。より良いモデル構築とデータセットの収集が必要である。

参考文献

- [1] Aggregating and Analyzing Articles and Comments on a News website [http://p-www.iwate-pu.ac.jp/~david/](http://p-www.iwate-pu.ac.jp/~david/publications/DavidRama-28June2016-LNAI-Final.pdf)

[publications/DavidRama-28June2016-LNAI-Final.pdf](http://p-www.iwate-pu.ac.jp/~david/publications/DavidRama-28June2016-LNAI-Final.pdf)

- [2] google-research/bert: TensorFlow code and pre-trained models for BERT <https://github.com/google-research/bert>
- [3] Toxic Comment Classification Challenge | Kaggle <https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge>
- [4] YouTube Spam Collection Data Set | Kaggle <https://www.kaggle.com/lakshmi25npathi/images>
- [5] Sentiment140 dataset with 1.6 million tweets | Kaggle <https://www.kaggle.com/kazanova/sentiment140>
- [6] News Headlines Dataset For Sarcasm Detection | Kaggle <https://www.kaggle.com/rmisra/news-headlines-dataset-for-sarcasm-detection>
- [7] Constructive Comments Corpus (C3) | Kaggle <https://www.kaggle.com/mtaboada/c3-constructive-comments-corpus>