D-001

BERT を用いた分類器によるクラウドソーシングの質の向上
Improving the Quality of Crowdsourcing using BERT

太田 奈那[1)]　鈴木 優[1)]
Nana Oota　Yu Suzuki

1　はじめに

クラウドソーシングとは，一人で行うには困難な作業をインターネット上で募集した大量の人に代わりに行ってもらうことである．一人で行うには困難な作業でも，分割して依頼することによって，単純な作業にすることができる．また，不特定多数の人に依頼をするため，費用や時間といったコストはかかるが，一人ひとりにかかる負担を減らすことができる．さらに，インターネット上で作業を行うことができるため，作業者は誰でも好きなときに簡単に作業を行うことができる．

しかし，不特定多数の人によって行われるクラウドソーシングでは意見のばらつきが見られる．また，行われた作業結果が正しいものであるとは限らない．なぜなら，スパムワーカーと呼ばれる作業の評価をわざと正反対に回答したり，適当に数だけこなそうとしたりする品質の低い作業者が一定数存在するためである．そこで，よりたくさんの作業者に作業を行ってもらい，その多数決をとる．その結果は一般的な評価と言え，作業依頼者が求める結果に近いと考えることができる．そのため，多数決をとるためのたくさんの作業結果が必要になる．しかし，たくさんの作業結果を調達するには費用と時間がかかる．そこで，人の作業を代替できる分類器を BERT[1] を用いた機械学習によって作成できれば，費用も時間もかけずに仮想的に作業を代替できるのではないかと考えた．しかし，人の作業を分類器で代替できないと意味がない．そのため，分類器が人の作業を代替できることを確認するために本研究を行った．

本研究は，作業数が最も多かった作業者一人に着目して行った．一人の作業者の作業を代替できる分類器を作成することができれば，その作業者以外の分類器も作成することができるのではないかと考えた．そして，その分類器が作成できれば，より多くのデータを集めることができる．多くのデータが集まれば大多数の意見で多数決をとることができ，より一般的で作業依頼者が求めるものに近い結果が得られる．そして，この結果を得ることは，クラウドソーシングの質を向上させることになるのではないかと考えられる．

2　関連研究

西らの研究 [2] では，ソーシャルネットワークを用いることによって，作業者の品質を向上させる研究を行っている．ここでの高品質な作業者とは，作業依頼者の求める正解に近い作業結果を多く残している人のことである．

芦川らの研究 [3] では，作業者に作業を行う適性があるかどうかのフィルタリングを，作業を依頼する前と作業途中で行っている．そのフィルタリングを行うことによって，品質の低い作業者を取り除く研究を行っている．

どちらの研究も，作業者の品質について着目している．本研究は，BERT を用いた機械学習によって作業者の作業を模倣するような分類器を作成してデータ数の増量を試みている．そのため本研究とは，クラウドソーシングの質を向上させる点で同じであるが，品質向上のためのアプローチが異なる．

3　提案手法

本研究は，BERT を用いた分類器を作成して，作業者の評価予測を行い，クラウドソーシングの質を向上させることを目的とする．手順は以下の通りである．

（1）クラウドソーシングによるデータ作成
（2）BERT による分類器の作成
（3）分類器を用いた作業者の評価予測
（4）作業結果の集約

本研究は，既にクラウドソーシングによって構築されたデータを使用するため，（1）の手順を省略している．（2）から（4）については 3.1 節から 3.3 節で説明する．

3.1　BERT による分類器の作成

BERT による分類器を作成する．データは，クラウドソーシングにより構築されたものを使用する．データの内容については 4.1 節で説明する．このデータを使用して作業者の作業を模倣するような分類器の作成を行う．

この分類器の作成は二つの異なる手法で行っている．一つ目は，一人の作業者が行った作業結果だけを取り出して作成したデータセットを用いて分類器を作成する方法である．二つ目は，全作業者が行った作業結果を使用して作成した分類器を，一人の作業者が行った作業結果でファインチューニングして，その作業者に合わせた分類器を作成する方法である．このとき，全作業者が行った作業結果には一つのタスクにつき複数の評価ラベルがついているものがある．それらのタスクは，多数決をとって一つの評価ラベルを与える作業を行う．これらの手法で作成した分類器を用いる．

3.2　分類器を用いた作業者の評価予測

分類器を用いて作業者の評価予測を行う．使用する分類器は，作業者の作業を模倣するように作成したものである．この分類器を使用して，作業者が回答していないタスクの評価を予測する．

このとき使用する分類器が，作業者の評価と同一の評価を得ることができれば，作業者の代替を BERT を用いた機械学習によって実現することが可能であると考えた．そのため，使用するデータの中からいくつかのデータをテストデータとして無作為に取り出し，そのテストデータを使用して評価予測を行い，分類器の精度を確認する．その精度が十分なものであると確認できれば，作成した分類器を使用して作業者の評価予測を行う．

3.3　作業結果の集約

作成した分類器を使用して行った作業者の評価予測の結果をもとに，新たなデータを作成する．新たに作成したデータと作業者が実際に行った作業結果全てを用いて多数決をとり，一つのタスクに対して評価ラベルを与え

1）　岐阜大学 工学部 電気電子・情報工学科

直す作業を行う．

この新しく与えられた評価ラベルが，データを増やす前の評価ラベルと比べて，作業依頼者の求める評価に近い評価になっていれば，品質の良い結果を得ることができたと考える．

4 実験

本実験では，3.1 節で述べた一つ目の手法を用いて，作業者の作業を模倣するような分類器を作成することを目的とする．また，この分類器はどれくらいの精度があるのか，その精度が十分なものであるかを確かめる．

4.1 実験手順

データとして我々が構築したセンチメント分析データを利用する．データは 180,182 件ある．このデータは，作業者番号，評価ラベル，ツイートの内部番号，ツイート内容の四つのカラムで構成されている．また，このデータは，ツイート内に含まれる「笑」がポジティブまたはネガティブな意味を持つか，特にネガティブな意味を持つときどのような場面で使われているのかを知ることを目的に作成された．このデータの内，作業数が一番多かった作業者が行った作業結果だけを取り出す．取り出した作業結果は全部で 6,027 件であった．この作業結果の評価ラベルは全部で 7 種類存在する．ラベルの内容は，表 1 に示す通りである．この取り出した作業結果の中から「その他」という評価ラベルを除いた 5,884 件の作業結果を使用する．この作業結果を使用して，ツイート内容と評価ラベルの二つのカラムで構成されるデータセットを作成する．

このデータセットを，訓練データが 6 割，検証データが 2 割，テストデータが 2 割となるように分ける．訓練データの評価ラベルごとのデータ数は，ラベル 0 が 186 件，ラベル 1 が 2,303 件，ラベル 2 が 120 件，ラベル 3 が 449 件，ラベル 4 が 415 件，ラベル 5 が 57 件であった．ラベル 1 のデータ数が他のラベルのデータ数と比べて圧倒的に多いことがわかる．そこで，ラベルごとのデータ数を揃えるために，最も多いラベル 1 に数を合わせるように他のラベルのデータの複製を行う．その結果，データ数は 13,818 件となった．このデータセットを訓練データとして用いる．そして，このデータを BERT で学習させて 6 クラス分類を行い，作業者の作業を模倣するような分類器の作成を行う．また，テストを行い，精度を確認する．

4.2 結果・考察

図 1 に示すグラフは，4.1 節で作成したデータセットを使用して行った 6 クラス分類の学習曲線である．この分類器を使用して行ったテストの正解率は 55.14%であった．この精度が十分であるかどうかを確かめるために，6 クラス分類でのチャンスレートと比較する．6 クラス分類のチャンスレートはおよそ 16.67%である．そのため，本実験で作成した分類器は十分な精度があると考えられる．

表 1 ラベル内容とラベルごとのデータ数

ラベル番号	ラベル内容	データ数 (件)
0	ネガティブ	304
1	ニュートラル	3815
2	攻撃性あり	186
3	攻撃性なし	753
4	自虐	722
5	ポジ+ネガ	104
6	その他	143

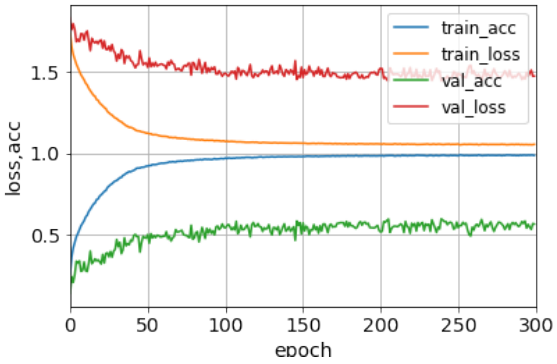

図 1 作業数が最多の作業者の分類器の学習曲線

図 1 を見るとわかるように，100 エポック付近を過ぎたあたりから検証データの正解率に大きな差が見られなくなってくる．このことから，今回はエポック数を 300 エポックと設定したが，200 エポックほどでも十分精度の高い分類器を作成することができると考えられる．

使用したデータを訓練するのにかかった時間は，1 エポックでおよそ 186 秒だった．そして，最終的に 300 エポックの学習を終えるのにはおよそ 16 時間かかった．また，使用したデータセットを作成するために行ったクラウドソーシングで，作業を行った全ての作業者数は 604 人であった．そのため，これら全ての作業者の作業を模倣するような分類器を作成するのにかかる時間はのべ 9736 時間であると予測できる．

5 おわりに

本研究は，クラウドソーシングによって構築されたデータを用いて作業者の作業を模倣するような分類器の作成を BERT を用いて行った．その結果，チャンスレートと比較して，十分な精度であることが確認できた．

今後の展望として，3.1 節で述べた二つ目の手法について実験を行い，4 章で行った実験と結果を比較したいと考えている．また，作業を模倣するような分類器を作業者の人数分作成し，3.2 節と 3.3 節で述べた手順を行う．この手法を行うことによって，作業依頼者が求める結果により近い結果を得られるかどうかを確かめることが必要であると考えている．

謝辞

本研究の一部は JSPS 科研費 19H04218 の助成を受けたものです．

参考文献

[1] Jacob Devlin, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proc.NAACL-HLT, Volume 1*, pp. 4171–4186, 2019.

[2] 西智樹, 小出智士, 大野宏司, 長屋隆之. ソーシャルネットワークを用いたクラウドソーシングの品質向上. 人工知能学会全国大会, 2013.

[3] 芦川将之, 川村隆浩, 大須賀昭彦. プライベートクラウドソーシングにおける精度向上. 人工知能学会全国大会, 2014.