

課題 3

Fine-tuning 以前および以後のスコアは以下の表のとおりであった(有効数字 17 桁).

Finetune	Pearson 相関係数	Spearman 相関係数
以前	0.7252414151720105	0.7066443476709469
以後	0.8946544003126126	0.8504891433302013

相関係数の結果より, BERT に対し fine-tuning を施すことで, STS の予測をより人間の精度/感覚に近づけることができるようになったと言える.

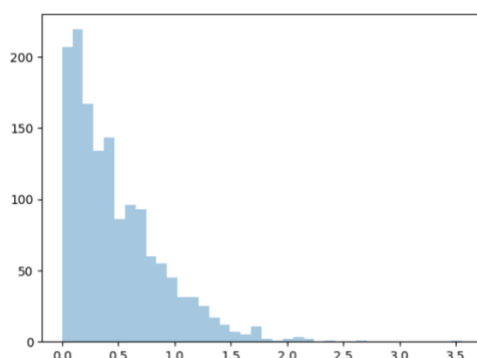


図 1. 予測結果と正解ラベルの値が乖離しているデータの個数の分布

また, テストデータについて, 予測結果と正解ラベルの値が乖離しているものについて図 1 および付録 1 に挙げる. なお, テストデータは JSTS データセットの `valid-v1.1.json` を使用している. 乖離の最大値は `3.5497367` だった. これを見ると, 予測結果と正解ラベルの値が乖離した文章対について以下の特徴のいずれかが見られることが分かる.

1. 2 文の意味が異なっているものについて少数の単語を置換し意味を大幅に変更している
2. 主語や目的語といった順番を入れ替えている

参考までに, 付録 2 に予測値とラベルが似ている文章対を見ると, 以下の特徴が確認できる.

1. 単語が意味的に全く異なるものに多数置換されている
2. ほぼ同じ単語を用いて同じ意味の文章になっている

上記からわかる結論として, BERT の STS スコアは各トークンのベクトルの平均の cosine 類似度を用いているため, 同じ意味を示すトークンが文対で離れたところにあったり, 1 単語で大きく意味が変わる部分の置換(目的語など)を行っている場合に対し人間が考える意味の乖離の感覚とは大きくずれた結果を出すことが確認できる. このため, 文章の順番の入れ替えがある文章や, 肯定文と否定文などの文対に対しては BERT の STS スコアの信頼性は落ちることが予想できる.

付録 1. 予測結果と正解ラベルの値が乖離したデータについて、
2.0 以上のデータを大きいものから降順に並べたもの。
文の横のブーリアンの値は、真理値は予測値がラベルより大きいかを示す。

{'sentence1': '建物の中で一頭の牛が人に引かれて立っています。', 'sentence2': '会場内で男性がホルスタインを連れています'} False
{'sentence1': '足元にキャメル色のカバンがあります。', 'sentence2': '絨毯の上に茶色い旅行トランクが置かれています。'} False
{'sentence1': 'きりんが草むらの中を歩いています。', 'sentence2': '草原でキリンが歩いているところです。'} False
{'sentence1': 'ベットのの上には大きなぬいぐるみが置いてあります。', 'sentence2': '窓のさんにぬいぐるみがかかっている。'} True
{'sentence1': '芝生の上に駐車しているバスの近くにきぐるみを着た人が立っています。', 'sentence2': '芝生の上に赤と緑色に塗り分けられたバスが止まっています。'} True
{'sentence1': '赤ちゃんが白いタオルの上に寝ています。', 'sentence2': '男の子がネクタイを締めて芝生の上に寝ています。'} True
{'sentence1': 'スマートフォンの上に猫の置物があります。', 'sentence2': '携帯電話が布の上に置かれています。'} True
{'sentence1': '赤い車の後ろに、犬が2匹います。', 'sentence2': '車の後ろで2匹のワンコが遊んでいます。'} False
{'sentence1': '台所はきれいに片づけられています。', 'sentence2': 'キッチンが生理整頓されており、きれいな状態です。'} False
{'sentence1': '小さな鏡が立てかけられ歯ブラシやらとにつくなどが乱雑に並んでいます。', 'sentence2': '洗面台の上に様々な物が並べられています。'} False

付録 2. 予測結果と正解ラベルの値の乖離が最も小さいものから 10 つ降順に並べたもの。

{'sentence1': 'ビールの入ったコップと携帯電話が手に持たれています。', 'sentence2': '人の手に飲み物の入ったグラスと携帯電話が持たれています。'} False
{'sentence1': '男の子が菓子袋を持って微笑んでいます。', 'sentence2': 'テーブルの上に花が活けられた沢山の花瓶が乗っています。'} True
{'sentence1': '男性がベンチに座って紙を広げています。', 'sentence2': '皿の上に牛肉と野菜が乗っています。'} False
{'sentence1': '道と道の間を流れる水路を小さな橋がつないでいます。', 'sentence2': '遠くに山が見える小さな川の橋の上に青い標識がたっています。'} False
{'sentence1': 'じゅうたんに雑貨が整理されて置いてあります。', 'sentence2': '部屋の地面に、自転車に乗るための道具が並んでいます。'} False
{'sentence1': '足元にキャメル色のカバンがあります。', 'sentence2': '青い車体のバスがこちらに向かっています。'} True
{'sentence1': '黒いトラックの前に、アメリカンバイクが二台止まっていて、バイクの横にはヘルメットを被った人物が立っています。', 'sentence2': '車やバイクが停まっている駐車場にヘルメットを被った人がいる。'} True
{'sentence1': 'サーフボードに乗って波乗りをしている人が一人います。', 'sentence2': '水の上でサーフィンをしている人がいます。'} True
{'sentence1': '白い箱の中にドーナツが入っています。', 'sentence2': '白い箱の中にたくさんのドーナツが入っていて、箱の横にはコーヒーが置いてあります。'} False
{'sentence1': '大型のバスが並んで止まっています。', 'sentence2': '男性たちが、フリスビーの試合をしています。'} False