

テキストから感情を読み取る AIの開発

東福岡高等学校 2年生 浦川遥斗

背景と目的

スマートフォンやパソコンの普及に伴い、テキストを用いたコミュニケーションが増えた。手紙と比べると、SNSなどは相手との意思疎通を手軽に行うことが可能である。一方で、手軽であるが故に、テキストに含まれる情報が少なく、意味の取り違えが生じやすい。

声のトーンや表情といった非言語情報が含まれず、言語情報も大きく制限されやすいテキストを用いたコミュニケーションでは、自分が相手に伝えたい感情と相手に伝わる感情に差が生まれやすい。

この感情ギャップによるトラブルは人々を悩ませている。これを解決する一歩目として、本研究では、文章から感情をリアルタイムで判定し、視覚的にフィードバックするAIシステムの開発を行う。



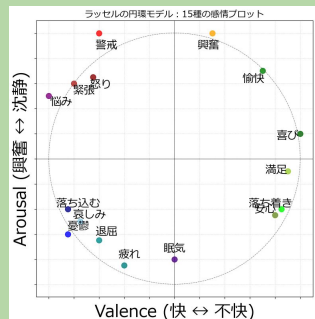
研究内容・方法

本研究では、ラッセルの円環モデル(図1)よりテキストデータから感情を快-不快と興奮-沈静の2軸で定量的に推定するAIモデルをpythonで構築する。

- ①日本語の単語や文章を集め、それぞれに快-不快と興奮-沈静の感情スコアを手動で定義した。このスコアをAIの教師データとして使用する。
今回は合計1000個以上の単語・文章を用意した。

- ②AIが文章を理解し学習するために、以下の手順でモデルを構築した。
1.用意した日本語の文章をJanomeというモジュールの形態素解析関数を使って単語に分解した。
2.TfidfVectorizerというモジュールを使用して、分解した各単語に対し文章の感情を判断する上での重要度(重み)を計算した。これによって各単語がどのような価値(感情)を持つかが設定される。

- ③最後にモデルで推定した感情の結果を視覚的に表示する。
1.作成したモデルを用いて、入力された新しい文章の快-不快と興奮-沈静の値を予測する。
2.快-不快と興奮-沈静の値を数学的にグラフ上の角度(感情の種類)と原点からの距離(感情の強さ)に変換する。
3.変換して得た感情の種類と強さを表示する。



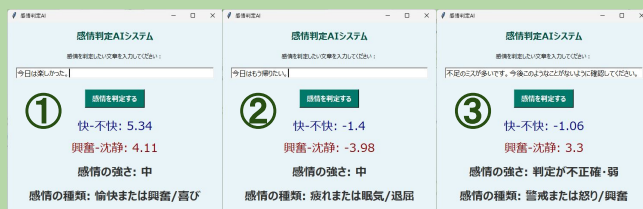
(図1)ラッセルの円環モデル(Russell, 1980)に基づき作成)

結果

作成したAIをテストするために感情を想定した文章を用意した。今回は変換した角度に近い3つの感情を表示し、これらの感情と想定した感情が一致したかどうかを調べる。例⑤に関しては皮肉を含む文章である。

- 例①今日は楽しかった。(喜び)
例②今日はもう帰りたい。(疲れ)
例③あなたは注意不足のミスが多いです。今後このようなことがないように確認してください。(怒り)
例④長期間にわたり、睡眠時間も削って必死に取り組んできた研究プロジェクトが、昨日ついに、学会会議で最優秀賞を獲得しました。これまでの苦労が一瞬で報われたような、魂が震えるほどの感動を覚えています。(喜び)
例⑤その努力、報われるといいですね。いつか... ね？(警戒)

結果として、用意した56個のテスト用文章のうち46個が一致した。しかし皮肉を含む文章は不一致の印象があった。



想定感情と一致

想定感情と一致

想定感情と一致



想定感情と一致

想定感情と不一致

テストデータへの適合率 : 82.1%

考察

本研究では8割程度の精度で感情の判定が行えるAIが作成できた。しかし、今回用意できた教師データ数は1000強であり、入力する文章によっては未学習の単語が多く、正確な感情判定が行えないことがある。また、今回採用したAIの作成方法では入力された文章が皮肉的表現である場合、うまく判定が行えないこともある。

そのため、AIの作成方法を変えたり教師データをもっと用意したりすることで、より高精度で感情を判定できるAIを作成できると考えられる。

今後は教師データ数を増やしてより高精度なAIを作成したり、SNSと連携させたりして、より手軽に感情の判定をできるようにしていきたい。

参考文献

Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178.

補足

○使用したモジュール

os,pickle,math,tkinter,janome,sklearn,matplotlib,japanize-matplotlib,numpy