

集合知型マーケティングサポートシステムにおける 顧客コメント解析

新井啓太^{†1} 櫻井義尚^{†1} 櫻井恵里子^{†2} 鶴田節夫^{†3}
 明治大学^{†1} 西武文理大学^{†2} 東京電機大学^{†3}

1. はじめに

サービス業においては、自社サービスの評価・改善を目的として、顧客にアンケート調査やインタビュー調査を行うことがある。リッカート尺度のような選択式のアンケートなどは定量的な集計が容易であるが、自由形式のコメントやインタビューなどは人が定性的に分析する必要があるため、人的コストが高くなってしまふ。しかし、深く顧客を理解するためには、これらの定性的な分析を欠かすことはできない。そこで、顧客のコメント分析を効率化するため、本研究では、自然言語処理を用いたコメントの評価指標を提案する。

大規模なサービス施設においては、一般的に顧客コメントは膨大な数になり、その中から重要なコメントを見つけるのは大変な労力を要する。例えば、複数あるテキストから着目するテキストを一つに絞り込むとき、一つずつ人間がコメントを読んで内容を確認する必要がある。このような作業をする上で定量的な指標が存在すると、重要なコメントの絞り込み作業を効率化することができる一般的な分類軸の例として、①多数派・少数派②ポジティブ・ネガティブ③意外性④流行りであるかどうか、などが挙げられる。本研究では①の多数派・少数派に着目した。なぜなら、③④の意外性や流行りを発見するきっかけになるからであり、応用性があると考えた。

本研究で扱うデータは、櫻井[1]によって収集された東京ディズニーランドの施設・サービス評価の結果を用いる。櫻井はキャプション評価法[2]を用いた意見収集実験を行なった。この実験では調査用スマートフォンアプリを用いて、調査員が実際に来園し気になった事をスマートフォンで撮影しその内容を自由記述で記録する。この実験でキャプション評価法を用いたテーマパーク顧客分析の有用性が確認された。

本論文では、この実験[1]で収集された 275 件

の施設やサービスの意見テキストデータを用いて、各々のコメントが多数的か少数的かを機械的に算出する手法を開発し、提案された手法と人間が行った結果が類似するかを確認した。

2. 関連研究

本研究では、2つの自然言語処理の既存手法を用いた。1つ目は単語埋め込みモデル(Word Embeddings Model)であり、Mikolovらによって提案された単語を低次元の実数ベクトルで表現する手法である。[3] 2つ目は文章間計算(Word Mover's Distance)であり、Kusnerらによって提案された文章間の類似度を1つ目の単語埋め込みモデルを用いて計算する手法である。[4]

3. 提案手法

本研究の提案手法は大きく2つのパートに分かれる。距離計算処理パートは、全てのコメントデータと各コメントが誰によって書かれたかが判別できるユーザーIDを入力とし、全てのコメントについてテキスト間の距離を返す。指標計算処理パートは、距離計算パートの結果を受け取り、任意に選択されたコメント(x)について多数的か少数的かの指標が計算されその結果が返ってくる。処理手順を下記に示す。

① 距離計算処理

1. ユーザーIDが紐づいたコメントを n 件入力する
2. 各コメント内のテキストを単語毎に分割する(分かち書き)
3. 単語毎に学習済み単語埋め込みモデル[3]から単語のベクトルを取得する
4. 全てのコメントのペア($n*(n-1)/2$ 件)について文章間類似度の距離[4]を計算する

② 指標計算処理

1. 全てのコメントのペアを距離の小さい順にソートする
2. 任意に設定した上位 $k\%$ のコメントのペアに"類似コメントフラグ"を付与する
3. 任意に選択されたコメント x を保持するペアを抽出する($n-1$ 件)
4. ②-2 で付与された"類似コメントフラグ"をユーザーID毎にカウントする
5. ②-4 で 1 以上カウントされたユーザー数を全ユーザー数で除算し、割合を算出する

Customer Comment Analysis
 in Collective Intelligent Marketing Support System
^{†1} KEITA ARAI, YOSHITAKA SAKURAI, Meiji University
^{†2} ERIKO SAKURAI, Bunri University of Hospitality
^{†3} SETSUO TSURUTA, Tokyo Denki University

コメントID	手動値	3%	6%	9%	12%	15%	18%	21%	24%	27%	30%
1	44.44%	0.00%	16.67%	33.33%	50.00%	55.56%	55.56%	77.78%	77.78%	83.33%	83.33%
2	33.33%	33.33%	38.89%	50.00%	50.00%	55.56%	55.56%	55.56%	55.56%	55.56%	55.56%
3	16.67%	61.11%	66.67%	72.22%	77.78%	83.33%	83.33%	83.33%	88.89%	88.89%	94.44%
4	11.11%	0.00%	11.11%	11.11%	11.11%	16.67%	16.67%	16.67%	22.22%	38.89%	38.89%
5	11.11%	50.00%	55.56%	55.56%	66.67%	66.67%	72.22%	77.78%	77.78%	77.78%	77.78%

表 1. k が 3 から 30%の時の実験結果

コメントID	手動値	0.3%	0.6%	0.9%	1.2%	1.5%	1.8%	2.1%	2.4%	2.7%	3.0%
3	16.67%	16.67%	16.67%	16.67%	33.33%	33.33%	33.33%	33.33%	38.89%	50.00%	61.11%
5	11.11%	11.11%	11.11%	22.22%	22.22%	27.78%	38.89%	44.44%	44.44%	44.44%	50.00%

表 2. k が 0.3 から 3.0%の時の実験結果

4. 実験

本実験で扱うデータは、[1]で収集された 275 個の日本語で書かれコメントを用いた。その内 5 個を無作為に選び、これらコメントに対し提案手法によって多数的か少数的かの指標を算出した。また、比較評価用に人間が提案手法の指標計算処理パートと同じような手順で算出した多数的か少数的かの指標を、同じく無作為抽出したテキストに対して用意した。

次に、実験条件として田口ら[5]が提案した朝日新聞社の 800 万記事を学習した学習済みモデルを用いた。本実験では GloVe を Retrofitting によって学習されたモデルを用いた。

また、本実験では提案手法によって算出された指標と人間が行なった指標を比較して妥当かどうかを検討する。そこで、提案手法②-2 について k を固定せずに 3~30%の 3%刻みと、0.3~3.0%の 0.3%刻みで行なった場合の結果を示した。

5. 結果

表 1,2 では、無作為抽出した 5 つのテキストに対して手動で指標算出した結果と提案手法による指標算出を行なった結果が比較されている。また、表 2 ではコメント ID が 1,2,4 の結果を割愛した。さらに、各コメントに対して手動結果と比較して前後 10%以内の提案手法結果は青色で表示されている。

今回表示した k の範囲では、5 つのコメントに対して提案手法による指標と人間による指標で近い数値を出す k が確認できた。しかしながら、すべてのコメントに対して同一範囲の k で確認できなかった。特に、コメント ID が 3,5 のものに関しては、 k が 3~30%の範囲では近い数値を確認できず、0.3~0.6%の範囲で確認できた。

6. おわりに

本研究では、複数コメントを多数的か少数的かに分類する作業を支援する事に着眼し、その分類軸における指標を設計した。結果として、提案手法により人間と近い数値を算出できる事が確認できた。また、任意に設定する k については、コメントによってブレがある事が観測されたことから、この事象の解明を今後の課題とし、どのようなコメントにおいても人間による数値と一致する共通な k の範囲に収束する手段を検討したい。

参考文献

- [1] 櫻井恵里子(2018),「テーマパーク調査におけるキャプション評価法を用いた顧客分析事例」『日本マーケティング学会カンファレンス・プロシーディングス』, Vol.7, pp.188-200.
- [2] 古賀誉章・高明彦・宗方淳・小島隆矢・平手小太郎・安岡正人(1999)「キャプション評価法による市民参加型景観調査—都市景観の認知と評価の構造に関する研究 その1」『日本建築学会計画系論文集』(517), pp.79-84.
- [3] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," NIPS, 2013.
- [4] M. J. Kusner, Y. Sun, N. I. Kolkin, K. Q. Weinberger, "From Word Embeddings To Document Distances," Proceedings of the 32nd International Conference on Machine Learning (ICML 2015), 2015.
- [5] 田口雄哉, 田森秀明, 人見雄太, 西島羽二郎, 菊田洗. 同義語を考慮した日本語単語分散表現の学習. 情報処理学会第 233 回自然言語処理研究会, Vol.2017-NL-233, No.17, pp.1-5. October 2017