

LSTM を用いた感情分析による授業アンケート自由回答文の解析

Analysis of Free Description in Lecture Questionnaires by Sentimental Analysis using LSTM

大谷 崇文^{*1}, 椎名 広光^{*2}
Takafumi OTANI^{*1}, Hiromitsu SHIINA^{*2}

^{*1}岡山理科大学大学院総合情報研究科
^{*2}岡山理科大学

^{*2} Okayama University of Science
Email: i18im02ot@ous.jp

あらまし：コメントデータの人手による評価には多大なコストが必要である。大学の授業アンケートにおいても同様であり、適した評価指標が必要である。そこで本研究では、自由回答文の一部に対して人手による6段階のランク評価を行い、LSTMを用いた感情分析による学習と未評価のランクの確率の評価を行った。

キーワード：自由記述解析，感情分析，アンケート分析，講義評価，LSTM

1. はじめに

講義アンケートの形式には選択式と自由記述式がある。特に、自由記述式の場合は選択式のように設問が限定されていないため、アンケート回答者の自由な意見調査が可能である。講義アンケートを実施することで教育の質の向上を図る大学にとって、教員や講義に対する学生の率直な意見は有用な情報であると考えられる。しかし、教育分野に特化した場合でのテキストを分析する手法が確立されていないことから講義アンケートの自由回答文の分析を行うには多大な人的コストや時間的コストが必要になる。また、自然言語処理分野では、文章分類や文章評価の研究は、商品レビューやアンケートの自由回答項目等の解析が行われている。

これまで筆者らは講義アンケートの自由回答文（以下、コメント）の評価からコメントを構成する単語を評価し、反対に単語から構成されるコメントの評価する互いに評価する方法を開発してきた⁽¹⁾。一方、ニューラルネットワークの機械学習を利用した評判分析が新しく行われてきており、LSTM (Long Short Term Memory, 以下 LSTM)⁽²⁾が利用されている。そこで本研究では、ニューラルネットの多層化とLSTMを用いた機械学習による授業アンケートの評判分析を行った結果について述べる。また、既存手法との比較を行う。

2. 使用したデータ

本研究では、岡山理科大学総合情報学部情報科学科で2014年春学期に実施した講義アンケートの1678件を利用した。1678件のうち12人の学生がそれぞれ100件のコメントを6段階で評価したものを、学習データのシードとして結果を利用している。

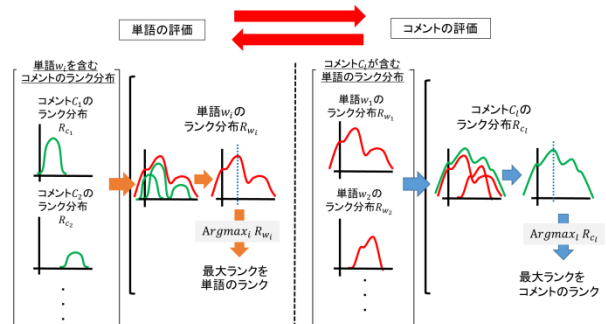


図1 コメントと単語の再帰的評価法

3. コメントと単語の再帰的評価法による解析

コメントと単語の再帰的評価法を、以下について以下に述べる（図1）。

- (1) 評価コメントを利用した単語評価推定：学習用のシードデータの評価コメントからコメントに含まれる単語の評価推定を行う。手順としては、評価コメント中の各単語の評価から単語ランク分布を作成し、単語ランク分布から単語の評価を推定する。
- (2) 単語ランク分布から未評価コメントの評価推定：評価コメントから作られた単語評価を直接使用し、それを利用して得られるコメント評価を再び単語評価に利用する。
- (3) 全コメントの評価推定：人手で評価したコメントとそれから推定された単語の評価、また未評価のコメントの評価を利用して、全コメントの評価を推定する。推定手法としては、次の(3-1)と(3-2)を全コメント評価の改善がなくなるまで繰り返す。
 - (3-1) 未評価コメント内に含まれる単語の評価分布から最大ランクをコメントの評価とする。
 - (3-2) 単語を含んでいるコメントの評価分布から最大ランクを単語の評価とする。

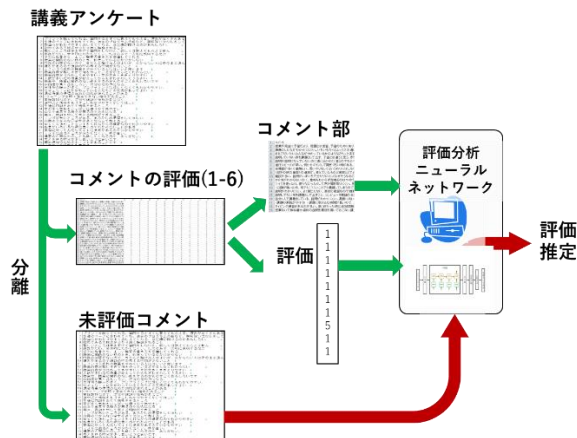


図2 コメント評価の学習と評価の概要

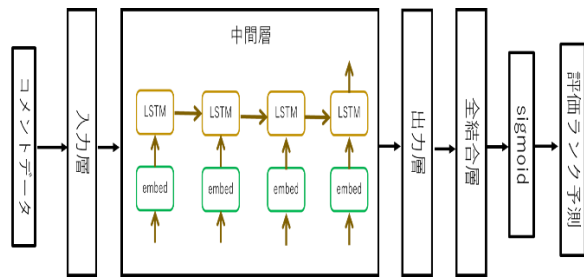


図3 ニューラルネットの構成

4. アンケート自由記述の LSTM による解析

図2に示すように、講義アンケート全体のうち100件を手で評価し、これをシードデータとしてLSTMを用いた機械学習に学習させる。学習モデルは、入力されたコメントに対して6クラス分類の結果の出力を行う分類モデルである(図3)。シードデータである100件のコメントのうち80件を学習データ、20件をバリデーションデータに用いて学習を行う。中間層はembed層、LSTM層で構成し、LSTM層にはdropoutを適応する。パラメータ更新の勾配法にはAdam（Adaptive moment estimation）を用いる。損失の計算は、出力された推定結果と評価ラベルとの平均二乗誤差より計算し、パラメータの更新を行っている。

5. アンケートの推定の評価

12人の評価者ごとに100件のコメントに対して、評価者の評価値と2つの手法による評価の推定値の相関係数の結果を表1に示す。クロズドテストの結果からは、コメントと単語の再帰的手法の精度の平均が0.569に対してLSTMを用いた機械学習の精度の平均0.914でありかなり良好と考えられる。

これに対して、1542件の未評価コメントを推定結果の例を表2に示す。推定結果からは、多くの未評価コメントにおいて、コメントと単語の再帰的评价法の評価平均よりLSTMを用いた機械学習の評価平均の方が低い傾向が見られた。

表1 クローズドテストによるアンケート評価と推定ランクとの相関

評価者	再帰的评价法	LSTM
A	0.773	0.906
B	0.482	0.901
C	0.359	0.937
D	0.573	0.896
E	0.521	0.928
F	0.475	0.913
G	0.284	0.925
H	0.734	0.924
I	0.779	0.871
J	0.535	0.916
K	0.657	0.927
L	0.661	0.919
平均	0.569	0.914

表2 未評価コメントの手法による評価の相違

コメント	再帰的手法による評価平均	LSTMによる評価平均
わかりやすいと思います	4.5	3.5
板書がよい	2.167	3.167
授業が分かりやすい	4.333	3.083
黒板を消すのが速い	2.333	3.000
CGの作り方を学べる	3.333	3.000
数学が実際にどのように利用されているかわかる	4.5	1.75
実技教科なので、演習や課題で技術が身につく	2.833	1.583
課題の答え合わせをしっかりとやってほしい	1.167	1.5
声が小さい。数字を入れた計算を教えてほしい	1.167	1.583
声がおとっていない、ききづらい、生徒をみない	1.583	1.0

6. 今後の課題

手順生成については、閾値の設定やヒューリスティックスが強く、改良の余地が多い。また、未評価コメントの評価推定結果が低くなりやすい傾向が見られる。学習量や未学習情報の影響が大きいと考えられる。また、重要な要素を講義との関連に応じて取り出せるようにしたい。

参考文献

- (1) 大谷, 椎名, “単語ランクに確率分布を用いた自由回答文解析”, 平成29年度(第68回)電気・情報関連学会中国支部連合大会, R17-27-05, pp.1-2 (2017)
- (2) Greff, K. et al, “LSTM: A Search Space Odyssey”, IEEE Transactions on Neural Networks and Learning Systems, Vol.28, Issue10, pp. 2222-2232 (2017)