

ニューラルネットワークによる繰込み学習と単語の共起評価によるアンケート自由回答文と単語の評価

大谷 崇文¹ 椎名 広光²

岡山理科大学大学院 総合情報研究科 情報科学専攻¹

岡山理科大学 総合情報学部 情報科学科²

i18im02ot@ous.jp¹, shiina@mis.ous.ac.jp²

1 はじめに

現在、各大学では、教員の教育・研究指導能力の向上を図るためにFD活動が行われている。FD活動の内容の1つとしてアンケートによる講義評価が行われている。アンケートには様々な種類があるが、そのなかでも自由記述は学生の多様な意見を取得できる重要な方法である。しかし、自由記述式のアンケート評価は難しいと考えられる。

これまで、アンケートの自由記述を評価する手法について、一部のアンケートのシードデータから全体を評価する手法として、コメントと単語の評価を相互に繰り返して評価する相互的手法 [1] やニューラルネットワークを利用したシードデータを学習したアンケートを評価する手法として、コメントの推定評価を新たにテストデータに繰込んで学習を行う手法 [2] を研究している。

しかしながら、単語のみの評価では、単語の現われる状況によって評価に相違があると考えられる。そこで、本研究ではコメント評価から単語の共起ペアを抽出し、共起ペアによる単語の評価とコメントの評価について述べる。

本研究で使用したデータは、岡山理科大学総合情報学部情報科学科の2014年度春学期(4月~9月)の中間段階(15回中8回目の時期)の講義アンケートで実施された自由回答項目を利用した。調査対象とした教員は15人、講義の科目数は41科目、アンケートの回答数は1678個である。なお、講義アンケートのコメントは評価得点の取得をしていない。そこで、2014年度のアンケートよりコメント100件を12人(A~L)の人手で6段階(1~6)のランクを付与して、シードデータとしている。

2 LSTMと学習データの繰込みによる評価

ニューラルネットワークのLSTMを利用してコメントの評価を行った。評価の1から6のランクを0から1に正規化ものとコメントのペアを学習させている。学習モデルはバッチサイズを10、初期の学習率を0.001、LSTM層数を1として、出力を0から1で極性値 p を得るようにし、 p を変換関数 $\lceil \frac{10 \cdot (p+0.2)}{2} \rceil$ で6クラスに分類し推定ランクとしている。

しかし、未評価コメントへのLSTMによる評価法は低い推定をする傾向があることから、ニューラルネットワークを用いて評価したコメントのうち k 個を取り出して、新しい学習データとして繰込み、学習と評価を繰り返していく、学習データの繰込み方式を提案する。以下に、学習データの繰込み方式による機械学習の手順を説明する。

- (1) 学習データの繰込み方式による機械学習の手順
 - (1-1) 初期学習としては、人手による評価の付いたコメントを学習器LSTMを用いて学習し、学習器LSTMで未評価コメントを評価する。
 - (1-2) 学習データに新たに評価した k 個のコメントを併合して新しい学習データを作る。併合できる場合は(1-3)の学習と推定を行う。新しい学習 k 個が学習データに併合できなくなった場合は停止する。
 - (1-3) 併合した新しい学習データを学習器LSTMで学習を行い、新しい学習データに含まれない未評価コメントの推定を行う。推定を行った後、(1-2)に戻る。

本研究では、新たに学習データに繰込むコメント数 $k=100$ 個として評価実験を行った。評価者12人ごとの人手による100件のコメントのシードデータに対するクロズドテストの結果を表1に示す。また、未評価コメントの全部の平均は、相互評価法とLSTMによ

表 1: コメント評価のクロズドテストにおける精度比較

評価者	相関係数				MSE			
	相互評価法	LSTM	線込み学習	共起評価	相互評価法	LSTM	線込み学習	共起評価
A	0.773	0.864	0.968	0.935	0.062	0.024	0.0092	0.440
B	0.482	0.813	0.968	0.855	0.120	0.120	0.0092	0.810
C	0.359	0.894	0.958	0.779	0.072	0.010	0.0092	0.850
D	0.573	0.813	0.968	0.895	0.102	0.023	0.0064	0.460
E	0.521	0.846	0.960	0.907	0.071	0.011	0.0064	0.380
F	0.475	0.850	0.964	0.879	0.114	0.014	0.0068	0.630
G	0.284	0.841	0.962	0.743	0.067	0.008	0.0044	0.690
H	0.734	0.907	0.945	0.838	0.075	0.019	0.0156	1.060
I	0.779	0.756	0.965	0.863	0.050	0.028	0.0076	0.740
J	0.535	0.877	0.962	0.879	0.120	0.017	0.0080	0.720
K	0.657	0.889	0.984	0.895	0.047	0.012	0.0024	0.410
L	0.661	0.841	0.841	0.980	0.058	0.016	0.0036	0.250
平均	0.569	0.850	0.965	0.868	0.080	0.025	0.0075	0.628

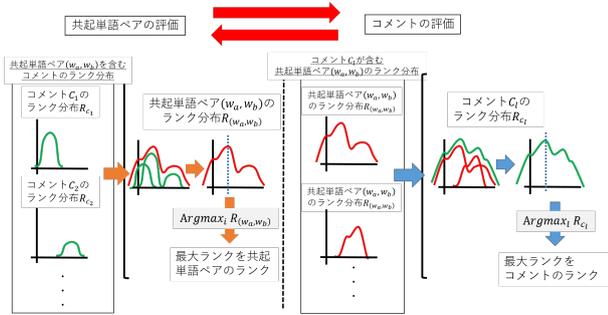


図 1: 共起単語ペアによる評価

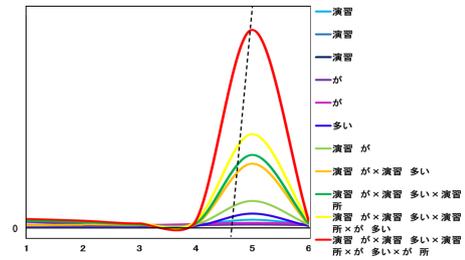


図 2: コメントランク分布の計算と分布例

る手法を単独で使用した場合は平均 2.790 と 2.316 に対して、学習データの線込み方式では 3.408 となっており、単独で使用するよりも高い評価であった。

3 コメントから共起単語ペアへの推定評価

評価済みのコメントに含まれる 2 つの単語 (以降、共起語ペア (w_s, w_t)) を抽出し評価を行う。本研究ではニューラルネットワークを用いた線込み学習による評価手法でのコメント評価を初期値として利用して共起単語ペア評価し、コメントの評価とコメントに含まれる単語の共起単語ペアの評価を交互に適応しながら評価する方法を用いている (図 1)。

(1) コメントを構成する共起単語ペアの評価推定：評価済みコメントに含まれる共起単語ペアの評価推定を行う。

(1-1) コメントから共起単語ペアを抽出し、線込み学習によるコメントの評価ランクを共起単語ペアの推定評価とする。

(1-2) 共起単語ペアごとに評価ランクの頻度から共起単語ペアの評価ランク分布を作成する。対象となる共起単語ペア (w_s, w_t) を含むコメントの評価ランクを $i(=1, \dots, M, M=6)$ として評価ランクの出現ごとにランク i を中心とした $\mu_i(=i)$, 分散を σ^2 とした正規分布 $\phi(x; \mu_i, \sigma^2)$ を作成する。

(1-3) 共起単語ペアの全評価ランク $(=1, \dots, M, M=6)$ の正規分布を結合して混合正規分布を作成し、共起単語ランク分布とする。混合正規分布の混合数 (ランク数と同じ) を M , パラメータ α_i をランク i に対する正規分布の重みとした混合正規分布を $p_{(w_s, w_t)}(x)$ で定義する。初期値については、 $\alpha_i = \frac{1}{M}, \sum_{i=1}^M \alpha_i = 1$ となるように定義する。また、共起単語ペア (w_s, w_t) がランク i となる回数を $N((w_s, w_t), i)$ とする。

$$p_{(w_s, w_t)}(x) = \sum_{i=1}^M \alpha_i \cdot \phi(x; \mu_i, \sigma^2) \cdot N((w_s, w_t), i)$$

表 2: 未評価コメントのランク推定の平均

コメント	相互的評価法	LSTMのみ	繰込み学習	共起評価
分かりやすく良いと思う	4.500	2.583	4.500	4.500
声が聞き取りやすい, 板書が見やすい	2.750	1.417	2.750	2.583
授業が分かりやすく良い.	4.500	2.000	4.500	4.500
黒板を消すのが速い	2.333	3.000	2.412	2.333
CGの作り方を学べる	3.333	3.000	3.333	3.333
数学が実際にどのように利用されているかがわかる	4.500	1.750	4.417	4.500
実技教科なので, 演習や課題で技術が身につく	2.833	1.583	2.750	3.083
課題の答え合わせをしっかりとってほしい	1.167	1.500	1.500	4.167
声が小さい. 数字を入れた計算を教えてほしい	1.167	1.583	1.167	1.167
声がおっていない, ききづらい, 生徒をみない	1.583	1.000	1.500	1.417
未評価コメントの平均	2.867	1.942	2.883	3.158

表 3: 単語のみと共起単語ペアの評価ランク

	単語例 1(授業)		単語例 2(黒板)		単語例 3(先生)	
単語のみ	4.50		3.67		4.66	
共起単語の評価	解説	1.25	講義	1.92	解説	1.25
	予習	1.25	声	2.00	強い	1.25
	雑談	1.58	大きく	2.17	人生	1.25
	あたらしい	1.75	字	2.17	ぐだぐだ	1.67
	速い	1.92	速い	2.33	やる気	1.92
	声	1.92	消し	2.42	たのしい	2.33
	見づらい	2.08	プリント	2.50	演習	2.75
	改善	3.25	文字	2.67	自慢	2.75
	質問	4.00	使い方	3.67	説明	4.08
	熱意	4.00	スクリーン	3.83	スライド	4.50
	黒板	4.08	授業	4.08	パソコン	4.50
	丁寧	4.33	課題	4.42	スクリーン	4.67
	提出	4.58	教科書	4.50	まとめ	4.67
	テスト	4.67	説明	4.50	丁寧	4.67
	演習	4.67	板書	4.50	板書	4.67
	教科書	4.67	資料	4.50	面白い	4.67
	資料	4.67	演習	4.67	楽しい	4.92
	使える	4.83	良い	4.75	実際	4.92
	社会	4.83	解答	4.75	親しみ	4.92
実際	4.92	意欲	4.75	笑顔	5.25	

表 4: 各手法ごとの単語の評価ランク

評価者	2014年					
	相互的評価		繰込み学習		共起評価	
	平均	分散	平均	分散	平均	分散
課題	4.33	1.22	4.42	0.91	4.50	0.75
丁寧	4.75	0.19	4.75	0.19	4.75	0.19
声	1.75	0.35	1.58	0.41	1.75	0.35
授業	4.50	0.75	4.50	0.75	4.50	0.75
話	3.00	2.17	3.00	2.17	3.00	2.17
簡単	4.50	0.42	4.58	0.41	4.58	0.41
板書	2.58	0.91	2.17	0.31	2.75	1.85
教員	2.67	2.56	2.17	0.64	2.42	0.41
速い	2.25	0.52	2.25	0.52	2.00	0.50
講義	4.50	0.75	4.75	0.19	4.67	0.22
面白い	4.92	0.58	4.92	0.58	4.75	0.35
プリント	4.67	0.22	4.67	0.22	4.67	0.22

(1-4) 共起単語ランク分布から最大ランクを共起単語ペアの推定評価ランクとする.

$$R_{(w_s, w_t)} = \operatorname{argmax}_{i=1, \dots, M} p_{(w_s, w_t)}(i)$$

(2) コメントの評価ランクの再推定: 共起単語ランク分布からコメントのコメントランク分布を作成する. コメントランク分布から, 最大ランクをコメントの推定値とする.

(2-1) コメント c_l のコメントランク分布 P_{c_l} は, コメントを構成している単語から作られる全ての共起単語ペア (w_s, w_t) に対する共起単語ランク分布の確率

$p_{(w_s, w_t)}$ に 1 を加算する. 次に共起単語ペアの各ランクでの確率同士を掛け合わせる.

$$\prod_{w_s, w_t \in c_l} (p_{(w_s, w_t)}(i) + 1)$$

次に, 確率同士を重みを付けて掛け合わせたものを混合正規分布で表し, その重みの合計 $\sum_{i=1}^M \beta_i = 1$ となるように正規化する. コメント例の「演習が多いところ」に対するコメントランク分布と共起単語ペアのランク分布との関係を図 2 に示す.

$$P_{c_l}(x) := \prod_{w_s, w_t \in c_l} \beta^k \cdot (p_{(w_s, w_t)}(i) + 1) = \sum_{i=1}^M \beta_i \cdot \phi(x; \mu_i, \sigma^2)$$

(2-2) コメントランク推定値 R_{c_l} を (1-4) と同様にコメントランク分布から最大ランクを計算する.

$$R_{c_l} = \operatorname{argmax}_{i=1, \dots, M} P_{c_l}(i)$$

(3) コメントと共起単語ペアの繰り返し評価: コメントに対する評価推定とそれを構成する単語からなる共

表 5: 単語「授業」の共起単語評価の評価者ごとの類似度

評価者	A	B	C	D	E	F	G	H	I	J	K	L	平均
A	1.000	0.908	0.898	0.885	0.916	0.878	0.865	0.911	0.921	0.903	0.903	0.908	0.900
B	0.908	1.000	0.962	0.921	0.962	0.943	0.947	0.933	0.927	0.955	0.945	0.919	0.938
C	0.898	0.962	1.000	0.924	0.972	0.949	0.964	0.932	0.928	0.947	0.948	0.932	0.941
D	0.885	0.921	0.924	1.000	0.929	0.929	0.931	0.894	0.923	0.919	0.918	0.939	0.919
E	0.916	0.962	0.972	0.929	1.000	0.960	0.960	0.934	0.954	0.959	0.944	0.935	0.948
F	0.878	0.943	0.949	0.929	0.960	1.000	0.955	0.921	0.939	0.939	0.926	0.915	0.932
G	0.865	0.947	0.964	0.931	0.960	0.955	1.000	0.897	0.926	0.921	0.947	0.292	0.931
H	0.911	0.933	0.932	0.894	0.934	0.921	0.897	1.000	0.912	0.939	0.915	0.902	0.917
I	0.921	0.927	0.928	0.923	0.954	0.939	0.926	0.912	1.000	0.928	0.946	0.933	0.931
J	0.903	0.955	0.947	0.919	0.959	0.939	0.921	0.939	0.928	1.000	0.928	0.913	0.932
K	0.903	0.945	0.948	0.918	0.944	0.926	0.947	0.915	0.946	0.928	1.000	0.943	0.933
L	0.908	0.919	0.932	0.939	0.935	0.915	0.929	0.902	0.933	0.913	0.943	1.000	0.924

起単語ペアに対する評価推定を交互に繰り返す。繰り返しが停止後、コメントランク分布 P_{c_l} と共起単語ランク分布 $p(w_s, w_t)$ から最大ランクをコメントと単語の最終的なランク推定値とする。

(3-1) 各コメントの評価ランクの再推定: (2) と同様に、単語の各ランクの確率に 1 を加算し、掛け合わせてランク分布とする。

(3-2) 共起単語ペアの評価ランクの再推定: (1) と同じように、全コメントに対するコメントランクを用いて、コメントを構成する単語からなる共起単語ペアランク分布を更新する。共起単語ペア (w_s, w_t) の単語ランク分布 $p(w_s, w_t)(x)$ をコメントランク分布 P_{c_l} に重みを付けて作成する。 $W((w_s, w_t))$ は、共起単語ペア (w_s, w_t) を含むコメント集合を表すものとする。また、混合正規分布で表され、その重みの合計 $\sum_{i=1}^M \gamma_i = 1$ となるように正規化する。

$$p(w_s, w_t)(x) = \sum_{c_l \in W((w_s, w_t))} \gamma^l \cdot P_{c_l}(x) \\ = \sum_{i=1}^M \gamma_i \cdot \phi(x; \mu_i, \sigma^2)$$

4 評価

(1) コメント評価: 共起単語ペアによる評価と他手法とのコメント評価を表 2 に示す。相互的評価法、繰込み学習との差はあまり見られない。表 1 のクローズドテストにおける精度比較では、MSE は平均が 0.628 と他の手法と比べて誤差が大きくなっている。

(2) 共起単語ペアの評価: 共起単語のペアの評価例を表 3 に、単語の評価例を表 4 に示す。表 3 の単語例 2 のように、「黒板」は単語のみでは評価ランクは 3.67

で中央値付近であるが、「演習」(4.67) のような推定評価ランクの高い単語との共起単語ペアでは、4.50 と評価が上昇する。一方、推定評価ランクの低い単語との共起単語ペアでは、低い評価となる。また、単語の出現状況によって評価が変化していると考えられる。

(3) 共起単語ペアによる単語評価の類似度: 共起単語ペアの評価から単語に対しての評価者ごとのコサイン類似度の例 (単語「授業」) を表 5 に示す。

出現した単語全てでは、最大値は 1、最小値は 0.84764、単語の類似度の平均は 0.977494、分散は 0.001078 であった。全体的に評価の類似度は高い傾向が見られ、評価者間の単語の評価の揺れが少ない。

5 今後の課題

単語・コメントの極性評価の精度の改善を行いたいと考えている。ニューラルネットワークから直接単語や共起単語ペアの極性評価を行うにあたり、BERT との関連を調査したい。

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, 2017.
- [2] 大谷, 椎名, 小林, “ニューラルネットワークと相互評価法を用いた自由記述アンケートからの単語評価”, 言語処理学会第 25 回年次大会 NLP2019, pp. 1375-1378, 2019.