

エントリーシートにおける不適切回答の自動添削手法

沖村 秀哉 目良 和也 黒澤 義明 竹澤 寿幸

広島市立大学大学院情報科学研究科

E-mail: okimura@ls.info.hiroshima-cu.ac.jp, {mera, kurosawa, takezawa}@hiroshima-cu.ac.jp

1. はじめに

大学生の就職率は近年低迷しており，そのような状況の下，就職面接に不安を感じている学生も多い．なかでも履歴書・エントリーシート(ES)の提出は就職活動において必須であり，多くの企業に提出する必要がある．

ESの作成を支援するシステムとして，三井所らはESの自動生成システムを開発している[1]．このシステムでは自己PR，志望動機，学生時代に取り組んだこと，長所の4つの設問と回答文字数，企業名を選択すると，その企業に合ったES候補が自動生成され表示される．ESの案を気に入ればES候補の編集を行う．

このシステムを利用することでESを短時間で作成することができる．ただ，就職活動はESの提出だけではない．自動生成されたESを提出するだけだと，面接で掘り下げた質問が来た場合に回答に詰まる可能性がある．自分で書き，添削を受け，修正を繰り返すことで頭が整理され，面接でもきちんと対応することができる．このような観点からESの添削は重要である．ESの添削は就職支援課で行われているが，学生の人数が多い大学では履歴書・ESの添削を全学生に対して何度も行うことは困難である．

そこで本研究では文章構成の不適切さに着目して，不適切な回答を検出し，添削文を自動生成する手法を提案する．本手法は回答を文単位に区切り，各文の機能にあったタグを自動的に付与する．例えば「喜びを感じた」という文には“考え”というタグが付与される．付与した後，タグの並びに着目することで回答の文章構成を捉え，不適切な回答を検出する．

2. タグの並びを用いた回答の自動添削手法

本研究では，回答中の各文に対して付与したタグの並びを解析することで回答の不適切な箇所を検出する手法を提案する．図1に本手法の処理の流れを示す．本手法の処理では，まず入力された学生の回答を文単位に分割する．その後機械学習器を用いてタグの自動付与を行う．タグは修辞構造理論[4]を参考に回答中の文の機能に注目し，その機能にあったタグを定義する．本研究では，複数の就職活動関連の書籍[5][6][7]を参考に，4種類のタグと9種類の詳細タグを定義した(表1)．

“結論”とは質問に対する自分の主張をまとめ

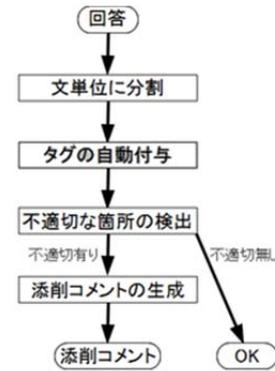


図1: 処理の流れ

表1: 付与したタグと定義

タグ	詳細タグ	定義
結論		質問に対する自分の主張
補足		結論の補足説明
具体例	状況	過去の経験の状況 または役割の説明
	考え	自分の考えていたことや 感じたこと
	行動	自分が直接行ったこと
	結果	状況の変化や結果の説明
締め	学習	何かを得たとき
	将来	未来の話をしたとき
	強調	結論を強調するとき

た文を表している．“結論”だけではわからない部分の説明をしているタグが“補足”である．“具体例”では主に過去の経験が語られる．“具体例”はさらに詳細にタグを付与することができるため，階層構造になっている．“具体例”に属する詳細タグは4つある．経験したことの状況や自分の役割の説明を述べている“状況”，自身の価値観や考え，感想を述べている“考え”，直接行った工夫や行動は“行動”，状況の変化を述べている文が“結果”となる．“締め”も“具体例”と同じように階層構造になっており，3つのタグが属している．“学習”は志願者が何かを得たことを述べた文に付与されるタグである．“将来”というタグは今後のことが語られた時に付与されるタグである．“強調”は“結論”を強調している文に対して付与されるタグである．“結論”と同様の言葉や類似する言い回しが

用いられている回答が多い。

次に付与されたタグの並びに着目し、不適切な箇所の検出を行う。不適切な箇所が検出された場合は不適切な箇所にあった添削文を生成し、表示する。生成する添削文については5章で述べる。

3. 機械学習を用いたタグ自動付与

本研究ではタグの自動付与に機械学習を用いる。学習のための素性としては、各タグの特徴語の出現回数を用いる。特徴語は tf*idf と情報利得の2種類の手法を用いて選定する。機械学習器は SVM, ニューラルネットワーク (NN), CRF の3種類の手法を比較する。各手法に用いる特徴語の数は、予備実験により決定した。

4. 添削文の自動生成

不適切な回答の内容に合った添削文を自動生成するために、インターネット掲示板から収集した回答と添削内容の対を分析することで不適切な回答の内容と生成する添削文の内容を決定する。まず正解となる適切な回答を収集、分析し、適切な回答の定義を“情報が過不足なく述べられており、結論→具体例→締めめのタグの並びである回答”とした。

本手法で生成される添削文の種類を表2に示す。情報の不足は“結論”, “状況”, “考え”, “行動”, “結果”, “締め”のいずれかのタグが不足している場合に生成される。情報の過多の添削文は同じ詳細タグが3つ以上連続して付与されており、同じ内容の文が長くなっている状態において生成される。タグの順序誤りで生成される添削文はタグの並びが適切な場合以外に生成される。長文は一文中の単語数が60語以上の場合に生成される。抽象度の高い表現とは人によって捉え方が異なる表現のことである。不本意な表現とは面接官に誤解を与えかねない表現のことである。注意喚起とは使用頻度が高く、使用の際注意が必要な表現のことである。バイト敬語とは接客場面において若い店員がお客様に対して使う特徴的な接客表現である。

表2: 生成される添削文の種類

添削文の種類	パターン数	
情報の不足	6	
情報の過多	1	
タグの順序誤り	4	
長文	3	
不適切語の使用	抽象度の高い表現	63
	不本意な表現	11
	注意喚起	22
	バイト敬語	5

5. 添削文の自動生成

5.1. 自動タグ付与実験

学習データとしてインターネット掲示板から収集した志願者の回答698回答(3980文), テストデータとしてインターネット掲示板から収集した回答193回答(1269文)に作業者1名(男子大学院生1名)がタグを付与したものを使用した。

本実験ではデータ数が少ない“補足”を除外して行った。“補足”以外のタグ3種類を“大分類”と呼び, “補足”以外の詳細タグ8種類を“詳細分類”と呼ぶ。自動タグ付与実験では特徴語の選定方法と各分類器の組み合わせを比較した。

5.1.1. 大分類の分類結果

表3に tf*idf, 表4に情報利得を用いて構築した素性について分類した結果をそれぞれ示す。

tf*idf, 情報利得どちらの特徴語収集方法を使用した場合でも各分類器の値にあまり差が見られなかった。これは3値分類のため同じような特徴語が多いことが原因と考えられる。

表3: tf*idf の分類結果

分類器		結論	具体例	締め	正解率
SVM	精度	0.80	0.84	0.68	0.83
	再現率	0.54	0.97	0.29	
	F値	0.64	0.90	0.41	
NN	精度	0.63	0.86	0.56	0.80
	再現率	0.53	0.89	0.58	
	F値	0.57	0.87	0.57	
CRF	精度	0.98	0.91	0.74	0.91
	再現率	0.90	0.98	0.43	
	F値	0.94	0.94	0.54	

表4: 情報利得の分類結果

分類器		結論	具体例	締め	正解率
SVM	精度	0.83	0.87	0.77	0.86
	再現率	0.62	0.96	0.51	
	F値	0.71	0.91	0.61	
NN	精度	0.74	0.90	0.70	0.86
	再現率	0.69	0.92	0.66	
	F値	0.72	0.91	0.68	
CRF	精度	0.97	0.96	0.58	0.92
	再現率	0.97	0.96	0.58	
	F値	0.97	0.95	0.63	

表 5 : tf*idf の分類結果

分類器		結論	状況	考え	行動	結果	学習	将来	強調	正解率
SVM	精度	0.65	0.45	0.56	0.40	0.78	0.52	0.50	0.00	0.51
	再現率	0.61	0.62	0.40	0.49	0.47	0.65	0.23	0.00	
	F 値	0.63	0.52	0.46	0.44	0.59	0.58	0.32	0.00	
NN	精度	0.50	0.45	0.53	0.36	0.58	0.53	0.00	0.00	0.46
	再現率	0.54	0.40	0.34	0.53	0.62	0.80	0.00	0.00	
	F 値	0.52	0.43	0.42	0.43	0.60	0.64	0.00	0.00	
CRF	精度	0.76	0.51	0.37	0.41	0.39	0.57	0.75	0.00	0.51
	再現率	0.90	0.65	0.14	0.52	0.46	0.38	0.46	0.00	
	F 値	0.82	0.57	0.20	0.46	0.42	0.45	0.57	0.00	

表 6 : 情報利得の分類結果

分類器		結論	状況	考え	行動	結果	学習	将来	強調	正解率
SVM	精度	0.77	0.49	0.65	0.48	0.67	0.78	0.66	0.75	0.59
	再現率	0.69	0.68	0.42	0.56	0.58	0.72	0.59	0.26	
	F 値	0.72	0.57	0.51	0.52	0.62	0.75	0.62	0.39	
NN	精度	0.70	0.46	0.55	0.47	0.66	0.68	0.53	0.00	0.55
	再現率	0.73	0.54	0.43	0.48	0.65	0.70	0.67	0.00	
	F 値	0.71	0.50	0.48	0.48	0.66	0.69	0.59	0.00	
CRF	精度	0.85	0.50	0.47	0.40	0.62	0.63	0.76	0.36	0.56
	再現率	0.97	0.74	0.18	0.46	0.53	0.46	0.56	0.39	
	F 値	0.91	0.59	0.26	0.43	0.57	0.53	0.65	0.38	

5.1.2. 詳細分類の分類結果

表 5 に tf*idf, 表 6 に情報利得を用いて構築した素性について分類した結果をそれぞれ示す。

大きな差がないまでも全体的に tf*idf よりも情報利得を用いた場合の方が高い結果が得られた。

以上のことから特徴語選定において情報利得を用いた方が有効であることが確認できた。また分類器では大分類では CRF が、詳細分類では SVM が有効であることが確認できた。

5.2. 添削文自動生成実験

5.2.1. 添削文の出力例

表 7 に添削文の出力例を示す。この入力回答では“考え”が不足している。そのため“考え”が付与されるように I の添削文が生成された。またこの回答には“締め”も付与されていない。そのため II の添削文が生成された。6 行目に“様々”という抽象度の高い表現が使用されている。そのため III のような具体化を指示する添削文を生成された。

5.2.2. 生成された添削文の傾向

自動生成される添削文の傾向を分析した。使用したデータは自動タグ付与実験で用いたテストデータと同一である。生成した添削文の内訳を表 8 に示す。

表 8 から情報の不足に関する添削文の割合が多いこと

表 7 : 添削文の出力例

<p>【入力回答】私は競争を恐れない人間です。私は中学から始めたバスケットボールを 10 年間続けています。しかし、大学での部活は簡単にはいきませんでした。今までとは比べ物にならないレベルの高さでした。わたしは他のプレイヤーと自分では何が違うのかを分析しました。得た結果を基に、様々なトレーニングを行いました。そして、スタメンを勝ち取ることができました。最上級生になった今では後輩との競争に負けないように、プレーだけでなく、チームの輪を大事にしています。</p>
<p>【添削文】</p> <p>I : “何故そうしたのか”というあなたの内面まで突っ込んで表現してみましょう。気持ちを添えることで、“あなたらしさ”がより伝わりますよ。</p> <p>II : 最後に経験を振り返っての”学び”を追加しましょう。</p> <p>III : 6 文目:様々:便利な言葉ですが、その多様性、中身について具体的に示すようにしてください。</p>

がわかる。これは字数制限のため本来書くべき文章を省略したためだと考えられる。

不適切語では抽象度の高い表現と注意喚起に関する

る添削文が多く生成された。これらは志願者が使うことが多いと添削文に記載してあった。この結果から抽象度の高い表現と注意喚起は志願者の使用頻度が高いということがわかる。

表 8：傾向分析のために生成した添削文の内訳

添削文の種類		生成した添削文数
情報の不足		323
情報の過多		13
タグの順序誤り		20
長文		25
不適切語	抽象度の高い表現	261
	不本意な表現	7
	注意喚起	161
	バイト敬語	29

5.2.3. 生成された添削文の精度

生成された添削文がどの程度正しいのか調査する実験を行った。添削文生成に使用した回答は被験者 4 名の履歴書から収集した回答 8 回答, 計 36 文に作業者 1 名がタグを付与したものである。実験に使用したデータは収集した回答に対して生成された添削文 30 文である。内訳は“情報の不足”が 23 文, “抽象度の高い表現”が 4 文, “注意喚起”が 3 文である。

人手での正解か否かの判定は本学の就職支援課の方々にご協力頂く。正解か否かの判定では“的を射ている”, “どちらかと言えば的を射ている”, “どちらでもない”, “どちらかと言えば的を射ていない”, “的を射ていない”の 5 段階評価を用いた。生成した添削文の精度は式(1)で求められる。

$$\text{精度} = \frac{\text{人手で正解と判定した添削文の数}}{\text{生成された添削文の数}} \quad (1)$$

6. おわりに

本研究では、就職面接における志願者の回答の文章構成の不適切さに着目して、不適切な回答を添削する手法を提案した。本手法では、まず入力された学生の回答を文単位に分割し、機械学習を用いて各文にタグの自動付与を行う。そして付与されたタグの並びに着目し、不適切な箇所の検出を行う。不適切な箇所が検出された場合は不適切な箇所に合った添削コメントを生成し、表示する。本研究では回答の文章構成を捉えるために 4 種類のタグ (“結論”, “補足”, “具体例”, “締め”) と 9 種類の詳細タグ (“結論”, “補足”, “状況”, “考え”, “行動”, “結果”, “学習”, “将来”, “強調”) を定義した。自動タグ付与では tf*idf と情報利得を用いて特徴語を定義し、特徴語の出現回数を機械学習することでタグの自動付与を行った。機械学習には 3 種類の分類器 (SVM, NN, CRF) を用い、比較実験を

行った。

自動タグ付与実験では特徴語の選定方法と各分類器の組み合わせを比較した。大分類での実験結果では情報利得と CRF の組み合わせの正解率(=0.92)が最も良いことが確認できた。詳細分類での実験結果では情報利得と SVM の組み合わせの正解率(=0.59)が最も良いことが確認できた。また不適切な回答を分析した結果に基づいて、5 種類の添削文(情報の不足, 情報の過多, タグの順序誤り, 長文, 不適切語の使用)を定義した。添削文の自動生成では入力された回答に 5 種類の添削文が該当するかを調査し、該当する箇所がある場合、該当箇所とそれに対応する添削文を表示した。

今後の課題としては、自動タグ付与については、特定の単語に依存しない新たな素性の検討を行い、さらに精度を向上させることが必要である。添削文の自動生成については、実験に用いたデータが少ないという課題があるため、多くのデータを用いた実験が必要である。

参考文献

- [1] 三井所健太郎ら, WEB インターフェースによる就職活動支援システムに関する研究, 情報処理学会研究報告会, Vol.17, pp.1-6(2009).
- [2] AttaliYigal, Burstein Jill. Automated essay scoring with e-rater v.2. Journal of Technology, Learning and assesment,(2007).
- [3] 石岡恒憲ら, 小論文およびエッセイの自動評価採点における研究動向, 人工知能学会誌, Vol.23, No.1,pp.17-24(2008).
- [4] Mann,W.C.,Thompson, and S.A. Rhetorical Structure Theory: Description and Construction of Text Structures. in Kempen, G.(ed.). Natural Language Generation: New Results in Artificial Intelligence, Psychology, and Linguistics, pp.85-96, 1987.
- [5] 横瀬勉, 人事のプロは学生のどこをみているか, PHP 研究所, 2010.
- [6] 高田裕明, 全員内定!, 2012.
- [7] 伊東明, 人気企業を目指す人のためのハイレベル面接術—面接官の心の裏を知り尽くした心理学者が書いた! 「100 人中の 1 人」になる極意, 2010.
- [8] 面接データベース よくある質問と回答例 (新卒・転職), <http://mennsetu.com/index.html> (2013/10/15 アクセス)
- [9] みんなの就職活動日記, [http://www.nikki.ne.jp/\(2013/10/15](http://www.nikki.ne.jp/(2013/10/15) アクセス)
- [10] V.Vapnik, "The Nature of Statistical Learning Theory", Springer, (1995).
- [11] J.Lafferty, A.McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," In Proc. of ICML, pp.282-289, 2001.