

就職面接における不適切回答の検出のための自動タグ付与手法

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 〒731-3194 広島市安佐南区大塚東 3-4-1

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

あらまし 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。そのため多くの大学では就職支援課など就職活動全般を支援する場所が設けられているが、支援課の人員も限られているため1人の学生に何度も指導することが困難である。本論文は回答中の各文に対してタグを付与し、その並びの特徴により不適切な回答を検出する手法を提案する。本論文では9種類のタグを自動的に分類するために、bag-of-words手法に基づいて作成した素性を用いて機械学習を行う。Support vector machine, ニューラルネットワーク, Conditional Random Fields(CRF)を比較した結果, CRFを用いた手法の正解率(55%)が最も良いことが確認できた。
キーワード 就職面接, タグ, 自動タグ付与, bag-of-words, CRF

Auto-tagging Method to Detect Inadequate Answer at Job Interview

Syuya OKIMURA Kazuya MERA Yoshiaki KUROSAWA Toshiyuki TAKEZAWA

Graduate School of Information Sciences, Hiroshima City University

3-4-1 Ozuka-higashi, Asa-minami-ku, Hiroshima, 731-3194 Japan

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

Abstract Although importance of job interview is increasing in employment exam, it is difficult to make adequate response for inexperienced job seekers. Therefore, many universities give opportunities to train job interview, however, there are not enough staff for simulation. We propose a method to point out inadequate answer according to grammatical features. This method annotates nine types of sentence tags automatically and checks the sequence of the tags. The tagging is carried out by using bag-of-word method. SVM, neural network, and CRF methods were compared and the accuracy of annotation by using CRF was the best (55%).

Keyword Job Interview, sentence tag, auto-tagging, bag-of-words, CRF

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。そのため多くの大学では就職支援課など就職活動全般を支援する場所が設けられているが、支援課の人員も限られているため1人の学生に何度も指導することが困難である。特に面接は練習を希望する学生が多いことから、近年、手軽に面接の練習ができる模擬面接システムに対する需要が高まっている。就職サイト「マイナビ 2014」では「リアル模擬面接シミュレーター モギメン」[1]と呼ばれる模擬面接を行うシステムが開発されている。このシステムは面接官の質問がランダムに出現し、その質問に使用者が答えるシステムである。また、模擬面接を行うアプリケーションとして「就活面接トレーナー」[2]が開発されている。このアプリケーションはスマートフォンのマイクでランダムに出題される質問に答えると自分の音声を波形で見ることができる。これらのシステムでは実際の質問を想定した練習や声の大きさ等を主観で評価することはできるが、回答の内容の評価や添削機能は存在しない。

そこで我々の研究グループでは、現在、就職面接における不適切な回答を検出する手法について研究を進めている。不適切な回答か否かを判断するには、一般

的にバーバルな情報とノンバーバルな情報の両方が参照される。バーバルな情報としては、若者言葉の使用、文章構成の不適切さ、主題の一貫性の無さなど、ノンバーバルな情報としては、声の大きさ、話速、アイコンタクトやボディランゲージの有無などが挙げられる。

本研究ではバーバル情報の一つである文章の構成の不適切さに着目して、不適切な回答を検出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. 従来研究

本研究は文字列で表した志願者の回答の文章構成の不適切さを検出することを目的としている。この課題は論理構成を評価する小論文の自動採点と類似している。

代表的な小論文の自動採点システムとして、英文を対象にした ETS の e-rater が挙げられる[3]。e-rater は、文章上から多数の素性を抽出し、重回帰分析によって得たモデルにしたがって小論文(essay)を採点する。そ

の精度は 97% である。一方、和文を対象にしたシステムとして、石岡らは e-rater を参考に jess というシステムを開発している [4]。jess は、接続表現を手掛かりにして隣接 2 文間の関係に着目することで、小論文を論理性に関して採点することができる。しかし、jess における手法では、接続表現を含まない文や遠隔的な論理展開を含む文章への対応が問題となる [5]。

また、修辞構造理論 (Rhetorical Structure Theory, RST) [6] は、テキスト中の文の機能を特定し文間の依存関係を特定するモデルである。テキスト中の文は他の文に対し、説明、例示、帰結などの機能を持って存在する。修辞構造理論は、文と文に結束性があると認められた文章の分析から、文と文の関係を形式化して記述した理論である [7]。

3. タグの並びを用いた不適切回答検出手法

前章で小論文の自動採点システムである jess を紹介した。しかし、就職面接の回答では接続表現を含まない文の方が多いため、接続表現を手掛かりにすることで論理性を評価することは難しい。

そこで本研究では、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。その後機械学習を用いてタグの自動付与を行う。タグは修辞構造理論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。本研究で定義したタグについては 4 章で述べる。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なければ添削コメントを生成せず終了し、検出された場合は不適切な箇所にあった添削コメントを生成し、表示する。

本論文では、本提案手法の処理のうち、タグの自動付与処理について詳述する。

本手法における適切な回答とは情報が不足なく述べられており、図 2 のタグの並びである回答と定義する。回答の初めには“結論”が付与される。“補足”は“結論”の補足説明を行う時に付与されるタグであるため、省略しても良い。その後“具体例”の“状況”が続く。“行動”または“考え”のタグが付与される。次に“結果”が付与され、最後に“締め”の中の“学習”、“将来”、“強調”のいずれかが付与される。

本手法を用いた不適切箇所検出の例を示す。表 1 に不適切な回答に対するタグの付与例とシステムの添削例を示す。表 2 に不適切でない回答に対するタグの付与例とシステムの添削例を示す。

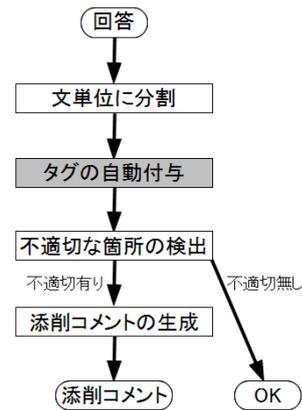


図 1：処理の流れ

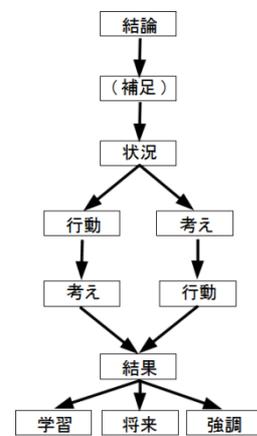


図 2：適切なタグの並び

表 1：不適切な回答に対するタグ付与例

発言者	タグ	詳細タグ	付与例
面接官	質問	質問	学生時代に最も力を入れたことは何ですか
	結論	結論	私は軟式野球サークルに入っていました。
学生	具体例	状況	同じリーグには長年のライバルチームがあり、いつもこのチームに勝ち越すことを目標にしていました。
	具体例	状況	大学最終シーズンは、1勝ずつの引き分けのまま最終戦を迎えました。
	具体例	結果	そしてその最後の試合ではみんなで団結し、なんとか勝つことができました。
システム	添削	添削	あなたが行ったことを述べてください。この経験から学んだことを述べてください。

表 2：不適切でない回答に対するタグ付与例

発言者	タグ	詳細タグ	付与例
面接官	質問	質問	学生時代に最も力を入れたことは何ですか
学生	結論	結論	私は古都散策会というサークル活動に所属しています。
	具体例	状況	入った当初はメンバーは5人でほとんど活動もなく、全く楽しめませんでした。
	具体例	考え	これではいけないと思い、積極的に活動することを決意しました。
	具体例	行動	私たちの活動を知ってもらおうと、展示会を開きました。
	具体例	結果	これが好評で、今では季節毎の展示会を行っています。
	具体例	結果	また、メンバーも50人ほどに増えました。
	締め	将来	今後もこの積極的な姿勢を仕事に活かしていきたいです。
システム	添削	添削	回答の文章構成は間違っていない

4. タグの定義

本研究では、複数の就職活動関連の書籍[8][9][10]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3).

表3: 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

“結論”とは質問に対する自分の主張をまとめた文を表している。“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。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。

表4に就職面接の回答とタグの付与例を示す。

5. 機械学習を用いたタグ自動付与

本研究では，回答を文単位に区切り，各文に対して前章で述べたタグを自動的に付与する。タグの自動付与には機械学習を用いる。学習のための素性としては，bag-of-words 手法に基づき，各タグの特徴を表す語（特徴語）の出現回数を用いる。本研究で用いる特徴語を定義するため，各タグにおける単語の出現傾向について調査を行った。調査ではまず以下の式を用い，学習コーパス中の各単語の出現度を計算した。

表4: 回答とタグの付与例

発言者	タグ	詳細タグ	付与例
面接官	質問	質問	自己PRをしてください。
	結論	結論	私の強みは人のよさです。
	具体例	状況	私は個人経営の喫茶店でアルバイトをしています。
	具体例	状況	お昼時でもあまりお客様が来られない所でした。
	具体例	考え	私はお客様に寄り添った接客を心がけるようにしました。
	具体例	行動	そこで私は店の前でチラシを配る工夫を行いました。
	具体例	結果	その結果、多くのお客様にきていただくことができました。
	締め	学習	この経験を通じて、人に丁寧に接する大切さを学びました。
	締め	強調	以上より、人のよさが私の強みです。
学生	締め	将来	今後もこの強みを活かし、御社に貢献したいと思います。

$$\text{出現度} = \frac{\text{タグ } t \text{ が付与された文中の単語 } w \text{ の出現回数}}{\text{タグ } t \text{ が付与された文の数}} \quad (1)$$

各詳細タグにおける全単語の出現度を計算した後，予備実験により算出した閾値(=0.05)以上の単語を特徴語候補とした。特徴語候補の内，別のタグの単語の出現度に2倍以上差があるものを特徴語として採用した。表5に採用した特徴語一覧を示す。

表5: 特徴語一覧

タグ	詳細タグ	特徴語	特徴語の数
結論		長所, 短所, 時代, 最も, 入れる, ところ, です	7
具体例	状況	大学	1
	考え	感じる, 心がける	2
	行動	行う, 練習, そこで, 勉強	4
	結果	その, 結果, くれる, いただく	4
締め	学習	学ぶ, 身, つける, を通して, を通じて, 大切, 重要etc	15
	将来	今後, 社会, 仕事, 活かす, 生かす, 御社, 貴社, 貢献etc	16
	強調	自ら, やる	2

6. 評価実験

学習データとしてインターネット掲示板(面接データベース よくある質問と回答例(新卒・転職)[11], みんなの就職活動日記[12])から収集した志願者の回答 698 回答, 計 3980 文に作業者(男子大学生 1 人)がタグを付与したものを使用した。なお, この回答は「自己PRをして下さい」(239 回答), 「学生時代に最も力を入れたことは何ですか」(283 回答), 「あなたの長所と短所を教えてください」(176 回答)という3つの質問に対する回答である。付与したデータの内訳を表6に示す。

表 6：付与したデータの内訳

タグ	詳細タグ	データ数(文数)
結論		819
補足		167
具体例	状況	834
	考え	616
	行動	807
	結果	449
締め	学習	186
	将来	185
	強調	84

テストデータとしてインターネット掲示板(みんなの就職活動日記)から収集した回答 193 回答, 計 1269 文に作業者(男子大学生 1 人)がタグを付与したものを使用した。なお, この回答は「自己 PR をして下さい」(57 回答), 「学生時代に最も力を入れたことは何ですか」(106 回答), 「あなたの長所と短所を教えてください」(30 回答)という 3 つの質問に対する回答である。付与したデータの内訳を表 7 に示す。

表 7：付与したデータの内訳

タグ	詳細タグ	データ数(文数)
結論		210
補足		25
具体例	状況	288
	考え	252
	行動	257
	結果	131
締め	学習	69
	将来	39
	強調	23

本研究では“結論”, “具体例”, “締め”の 3 種類を分類することを“大分類”と呼び, “結論”, “状況”, “考え”, “行動”, “結果”, “学習”, “将来”, “強調”の 8 種類を分類することを“詳細分類”と呼ぶ。なお, 補足については数が少ないため除外した。

収集した回答を使用し, “大分類”, “詳細分類”が判別できるかどうか実験を行った。分類手法として SVM(Support vector machine)[13], ニューラルネット, CRF(Conditional Random Fields)[14]を比較した。SVM, ニューラルネットでは各特徴語の出現回数を素性値とし, CRF では各特徴語が出現するか否かを素性値としている。

6.1. 大分類の分類結果

構築した素性について各分類器により分類した結

果を表 8, 表 9, 表 10 にそれぞれ示す。人手で付与したタグを上部に, 左部にシステムが出力したタグを示す。例えば表 8 で人手は具体例, システムは結論の結果は 71 である。ここで精度と再現率の算出について表 8 を例にして示す。結論の精度は式(2)で求められる。また結論の再現率は式(3)で求められる。

$$\text{精度} = \frac{114}{114+71+6} \quad (2)$$

$$\text{再現率} = \frac{114}{114+91+5} \quad (3)$$

表 8：SVM を用いた分類結果

		正解			精度
		結論	具体例	締め	
出力	結論	114	71	6	0.60
	具体例	91	827	74	0.83
	締め	5	30	51	0.59
再現率		0.54	0.89	0.39	正解率
F 値		0.57	0.86	0.47	0.78

表 9：ニューラルネットを用いた分類結果

		正解			精度
		結論	具体例	締め	
出力	結論	126	83	11	0.57
	具体例	84	817	71	0.84
	締め	0	28	49	0.64
再現率		0.60	0.88	0.37	正解率
F 値		0.59	0.86	0.47	0.78

表 10：CRF を用いた分類結果

		正解			精度
		結論	具体例	締め	
出力	結論	187	2	2	0.98
	具体例	21	883	58	0.92
	締め	2	43	71	0.61
再現率		0.89	0.95	0.54	正解率
F 値		0.93	0.93	0.57	0.90

CRF では他の 2 つの分類器に比べ, “結論”の結果が高い。これは“結論”タグの出現する位置が最初にあることが多いことを考慮できた結果だと考えられる。全分類器において“具体例”が高い値を示している。これは学習データが多く十分に学習できた結果だと言える。またどの分類器でも“締め”の値が他のタグに比べて低い。これは“締め”が“具体例”に誤分類されたためである。

6.2. 詳細分類の分類結果

構築した素性で各分類器により詳細分類について分類した結果を表 11, 表 12, 表 13 にそれぞれ示す。

人手で付与したタグを上部に、左部にシステムが出力したタグを示す。

表 11 : SVM を用いた分類結果

		正解								精度
		結論	状況	考え	行動	結果	学習	将来	強調	
出力	結論	135	53	31	15	6	3	1	7	0.54
	状況	46	151	51	104	36	6	1	7	0.38
	考え	1	4	57	8	5	6	2	3	0.66
	行動	19	50	54	98	16	12	6	4	0.38
	結果	4	23	21	22	61	4	2	0	0.45
	学習	3	4	17	6	4	37	1	2	0.50
	将来	2	3	21	4	3	1	26	0	0.43
	強調	0	0	0	0	0	0	0	0	0.00
再現率		0.64	0.52	0.23	0.38	0.47	0.54	0.67	0.00	正解率
F 値		0.59	0.44	0.34	0.38	0.46	0.49	0.53	0.00	0.45

表 12 : ニューラルネットを用いた分類結果

		正解								精度
		結論	状況	考え	行動	結果	学習	将来	強調	
出力	結論	127	42	25	13	4	5	1	8	0.56
	状況	54	172	59	109	36	6	1	6	0.39
	考え	4	7	62	10	8	6	3	3	0.60
	行動	17	46	53	94	18	12	6	5	0.37
	結果	3	15	15	21	58	4	1	0	0.50
	学習	3	4	17	6	4	35	1	1	0.49
	将来	2	2	21	4	3	1	26	0	0.44
	強調	0	0	0	0	0	0	0	0	0.00
再現率		0.60	0.60	0.25	0.37	0.44	0.51	0.67	0.00	正解率
F 値		0.58	0.47	0.35	0.37	0.47	0.50	0.53	0.00	0.45

表 13 : CRF を用いた分類結果

		正解								精度
		結論	状況	考え	行動	結果	学習	将来	強調	
出力	結論	193	2	1	2	1	2	0	0	0.96
	状況	6	204	75	66	23	8	1	1	0.53
	考え	6	13	66	26	14	11	1	3	0.47
	行動	2	43	60	127	21	10	2	1	0.48
	結果	2	17	16	27	63	10	4	5	0.44
	学習	1	5	17	6	5	20	6	5	0.31
	将来	0	2	14	1	4	5	25	6	0.44
	強調	0	2	3	2	0	3	0	2	0.17
再現率		0.92	0.71	0.26	0.49	0.48	0.29	0.64	0.09	正解率
F 値		0.94	0.61	0.34	0.49	0.46	0.35	0.52	0.11	0.55

CRF では他の 2 つの分類器に比べ、“結論”の結果が高い。これは“結論”タグの出現する位置が最初にあることが多いことを詳細分類でも判定できた結果だと考えられる。

次に“状況”タグの結果を見てみる。“状況”では再現率は高い値を示しているが、精度は良くない。これはシステムが誤って本来“考え”や“行動”のタグを付与される文を誤分類したためだと言える。逆に“考え”では再現率が低い。これは 1 文が長くなった場合に“状況”や“行動”の特徴語が含まれているため誤分類したものだと考えられる。“行動”でも 1 文が長くなった場合に“状況”や“考え”の特徴語が含まれているため精度、再現率共に望ましい値であるとは言えない。

“結果”と“将来”では全分類器で値の差があまり見られない。この結果からこれらのタグは前後のタグに関係なく出現していると言える。“学習”では CRF で低い値を示した。これは“学習”タグの後に自身の考えを示す“考え”タグが出ている回答があったため、前後を考慮し誤ったものと考えられる。“強調”ではどの分類器でも低い値を示した。これは特徴語や前後のタグを考慮する素性だけでは分類が難しいためであると言える。

正解率を比べると特徴語の出現頻度のみで判定している SVM やニューラルネットよりも前後のタグを考慮した CRF の方が正解率が高い結果が得られた。

7. 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、不適切な回答を検出する手法を提案した。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。予備実験より計 51 語の特徴語を定義し、3 種類の分類器で比較実験を行った。大分類での実験結果、詳細分類での実験結果共に SVM やニューラルネットよりも前後のタグを考慮した CRF の方が高い結果を示した。

今後の課題としては、特定の単語に依存しない新たな素性の検討を行う必要がある。また本手法では bag-of-words 手法により各タグの特徴語を求めたが、情報利得等を用いて算出した特徴語で作成した素性と比較する必要がある。本研究ではタグの自動付与のみを行った。今後は自動的に付与されたタグの並びに着目し、不適切な回答を検出する必要がある。

参考文献

- [1] リアル模擬面接シミュレーター モギメン、マイナビ 2014, <http://job.mynavi.jp/conts/2014/mogimen/> (2013/10/15 アクセス)
- [2] 就活面接トレーナー, タブロイド, <http://www.tabroid.jp/app/business/2012/08/mensetsuTrainer.com.html>(2013/10/15 アクセス)
- [3] AttaliYigal, Burstein Jill. Automated essay scoring with e-rater v.2. Journal of Technology, Learning and assesment,(2007).
- [4] 石岡恒憲, 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).
- [5] 藤田彬, 田村直良, 文章構造解析に基づく小論文の論理性の自動採点, 情報科学技術フォーラム講演論文集, pp.41-44, 2010
- [6] 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.
- [7] 横山憲司, 難波英嗣, 奥村学, Support Vector Machine を用いた談話構造解析, 情報処理学会自然言語処理研究会, NL-155, pp.193-200, 2003.
- [8] 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.
- [9] 高田裕明, 全員内定!, 2012.
- [10] 伊東明, 人気企業を目指す人のためのハイレベル面接術一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.
- [11] 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)
- [12] みんなの就職活動日記, <http://www.nikki.ne.jp/>(2013/10/15 アクセス)
- [13] V.Vapnik, "The Nature of Statistical Learning Theory", Springer, (1995).
- [14] 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.