

AI 予測とテキストマイニング分析による短大生の就職活動支援方法の研究

片瀬拓弥

Study of job-hunting support method for junior college students by AI prediction and text mining analysis

Takuya Katase

要旨

本研究では、短大生に「AI 予測を活用した就職活動支援、就職活動の具体的支援方法、学校満足度調査」に関する調査を行い、AI 予測の活用や具体的支援方法について、単相関分析・テキストマイニング分析を実施して検討した。その結果、AI 予測の活用では、対象者の AI 予測に対する考え方が支援に影響する可能性があることを示した。さらに、学校満足度調査結果に対応した具体的な就職活動支援方法が明確になった。

キーワード：AI 予測、学校満足度調査 (QU)、テキストマイニング、就職活動支援

1. はじめに

近年、人工知能 (AI) や機械学習を活用した様々なイノベーションや予測などが取り上げられている。例えば、鹿島 (2015) は「いまや機械学習という言葉がビジネスの文脈でも当然のように使われ、その言葉を発する人の種類も技術者に留まらずさまざまである」といっている。つまり、研究者だけでなく、世間一般の人々までが「AI」や「機械学習」という用語自体に右往左往している感が否めない。そもそも、「AI」と「機械学習」の関係は、「AI」の一部手法として「機械学習」が存在するのであり、その専門用語の関係をしっかりと理解せずに使用することが、一層の混乱を招いているように思える。一方、今回の第三次 AI ブームの背景には、「ビッグデータを扱うためのネットワーク/クラウド環境などのデータ基盤と統計/パターン認識/機械学習などの高度なデータ解析技術の進展により、ビッグデータから知が生み出されたことにある」という (中林・和田 2016)。すなわち、ビッグデータ活用とパターン認識技術の融合が、AI ブームを実質的に後押ししていると考えられる。

さて、学校業界においても、この AI ブームは例外ではない。2017 年 5 月、経済産業省は「新産業構造ビジョン」を取りまとめ、「IoT、ビッグデータ、AI、ロボットなど、第 4 次産業革命によるさまざまな技術を社会実装し、社会が抱える様々な課題を解決しようという、我が国の政府が推進する『Society 5.0』を実現することを目指している」と述べている (堀田 2019)。さらに、AI ブームの火付け役として、社会的インパクトがかなり強かったものとして、英国・オックスフォード大学で機械学習の研究をしているマイケル A. オズボーンと経済学者のカール・ベネディクト・フレイが、野村総合研究所と共同で行った研究がある (野村総合研究所 2017)。この研究では、「我が国の約 600 種類の職業のうち、AI やロボットが代替して遂行できる確率を求めて公表し、実に日本の労働人口の 49% が、AI によって代替可能と推定」と報告した (堀田 2019)。つまり、現代に生きる学生達の将来的な職業は、ますます見えにくくなっていると考えられる。

このような状況の中、短大生の就職活動は、従来とあまり変わらない方法で実施されている。つまり、合同企業説明会、企業説明会、書類選考、筆記試験 (SPI など)、面接試験、役員面接、内々定、

内定という流れである。しかし、本格的な AI 時代が到来すれば、このような就職活動の流れ自体が変化する可能性は十分にある。現に、2019 年 8 月、就職情報サイトの運営会社が、就職活動中の学生が内定を辞退する確率を AI 予測し、38 社に販売した事件があった（時事通信社 2019）。この事件は、AI が就職活動に適用された事例であり、今後、学生がインターネット等を介して質問に回答するだけで、企業と学生とのマッチングを AI が判断するような時代になるかもしれない。ただし、現時点では、前述のような就職活動の流れが一般的であり、本研究でも、その就職活動の流れを前提に検討を行うものとする。

次に、就職活動と機械学習との関連についての研究を挙げる。片瀬（2019）は、短大 1 年 8 月までのデータから、短大 2 年時の就職活動終了時期を予測するモデルを開発した。この予測モデルでは、「教師なし機械学習」の一種であるクラスター分析を用いている。入力データとして、出身高校偏差値、日本語プレースメントテスト、短大 1 年 7 月時の学期 GPA、さらに、河村（2010）によって開発・標準化された学生生活に関するアンケート調査（学校生活満足度調査：以下、QU とする）などを用いていた。この研究では、短大 1 年時 QU において学校生活に満足していない（承認点が低く、被侵害点が高い）学生群の就職活動終了時期（短大 2 年時）が、最も遅いことを予測している。このことから、QU 結果の各所属群に対応した就職活動の具体的支援策が求められている。

そこで、本研究では、短大生に「AI 予測を活用した就職活動支援、就職活動の具体的な支援方法、QU」に関する調査を行い、これらの結果を比較分析することにより、適切な就職活動支援方法を検討することを目的とする。ただし、自由記述式の回答は、テキストマイニング分析を行う。

2. 調査方法

2. 1 調査対象

調査対象は、X 年度に本校国際コミュニケーション科に所属する短大生とした。QU については、X 年度 7 月時に実施した。さらに、就職活動に関する調査については、X 年度 10 月～11 月にかけて実施し、研究協力に対する同意を得てから調査した。

2. 2 学校生活満足度調査

学校生活満足度調査（QU）は、市販されている hyper-QU 大学版を活用する。hyper-QU 大学版は、河村（2010）によって開発・標準化された学生生活に関するアンケート調査のことである。河村（2010）が開発・標準化した学校生活満足度尺度は、2つの下位尺度（X 軸：被侵害点、Y 軸：承認点）から構成され、XY 座標上の位置から 4 群（満足群、非承認群、不安定群、不満足群）に分けられる。また、QU には、13 項目（学習、友人関係、将来、大学生活、家庭、健康、情緒、容姿、性格、異性、金銭、精神、人生）の「悩みに関するアンケート」が付随している。この悩みの合計、すなわち、悩み度合いも分析対象としたい。そこで、本研究では、学校生活満足度尺度の下位尺度である「承認点」と「被侵害点」を素点データのまま使い、13 項目の悩みに関しては、悩み 13 項目の合計値を分析に用いることにした。また、分析では、これら調査項目の表記を「QU 承認、QU 被侵害、QU 悩み」と略記する。

2. 3 「AI 予測を活用した就職活動支援」に関する調査

項目①：AI 予測を就職活動支援に活用する意志

項目①は、事前に入手可能な卒業生データを使って、短大 2 年時における就職活動終了時期を AI 予

測できるとしたら、その予測を活用したいと思う意思があるかどうかを尋ねた。具体的には「あなたの就職活動終了時期を AI 予測（1 年秋学期開始時）できるとしたら、あなたはそれを活用したいと思いますか」と質問し、選択肢を「5. そう思う, 4. どちらかというと思う, 3. どちらともえない, 2. どちらかというと思わない, 1. そう思わない」の 5 件法により回答させた。また、分析では、この項目の表記を「AI 活用」と略記する。

項目②：「AI 予測を就職活動支援に活用する意志」に対する理由（自由記述）

項目②は、項目①のいずれかの選択肢を回答した理由を自由記述により回答させた。この項目は、自由記述形式のため、テキストマイニング分析を行って検討する。

2. 4 「就職活動の具体的な支援方法」に関する調査

項目③：就職試験の受験回数（予想と実際）と内定月（予想と実際）に関する調査

項目③では、以下のように調査した。

1 年生（予想）

- ・「あなたが、最初の内定を頂くまでに、いくつ受験すると思いますか」

選択肢：1 つ、2 つ、・・・9 つ、10 以上、分からない（計算上、11 として扱う）

- ・「あなたが、最初の内定を頂くのは、何月だと思いますか」

選択肢：3 月、4 月、・・・、10 月、11 月以降、分からない（計算上、12 月として扱う）

2 年生（実際）

- ・「あなたが、最初の内定を頂くまでに、何社受験しましたか」

- ・「あなたが、最初の内定を頂いたのは、何月でしたか」

各選択肢は、1 年生と同様とした。また、分析において、これらの項目の表記を 1 年生については、「受験回（予想）、内定月（予想）」、2 年生については「受験回（実際）、内定月（実際）」と略記する。

項目④：就職活動に対する具体的支援方法に関する調査

項目④では、1 年生に対しては「あなたが就職活動・進学するにあたり、誰からどのような支援を受けたいと思いますか（いつ、どこで、だれに、なにを、どのように）」と質問した。また、2 年生に対しては「あなたが就職活動を行うにあたり、誰からどのような支援を受けましたか（いつ、どこで、だれに、なにを、どのように）」と質問した。両学年とも、回答は 150 字以上の自由記述形式とした。就職活動終了時期と関係がある QU 結果と比較しながら、テキストマイニング分析を行って検討する。

3. 分析結果

3. 1 分析対象者

分析対象者は、X 年度に本校国際コミュニケーション科に所属する短大生の中から、研究に同意した上で、全データが揃っている学生とした。これらの手続きにより、分析対象者は、1 年生：61 名、2 年生：43 名となった。さらに、教務学生部から提供された情報は、個人情報が入り込まないように匿名化処理を行った。

また、QU 結果の人数は、（満足群、非承認群、不安定群、不満足群）の表記として、1 年生（34 名、6 名、5 名、16 名）、2 年生（22 名、9 名、5 名、7 名）であった。分析対象者では、1 年生の満足群が 55.7%、2 年生の満足群が 51.2%となっており、満足度に統計的有意差は無かった。

3. 2 データの記述統計と単相関分析

第一に、数値分析が可能な各 QU 得点、項目①及び③の調査に関する記述統計量と単相関係数を表 1 及び表 2 に示す。

表 1 によれば、受験数（予想）／内定月（予想）、QU 被侵害／QU 悩みの間に有意な正の相関関係があり、QU 承認／QU 被侵害、QU 承認／QU 悩みの間に有意な負の相関関係があった。また、受験数（予想）／QU 被侵害、QU 被侵害／AI 活用の間に有意傾向な弱い正の相関関係があり、内定月（予想）／AI 活用の間に有意傾向な弱い負の相関関係があった。次に表 2 によれば、受験数（実際）／内定月（実際）、受験数（実際）／QU 被侵害、受験数（実際）／QU 悩み、内定月（実際）／QU 悩み、QU 被侵害／QU 悩みの間に有意な正の相関関係があり、QU 承認／QU 被侵害の間に有意な負の相関関係があった。また、QU 承認／QU 悩み、QU 悩み／AI 活用の間に有意傾向な弱い負の相関関係があった。

両学年の表 1 及び表 2 の結果から、受験数（予想・実際）／QU 被侵害の間に正の相関関係があったことは、片瀬（2019）が QU 結果と関連があるとした結果を裏づける結果となった。さらに、1 年生では、「内定月（予想）／AI 活用、QU 被侵害／AI 活用」、2 年生では「QU 悩み／AI 活用」に有意傾向な相関関係が見つかった。このことは、AI 予測を活用した就職活動支援には、対象者の AI 予測に対する考え方（積極的、消極的）が支援に影響する可能性があることを示した。つまり、AI 予測を活用した就職活動支援を誰もが望むわけではないことが判明した。

表 1 1 年生の記述統計と単相関分析 (n=61)

変数名	M	SD	単相関係数				
			受験数 (予想)	内定月 (予想)	QU 承認	QU 被侵害	QU 悩み
受験数（予想）	5.8 (3.5)		—				
内定月（予想）	9.1 (2.6)		.38 **	—			
QU承認	51.7 (8.7)		-.06	-.07	—		
QU被侵害	25.1 (8.1)		.22 †	.12	-.66 **	—	
QU悩み	20.1 (5.4)		-.03	.00	-.46 **	.53 **	—
AI活用	3.4 (0.9)		-.04	-.22 †	-.14	.23 †	.10

† $p < .1$, * $p < .05$, ** $p < .01$

表 2 2 年生の記述統計と単相関分析 (n=43)

変数名	M	SD	単相関係数				
			受験数 (予想)	内定月 (予想)	QU 承認	QU 被侵害	QU 悩み
受験数（実際）	3.8 (2.9)		—				
内定月（実際）	6.6 (2.0)		.66 **	—			
QU承認	53.0 (8.6)		-.13	-.09	—		
QU被侵害	23.7 (7.8)		.30 *	.18	-.44 **	—	
QU悩み	16.9 (4.5)		.48 **	.33 *	-.26 †	.47 **	—
AI活用	3.2 (1.4)		-.03	.06	.10	-.11	-.27 †

† $p < .1$, * $p < .05$, ** $p < .01$

3.3 「AI活用の理由」に対するテキストマイニング分析

前節の結果より、AI予測を活用した就職活動支援には、対象者のAI予測に対する考え方が支援に影響する可能性が存在する。そこで、「AI活用の理由」に対する自由記述をテキストマイニング分析し、さらに詳しい検討を行う。

テキストマイニング分析を行うためには、自由記述されたテキストデータを形態素に分解する必要がある。形態素とは、文章を各品詞に分解して最小単位にすることである。この形態素を取り出すことを形態素解析という。本研究では、形態素解析を行うテキストマイニングソフトとして、*KH Coder*を活用した。*KH Coder*は、樋口（2014）が開発したものである。本研究では、文字データを「名詞、サ変名詞、形容動詞、複合語、動詞」に分解した。複合語とは、「就職、活動」のように別々の名詞を「就職活動」のように複合させたものである。複合語の候補は、*KH Coder*を使った別分析を行い、出現頻度に応じてリストアップした。さらに、形態素解析が完了した文字データについて、抽出語の出現頻度を確認しながら、抽出語間の関連性を分析した。*KH Coder*では「抽出語間の関連性」を探索する一つの手法として、対応分析がある。図1にAI活用理由と各学年のAI活用意志（○：予測活用に積極的な者、×：予測活用に消極的な者）の関連性を示す。ただし、「○」は、AI活用について「4以上（平均以上）」と回答した者とした。それ以外は「×」に分類した。

図1では、各抽出語と各学年○×の位置関係が近いほど、文章に出現する時の関連性が高いことを示している。1年AI活用は、○と×で位置関係が離れている。「○」の近くにある抽出語は、「活用、改善、判断、目安、早め」などがある。一方、「×」の近くにある抽出語は、「変わる、未来、信頼、不安」などがある。これらのことから、1年AI活用「○」の学生は、就職活動終了時期についての

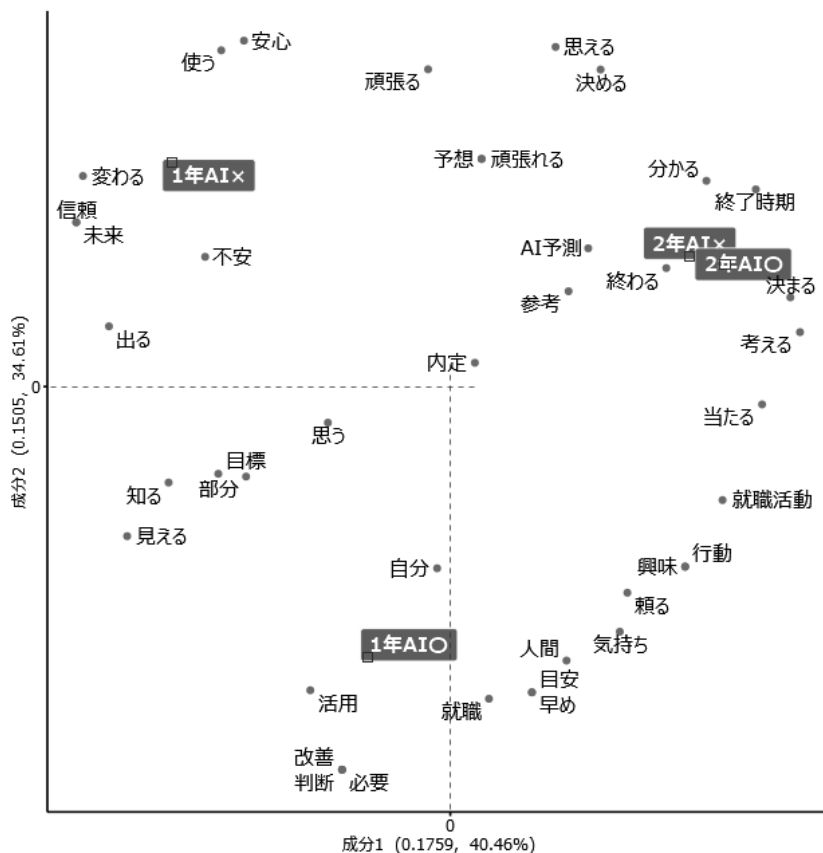


図1 AI活用の意志と理由の対応分析

「目安」や「早めに知ることによって改善や判断材料」に使っている可能性がある。他方、1年AI活用「×」の学生は、就職活動終了時期についての「予測の信頼性」や「予測が当たった時の不安」から、あまりAI活用に積極的ではないと考えられる。事実、以下のような理由が記述例としてあった。

「○」学生の記述例

- ・AI予測を知ることによって改善できる部分や良い部分があって改善していけるなら活用したいと思ったから。
- ・絶対ではないと思うけれど目安を知りたいし、自分の目標が達成できる時期を知っていたほうが自分自身頑張れるから。

「×」学生の記述例

- ・AIについてあまり詳しく知らないから、信頼性が欠けているから。
- ・参考にすることはできると思いますが、その結果によって不安や焦りが増えることになったら怖いと感じたから。

また、「×」学生の中には、「AI予測で結果が出てしまったら、それがたとえどんな結果であれ、自分は何があってもこうなってしまうし、なにも頑張らなくなってしまうから」と回答した学生もあり、逆効果になるケースも散見された。

一方、2年AI活用は、○と×で位置関係が離れておらず、結果の解釈が読み取り難い。よって、代表的な理由の記述例を示すことに留める。

「○」学生の記述例

- ・AI予測はほぼ当たると言われているので自分の終了時期がわかれば、終わりがあるならそこまで頑張ろうと思えると思ったから。
- ・AI予測ができるのであれば、それを活用して就職活動のスケジュールが立てやすいと思うから。

「×」学生の記述例

- ・終了時期に実際終わっていなかったら怖いし、その時期が近づくまで自分は何もしなくなるんじゃないかと思うから。
- ・就職活動は自分次第なのでAI予測はいらない。企業側も人間なので、AIなんかの予想を信じられない。

「AI予測を積極的に捉えようとする学生」と「AI予測があると逆効果になってしまう学生」が存在することが伺えた。このことは、本研究のAI活用に限らず、AI予測を活用すること自体に対し「共通的な別要因」があるのかもしれない。しかし、この新たな仮説は、本研究では検証できないため、これ以上の言及はしない。

3.4 「就職活動に対する具体的支援方法」に対するテキストマイニング分析

片瀬(2019)によれば、就職活動終了時期とQU結果には関連性がある。さらに、本研究の表1及び表2の結果からも、受験数(予想・実際)/QU被侵害の間に正の相関関係があった。したがって、「就職活動に対する具体的支援方法」の自由記述に対するテキストマイニング分析を行い、QU結果と抽出語の関連性を対応分析により検討する。

図2に1年「QU結果」と「就職活動に対する具体的支援方法の抽出語」との対応分析を示す。ただし、QU結果について、「満足群、非承認群、不安定群、不満足群」と略記した。また、KHcoderの表示機能にある「中心点付近を係数2で拡大表示」を行っている。表示に使った抽出語は、出現回数の平均値3.6回以上の「84語」であった。

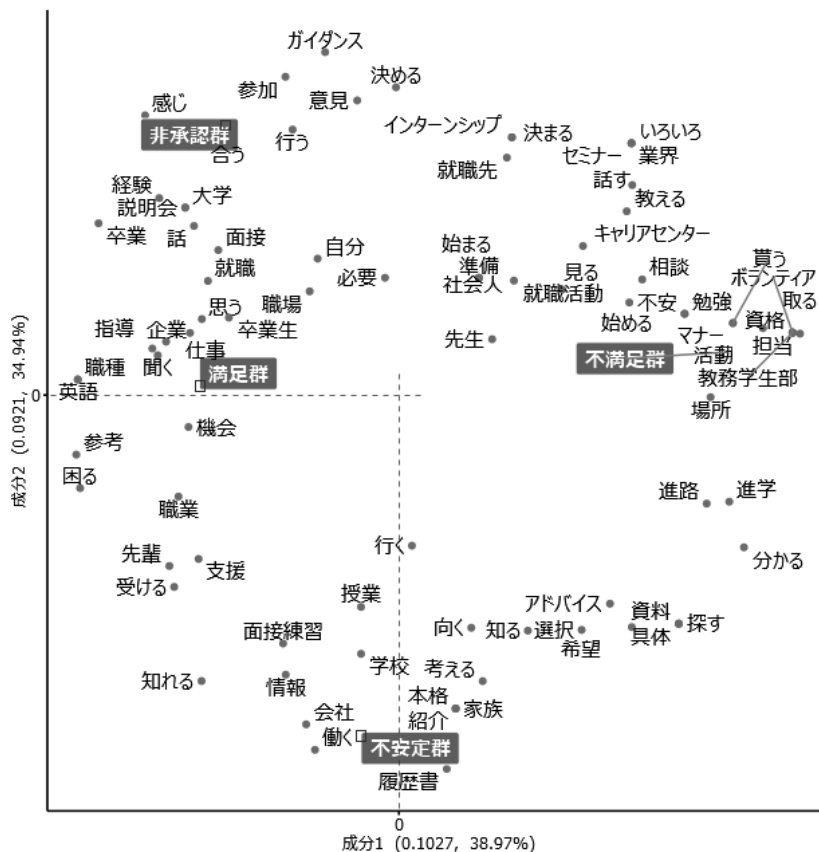


図 2 就職活動の具体的支援方法と QU 結果の対応分析（1 学年）

さて、項目④では「・・・誰からどのような支援を受けたいと思いますか（いつ、どこで、だれに、なにを、どのように）」と質問しているため、「5W1H に対応する抽出語」と「同時出現する動詞」の関連性に着目し、代表的抽出語として、以下を読み取った。

図 2 から読み取った代表的抽出語

満足群の近傍：「だれに：卒業生、先輩」、「どこで：職場、企業」、「なにを：仕事、職種」

非承認群の近傍：「どこで：大学、説明会、ガイダンス」、「なにを：経験、参加」

不安定群の近傍：「だれに：家族」、「どこで：学校、授業」、「なにを：履歴書、面接練習」

不満足群の近傍：「だれに：教務学生部、キャリアセンター」、「なにを：相談、勉強、資格、マナー」

また、「だれに」の観点では、「先生（学科教員）」が比較的中心に位置していた。

次に、図 3 に 2 年「QU 結果」と「就職活動に対する具体的支援方法」との対応分析を示す。図 3 も図 2 と同様の表示処理を行っている。そして、図 2 の結果と同様に代表的抽出語として、以下を読み取った。表示に使った抽出語は、出現回数の平均値 3.2 回以上の「63 語」であった。

図 3 から読み取った代表的抽出語

満足群の近傍：「だれに：センター職員」、「どこで：企業、企業説明会」、「なにを：相談、職種」

非承認群の近傍：「だれに：家族、友人、自分」、「なにを：履歴書、添削、就職活動」

不安定群の近傍：「どこで：合同説明会、授業」、「なにを：面接練習、不安」

不満足群の近傍：「だれに：所属ゼミ」、「どこで：説明会、キャリアデザイン」、「なにを：試験、心配」

また、「だれに」の観点では、図 2 と同様、「学科教員（先生）」が比較的中心に位置している。一方、

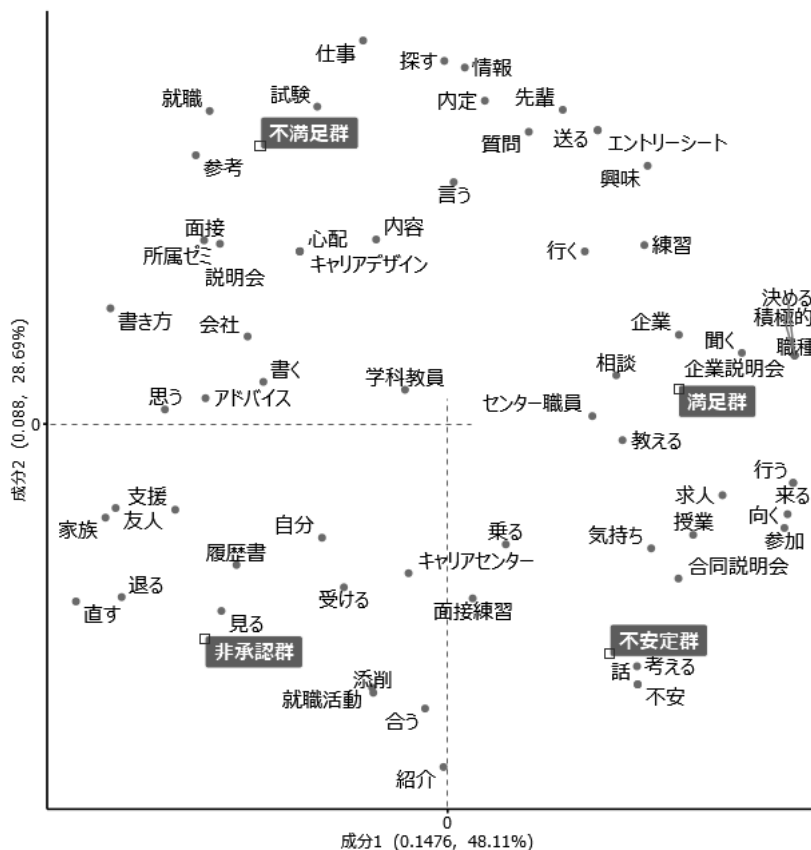


図3 就職活動の具体的支援方法とQU結果の対応分析（2学年）

「キャリアセンター」は、非承認群と不安定群の中間に位置している。

ここで、1年（図2）と2年（図3）の対応分析を統合し、QU結果と各図の代表的抽出語等を取捨選択しながら、具体的な就職活動支援方法を模索したい。そこで、代表的抽出語が「どのようにオリジナルのテキストデータで使用されているか」を詳細に確認し、以下のようなコーディングルールを作成した。コーディングルールとは、ほぼ同じ意味の抽出語をまとめて分析するための手法である。

コーディングルール

だれに：自分、先輩（卒業生）、家族（親戚）、友人（友達）、学科教員（先生、所属ゼミ）、センター職員、教務学生部

どこで：キャリアセンター、学校（大学、学校）、説明会（企業説明会、ガイダンス、合同説明会）

なにを／どのように：履歴書添削（履歴書&【添削 or 見る】）、面接練習（面接&練習、マナー、お辞儀）、相談（相談、アドバイス、面談）、学校教育（授業、キャリアデザイン、勉強、資格）、心配、不安、就職活動

図4に両学年の「代表的抽出語のコーディング」と「QU結果」の対応分析を示す。図4より、QU結果の各所属群に対応した具体的な就職活動支援方法がさらに明確になったと考えられる。つまり、図2から図4を総合的に検討した結果、「満足群：面接練習、非承認群：就職活動の心構えと適職指導、不安定群：不安低減と履歴書添削、不満足群：学科教員・教務学生部による相談」に重点を置いた支援が適切ではないかと推測した。ただし、1年生の場合、1年時におけるキャリアセンターとの接点は、まだ多くない。そのため、「教務学生部に相談」となっている部分は、2年時になると「キャリアセンターの職員」に引き継がれることになると考えられる。

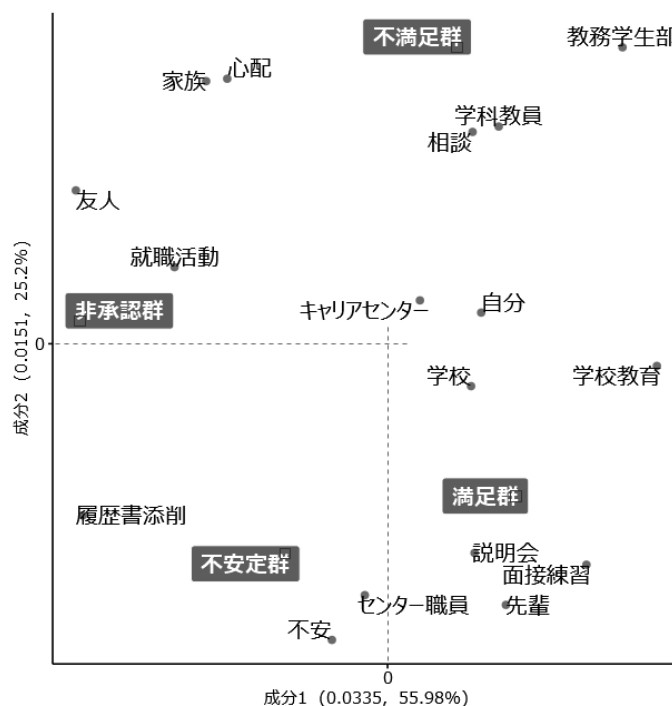


図4 代表的抽出語のコーディングとQU結果の対応分析（両学年統合）

4. まとめと今後の課題

本研究では、短大生に「AI 予測を活用した就職活動支援、就職活動の具体的支援方法、QU」に関して調査し、AI 予測の活用や具体的支援方法について、単相関分析・テキストマイニング分析を行って検討した。その結果、AI 予測の活用では、対象者の AI 予測に対する考え方が支援に影響する可能性があることを示した。さらに、QU 結果に対応した具体的な就職活動支援方法が明確になった。

本研究の結果を考察する。単相関分析により、QU 被侵害が高いと就職活動において受験数（実際）が多くなる可能性がある。QU 被侵害が高い群とは、すなわち、「不安定群」と「不満足群」である。さらに AI 活用意思と QU 被侵害の関係では、1 年（就活前）と 2 年（就活後）で違いがあった。そこで、この点を詳しく検討するため、AI 活用（○×）と QU 結果との関係性について割合を調べた。すると、1 年 AI 活用「○」の割合は、（満足群、非承認群、不安定群、不満足群）の表記として、（50.0%、66.7%、80.0%、75.0%）であった。一方、2 生は（59.1%、55.6%、40.0%、42.9%）となった。つまり、QU 被侵害が高い「不安定群」と「不満足群」では、AI 活用が就活前後で変化する可能性を示唆している。よって、これらの群が AI 予測を活用しやすい「就活前 1 年時」に使用して支援することが適切であると考えられる。さらに、その具体的な支援方法は、テキストマイニング分析により、「不安定群では、不安低減と履歴書添削」、「不満足群では、学科教員・教務学生部（キャリアセンター）による相談」に重点を置くべきだと考えられる。

今後の課題としては、本研究の結果は、あくまでも X 年度の短大生データを元に分析しているため、汎用性には欠ける。よって、他校種においてもデータを収集し、同様の結果が出るのか再検証する必要がある。また、AI 予測を就職活動支援に活用することに対して、「積極的に捉える学生」と「消極的に捉える学生」が存在した。このことは、本研究に限らず、AI 予測を活用すること自体に「共通的な別要因」があるのかもしれない。この仮説に対しては、今後の研究が望まれる。

謝辞

本研究は JSPS 科研費 JP18K02882 の助成を受けた。また、データ収集・提供に際しては、本校の学生や教務学生部職員の協力を受けた。

参考文献

- 鹿島久嗣 (2015) 『機械学習が起こすイノベーション < 特集 > 編集委員会企画-社会と AI の羅針盤 2015』, 人工知能, 30(1), pp.23.
- 中林暁男, 和田英彦 (2016) 『第 3 次人工知能ブームを機に見る製造業におけるデータ解析技術の変遷』, 横河技報, 59(1), pp.3-6.
- 堀田龍也 (2019) 『AI 時代の学校教育の在るべき姿とは』, 学術の動向, 24(2), pp.56-65.
- 野村総合研究所 (2017) 『AI 時代に求められる豊かな個性と人材のダイバーシティ』, <https://www.nri.com/jp/journal/2017/0503> (Accessed 2020.01.07)
- 時事通信社 (2019) 『リクナビ、内定辞退を AI で予測・販売=就活生への説明不足』 <https://www.jiji.com/jc/article?k=2019080200139> (Accessed 2020.01.07)
- 河村茂雄 (2010) 『hyper-QU(大学版)』, 図書文化社, 東京
- 片瀬拓弥 (2019) 『短大生の就職活動終了時期の予測モデルの開発』, 日本教育工学会研究会報告集 高等教育における FD・SD・IR・学修支援/一般 (JSET19-3), pp.27-34.
- 樋口耕一 (2014) 『社会調査のための計量テキスト分析』, ナカニシヤ出版

SUMMARY

In this study, I conducted a survey on junior college students about "job-hunting support using AI prediction, concrete support methods for job-hunting, and school satisfaction surveys", and conducted simple correlation analysis and text mining analysis was performed. As a result, it was shown that the subject's attitude toward AI prediction may affect the support of using AI prediction. In addition, the specific job-hunting support methods corresponding to the school satisfaction survey results were clarified.

Keywords: AI prediction, school satisfaction survey (QU), text mining job hunting support