

就職活動支援システムにおける企業情報推薦機能の開発 - 企業の採用項目と就活生のアピールのギャップに注目して -

何 陽^{†1} 長谷川 忍^{†2}

概要：本研究では、就職活動中の学生が関連した企業の情報を発見することを支援することを目的として、学生の個人情報や検索履歴、協調フィルタリング、内容ベースフィルタリング等の手法を利用した関連企業の推薦機能の開発を目指す。特に、企業の採用項目と就活生のアピールポイントのギャップを考慮した上で、それぞれの学生の企業選択項目に対する重要度を反映できる推薦機能を実現する。

キーワード：就職活動，企業推薦機能，個人差

Development of Company Information Recommendation Function in Job Hunting Activity Support System - Focusing on the gap between companies' viewpoints and students' appealing points in job hunting process-

YANG HE^{†1} SHINOBU HASEGAWA^{†2}

Abstract: This research aims to develop a recommendation function to help students search for related/unknown companies based on content-based/collaborative filtering methods with student personal information. The originality of this research is to deal with the gap between the companies' viewpoints and students' appealing points in a job hunting process. The research also focuses on individual difference about criteria/importance in their job-hunting activities. In other words, the proposed function consists of two recommendation algorithms which can reflect on each student interests.

Keywords: job hunting activity, recommendation function, individual differences

1. はじめに

インターネットに代表される情報化の進行により、我々が触れることができる情報量は爆発的に増加している。一方、こうした膨大な情報から有用な情報を検索・抽出することは非常に難しい。特に就職活動中の学生（以後、就活生）はこの問題に直面する。例えば、中国においては、大卒就活生の数がここ数年で10倍以上に増加したこともあり、競争が一層激しくなっている。ただし、就活生の関心は主に国営大手企業に限られており、民間企業の情報収集やエントリーには消極的である[1]。一方、日本における就職活動の現状を見ると、就活生は就職活動に非常に多くの時間と労力を掛けていることが大きな課題となっている[2]。

こうした問題を解決するため、近年、就職活動支援に関する研究が盛んに行われており、就活生の指導や支援を目的とした就職支援システムが提案されている[3][4][5]。特に、就活生自身が「将来何をしたいか」、「社会で何ができるのか」、「どのような人になりたいか」などといった自己分析は重要なタスクである。加えて、対象企業の数が多すぎて適切な企業情報の検索が困難であるという課題も存在する。本研究では、個々の就活生が重視する観点に基づいた企業推薦の可能性について検討する。

筆者らの所属大学においても、大学に対する企業からの求人情報や推薦情報を検索する就職活動支援システムが運

用されてきたが、2015年度からより多様な機能を備えたJAIST Career Support Systemが運用されている[6]。以下の図1にJAIST Career Support Systemの機能図を示す。

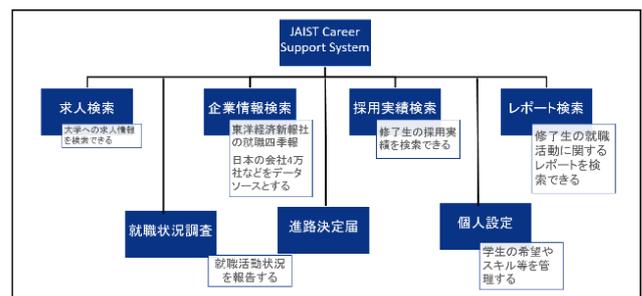


図1 JAIST Career Support Systemの機能

JAIST Career Support Systemは一般的な企業情報の検索機能は備えているが、競合する企業を検索したり、企業名やキーワードが思い浮かばない企業を検索したり、学生の検索プロセスに基づいて関連企業を推薦する機能は実装されておらず、企業検索の有用性が十分に発揮できているとは言い難いのが現状である。また、多くの既存の就職支援システムも同様に企業情報の検索はキーワード検索に基づいている。しかしながら、単にキーワードに基づいて企業情

^{†1} YANG HE, Japan Advanced Institute of Science and Technology
^{†2} SHINOBU HASEGAWA, Japan Advanced Institute of Science and

Technology.

報を照会するという従来の機能はますます学生のニーズを満たすことができなくなっていると考えられる。

これらの課題を解決するために、本研究では関連した企業情報を提示するために学生の個人情報や検索履歴、内容ベースフィルタリング等の手法を利用した関連企業の推薦機能の開発を目指す。特に、企業の採用時に重視する項目と就活生のアピールポイントのギャップを考慮した上で、企業選択時の観点に対する学生の重要度を反映できるアルゴリズムを検討し、企業推薦機能の開発や類似度計算アルゴリズムの改良及び有効性の評価を行う。

具体的には、推薦アルゴリズムを就職支援システムの企業推薦機能として適用することを試み、学生間の類似度計算と企業間の類似度計算の結果の精度を比較することにより、企業推薦機能における2つの推薦方法の妥当性と適用可能性を検討した。加えて、従来の推薦アルゴリズムと、提案する個人差を反映した重み付け特徴ベクトルに基づく類似度計算手法の推薦精度の比較を行った。単に業種・職種・企業規模などの項目に基づいて企業間と就活生間それぞれの類似度を計算する推薦システムと、就活生の特性の個人差及び、就職活動における企業と就活生の双方の観点が反映された企業推薦の有効性を比較することにより、キーワード検索だけでは十分に発見できなかった企業が発見できるかどうかを評価する。

2. 就職活動の現状

まず、就職活動で一般に行われている具体的な活動を図2に示す[7]。日本の大学院生と大学生のスケジュールは一般にほぼ同様である。

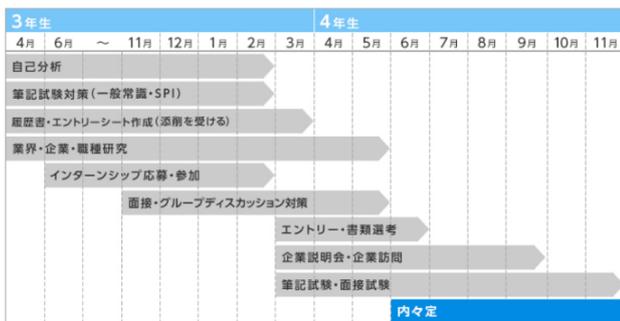


図2 就職活動の流れ図[7]

就職活動においては、「世の中にはどんな仕事や企業があるのか？」などといった業界・企業・職種研究のみならず、「自分はどんな仕事がしたいのか？」などの問題を考慮し、自己分析を行う必要がある。自己の特徴や適性を明確に把握することは容易ではなく、希望する業種や職種を決められずに多くの企業にエントリーする例も少なくない。

経団連の「採用選考に関する指針」によれば、「広報活動：卒業・修了年度に入る直前の3月1日以降」、「選考活動：卒業・修了年度の6月1日以降」となっている[9]。しかし、実際には広報活動はほぼ3月に集中し、選考活動は6月以前から始まっている[8]。

さらに、図3の「マイナビ2017年卒の就活生のエントリー数調査」によれば、2017年卒業の就活生の3月時点のエントリー数の平均は30.6社となっている[2]。2016年度以降、採用選考の指針の影響により、就活生のエントリー時期も3月に集中した。なお、2016年卒の3月時点のエントリー数

の平均が35社であったことに対して、2017年卒の同時期のエントリー数が平均30.6社まで減少している原因は、就職活動が売り手市場の傾向となり、就活生に有利となったことが考えられる。いずれにしても、就活生は就職活動に対して非常に多くの時間と労力をかけていると言える。

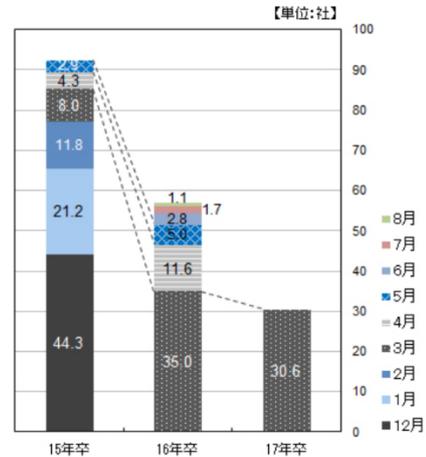


図3 マイナビ2017年卒の就活生のエントリー数調査[2]

本研究では、企業と就活生の双方の観点の違いを検討するために、図4で示す企業が採用時に重視する項目と学生が面接等でアピールしている項目についての調査結果[8]に注目する。

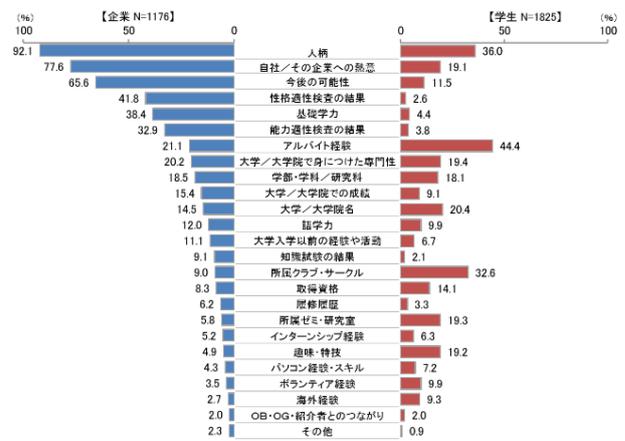


図4 企業が採用時に重視する項目と学生がアピールする項目の調査結果[8]

この調査結果によれば、就職活動において企業が重視する項目と就活生がアピールしたい項目の間には大きなズレがあることがわかる。つまり、就活生はこのギャップを正確に把握できていないことを示している。例えば、企業は就活生の「将来の可能性」、「性格」、及び「基礎学力」などの項目を非常に重視していることと比較して、就活生は自分の「アルバイト経験」や「所属クラブ・サークル」などの項目を重視する傾向にある。このように、企業と就活生の双方の就職活動に関する観点のギャップは大きなものとなっている。そこで本研究では、学生間の類似度計算と企業間の類似度計算の双方で、各観点に対する重要度を総合的に考慮することで、推薦精度の向上が期待できるのではないかと仮説を立てた。

3. 内容ベースフィルタリングと協調フィルタリング

一般的な推薦システムに利用されている内容ベースフィルタリングと協調フィルタリングの長所と短所[10]を図5に示す。

分類	協調フィルタリング	内容ベース
多様性	○	×
ドメイン知識	○	×
スタートアップ問題	×	△
利用者数	×	○
被覆率	×	○
類似アイテム	×	○
少数はの利用者	×	○

図5 内容ベースフィルタリングと協調フィルタリングの長所・短所分析図[10]

協調フィルタリングとは、ユーザーの行動履歴を元に推薦を行う手法であり、Amazonの商品推薦機能などに代表される。内容ベースフィルタリングとは、企業の特徴ベクトルの類似度計算により推薦する方法である。企業推薦機能を検討する上で、協調フィルタリングで最も注目される特性は多様性である。つまり、協調フィルタリングを利用することで、推薦が必要な就活生と最も類似した就活生の検索履歴から、関連する企業を推薦することが実現できる。一般に、協調フィルタリングは多様性を有するため、企業名がわからない企業や異業種の企業の推薦確率が高くなると推測される。一方で、内容ベースフィルタリングは「少数の利用者」や「被覆率が高い」という特徴により、本学のような小規模な機関において推薦機能を実現する上で、重要な要素となる。これらのことから本研究では、協調フィルタリングと内容ベースフィルタリングの双方を組み合わせた推薦機能の実現を目指す。

4. 提案する推薦機能の原理

4.1 類似度計算：コサイン類似度 (Cosine Similarity)

本研究で提案した企業推薦方法の理論的根拠は、主に内容ベースフィルタリングに基づいており、就活生と企業の特徴を表現するために、就活生の個人情報や就職活動に対する意向などの情報と企業の業種や規模、募集要項などの情報からそれぞれに各項目を抽出し、相互の類似度を計算する。本研究では、関連する企業情報を推薦するために企業情報および就活生の特徴ベクトルを構成し、その類似度を計算することで推薦に必要な情報を得るアプローチを採る。企業情報に関する特徴ベクトルは企業規模や募集要項など、企業を特徴づけるデータの集合として構成される。就活生の特徴ベクトルは、個人スキルや希望の職種及び、希望の勤務地などの就職希望に代表される就職活動に関する考え方を情報源とする。

特徴ベクトル間の類似度の計算に当たって、本研究ではコサイン類似度[11]を用いる。以下に、二次元ベクトルにおけるベクトルaとベクトルbのコサイン類似度の概念図を図6に示す。

コサイン類似度では、2つのベクトル間の角度の余弦によりそれらのベクトル間の類似度を定義する。角度のコサイン値は[-1~1]の間に分布している。したがって、2つ

のベクトル間の角度の余弦値によって、2つのベクトルの方向の類似度を示すことができる。すなわち、2つのベクトルが同じ方向を指している場合には、コサイン類似度の値は1であり、2つのベクトル間の角度が90°の場合には、コサイン類似度の値が0となる。さらに、2つのベクトルが反対の方向を指している場合には、コサイン類似度の値が-1となる。なお、コサイン類似度は通常正の空間のみで使用されるため、結果の値は[0~1]の間となる。この計算方法は任意の次元のベクトル空間にも適用することが可能であるため、本研究における企業特徴ベクトル間、及び就活生特徴ベクトル間の類似度をコサイン類似度によって決定することができる。

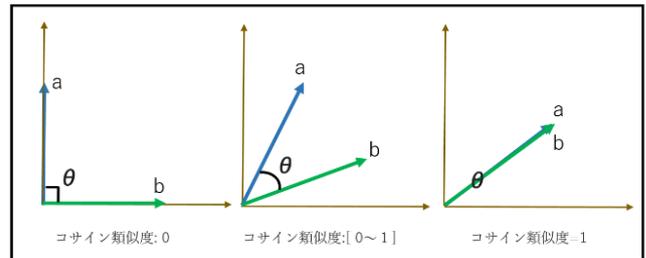


図6 二次元座標におけるコサイン類似度

企業間の特徴ベクトルの類似度計算により、就活生が興味を持っている企業と類似度が高い企業を計算することができる。これにより、システムは企業データベースの中から就活生が興味を持つ企業に類似した他の企業を推薦することが可能となる。二次元座標における具体的な企業特徴ベクトルの例を図7に示す。

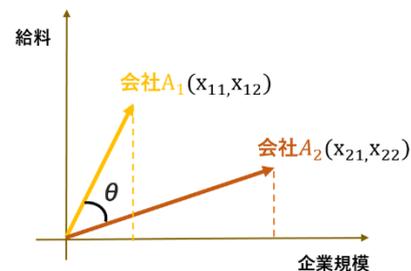


図7 二次元座標における企業特徴ベクトル

図7は、「給料」と「企業規模」の項目に基づき、企業A1と企業A2の類似度を例として表示したものである。x11とx12はそれぞれ企業A1の「給料」と「企業規模」の項目における属性値を表す。属性値とは、特徴ベクトル間の類似度計算を行う際に、実際の値（給料:24万円、企業規模:3000人など）をベクトル化したパラメータである。

実際の特徴ベクトルは(式1)で定義する。ここで、x_{ji}は各項目の属性値の数を表す。ある2つの特徴ベクトル類似度は、N次元空間におけるベクトルのコサイン類似度計算式(式2)を用いて計算される。例えば、図7のように、企業情報が「企業規模」と「給料」から構成されると仮定した場合、企業A1と企業A2のコサイン類似度は式(3.2)によって計算でき、企業A1と企業A2のコサイン類似度の結果は
$$\frac{x_{11}x_{21}+x_{12}x_{22}}{\sqrt{x_{11}^2+x_{12}^2}\sqrt{x_{21}^2+x_{22}^2}}$$
である。

$$A_j = \{x_{j1}, x_{j2}, \dots, x_{jn}\} \dots \dots \dots (式 1)$$

$$\text{Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \cdot \sqrt{\sum_{i=1}^n (B_i)^2}} \dots (式 2)$$

4.2 就活生間類似度計算による企業推薦

企業間の類似度計算と同様に、就活生の特徴ベクトル間の類似度計算を行う。就活生の各評価項目(性格, アルバイト経験, 勤務地の希望など)に基づいて構成された特徴ベクトルを利用して計算したコサイン類似度の結果により、システムを使用中の就活生の中で自分の状況と類似した就活生を抽出することができる。こうした就活生の検索履歴の中には、自身が発見しておらず、かつ自身の興味関心に近い企業が含まれていることが期待でき、効果的な企業推薦が行える可能性がある。

ここまでで述べた、企業間類似度計算と就活生間類似度計算の 2 つの方法を組み合わせることにより、名前を知らない企業や目標とする企業の競合企業、さらに異業種の企業を推薦することが期待できる。

4.3 個人差の重み付け類似度計算手法

前節で述べた類似度計算手法により、企業推薦を行うことは可能である。しかしながら、就活生が就職活動において重視する観点は一般に異なっている。例えば、勤務地を重視する就活生もいれば、仕事内容を重視する就活生も存在する。類似度の計算にあたって、こうした就活生の特徴企業に対する観点の重要度を反映させない場合、推薦結果に対する就活生の満足度が向上しないことが予想される。

このような就活生の観点の重要度の推定に当たっては、就活生の検索履歴を利用することが考えられる。しかしながら、就職活動の期間は限定的であり、また一般的な検索行動と比較すると検索回数も少ないため、システム利用の初期の段階で就活生の適切なモデルを構築することが難しいスタートアップ問題が発生する。さらに、就職活動において企業が就活生に期待する項目と就活生がアピールしようとする項目はしばしば異なっていることが指摘されている[2]。そこで本研究では、就活生と企業のズレを避けるのみならず、推薦機能の精度の向上を目的として個人差を考慮した重み付けコサイン類似度計算を利用した企業推薦手法を提案する。

企業間の類似度を計算する特徴ベクトルの構成要素として、前節で述べた就職活動における企業情報において企業と就活生の双方が重視する項目と本学の就職支援システムのデータベースの情報を総合的に考慮した結果、表 1 に示す項目に対して、重要度を重みとする企業の特徴ベクトルとして利用することとした。

表 1 企業間の類似度計算用の項目及び重み

	図 3.4 に対応する項目番号	重要度の重み	パーセンテージ標準換算
重要度上位の項目			
勤務地	2	49.8	入力
職種	3	48.0	入力
初任給	5	47.0	入力
有給年休	6	43.4	入力
業種	7	40.8	入力
月残業時間	11	36.2	入力
離職率	10	$x_1=36.2$	$y_1=12.25$
平均勤続年数	13	$x_2=32.3$	$y_2=10.93$
二年間採用実績	18	$x_3=29.0$	$y_3=9.81$
上場状況	19	$x_4=28.7$	$y_4=9.71$
年商	19	$x_5=28.7$	$y_5=9.71$
平均年齢	20	$x_6=28.3$	$y_6=9.57$
平均年収	24	$x_7=23.7$	$y_7=8.02$
企業規模	28	$x_8=22.1$	$y_8=7.48$
男性役員数	28	$x_9=22.1$	$y_9=7.48$
女性役員数	28	$x_{10}=22.1$	$y_{10}=7.48$
女性役員割合	29	$x_{11}=22.0$	$y_{11}=7.45$

情報源から抽出された部分の項目に対しては、パーセンテージの再調整が必要である。重要度下位の項目については、全体の重みが 100 となるように以下の通り重みの正規化を行った。

$$\text{Sum} = \sum_{i=1}^{11} x_i = 295.5 \quad \text{Std} = \frac{\text{Sum}}{100} = 2.955 \quad y_i = \frac{x_i}{\text{Std}}$$

就職活動における企業の選択時に、就活生の希望を反映した企業を推薦するために本研究では、企業の特徴ベクトル間の類似度の計算時に就活生の個人差を考慮する。つまり、就活生が企業に対して重視する観点を、就活生の特徴を把握するための重要な項目として取り扱う。こうした観点は、企業の特徴ベクトル間の類似度を計算する上でも大きな影響を与える。そこで本研究では、企業間の類似度計算において、特に就活生が強く意識するとされる「勤務地」、「職種」、「初任給」、「有給年休」、「業種」、「月残業時間」の項目について、各就活生がどの程度重視するかをパラメータとして入力させることにより、就活生の視点に基づく企業推薦を行えるようにした。

また、就活生に最も重視される上位の 6 個の観点は、企業間類似度計算のみならず、就活生の特徴ベクトル間の類似度計算にも利用することが可能である。就活生間の類似度を評価する場合には、一般に就活生が把握することが難しいと考えられる、企業の視点で就活生の採用時に重視している観点を抽出し、就活生が企業を選択する際に重視する項目を加えることによって、就活生の特徴ベクトル間の類似計算に用いられる重みを設計した。前章の図 2 に示した「企業が採用項目で重視する項目」に基づき、企業側の採用項目と就活生のアピールポイントのズレを避けることで、両方の視点を総合的に考慮した本提案手法の就活生間類似計算用の各項目及び重みを以下の表 2 に示す。

表 2 就活生間の類似度計算用の項目及び重み

	図 1.3 に対応する項目番号	重要度の重み	パーセンテージ標準換算
性格		$x_1=41.8$	$y_1=19.11$
学歴		$x_2=38.4$	$y_1=17.61$
SPI 成績		$x_3=32.9$	$y_1=15.04$
アルバイト		$x_4=21.1$	$y_1=9.65$
研究の専門領域		$x_5=20.2$	$y_1=9.24$
研究科		$x_6=18.5$	$y_1=8.46$
講義の平均成績		$x_7=15.4$	$y_1=7.04$
TOEIC 成績		$x_8=12.0$	$y_1=5.49$
資格		$x_9=8.3$	$y_1=3.80$
インター経験		$x_{10}=5.2$	$y_1=2.38$
趣味		$x_{11}=4.9$	$y_1=2.24$

5 システムの設計と開発

5.1 企業推薦機能のアーキテクチャ

本研究で提案する企業推薦機能のアーキテクチャを図 8 に示す。本システムは主にユーザーモジュール、推薦オブジェクトモジュール、類似度計算モジュールから構成され、ユーザーモジュールは就活生の個人情報と就職希望に関するデータをベクトル化し、就活生の特徴ベクトルおよび検索された企業の特徴ベクトルを生成する。次に推薦オブジェクトモジュールから生成された比較対象となる特徴ベクトルとの類似度を類似度計算モジュールで計算する。つまり、類似度計算アルゴリズムを用いてスクリーニングを行い、就活生の興味に近いと判断された関連企業を推薦対象として推薦する。

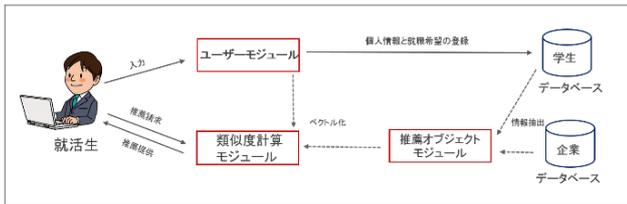


図8 企業推薦機能のアーキテクチャ

5.2 システムの開発環境

前節に示した就職支援システムにおける企業推薦機能を実現するにあたって、筆者らは以下の表3で示す環境を用いて開発した。

表3 システムの構築環境

システムの構築環境。	
OS。	Windows 10 Enterprise 64bit。
開発言語。	Java。
開発環境。	JDK 開発環境。
データベース。	MySQL。
開発ツール。	Eclipse(4.9)。

5.3 データベースの設計

企業推薦機能を実現するために、前章で説明した企業の特徴ベクトルと就活生の特徴ベクトルの生成に用いられるデータベーステーブルを設計した。企業特徴ベクトルに関しては、既存のJAIST Career Support Systemに使用されている企業データベースのカラムから項目を抽出した。一方、就活生の特徴ベクトルについては既存システムには蓄積されていないため、前章の議論に基づいて図9に示すようなテーブルを設計した。

studentID	ips	english	pr	research	area	profession	internship	requirement	entrycom	academic	business	parttime	field	result	interest	qualification	pay	holiday	overtime
SPI成績	TOEIC成績	性格	研究科	勤務地の希望	職種の希望	インター経験	入力した重視度	興味深い四季報コード	学歴	業種	アルバイト経験	研究領域	講義の成績	趣味	各種類の資格	初任給の希望	年休の希望	残業時間の希望	
int	decimal	decimal	varchar	decimal	varchar	varchar	varchar	varchar	decimal	varchar	varchar	decimal	decimal	varchar	varchar	decimal	decimal	decimal	
10	10	10	255	10	255	255	255	255	10	255	255	10	10	255	255	10	10	10	
0	1	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	
<input checked="" type="checkbox"/>	<input type="checkbox"/>																		

図9 就活生情報テーブルの構造

企業間類似度計算と就活生間類似度計算に用いられる各項目の具体的な選択肢については図10に示す。

5.4 データのベクトル化手法

各項目の値については、純粋な数字型データ（英語成績、講義平均成績、企業の役員人数などの項目のデータ）、単一選択肢型データ（学歴、研究科、研究領域などの項目のデータ）、および複数選択肢型データ（業種の希望、勤務地の規模、趣味などの項目のデータ）の三種類の属性値の表現方

法がある。特徴ベクトルのデータに変換する際には、それらの表現方法に応じて、それぞれ異なる処理方法が用いられる。

職種の希望	
id	name
1	営業
2	管理事務・業務
3	企画開発・調査
4	SE・プログラマ
5	研究
6	開発
7	生産管理
8	品質管理
9	生産技術
10	設計
11	教師・インストラクター
12	その他

業種の希望		
id	name	
1	農林水産業	21 陸運業
2	鉱業	22 海運業
3	建設業	23 空運業
4	食料品	24 倉庫・運輸関連
5	繊維・衣服	25 情報・通信業
6	パルプ・紙	26 卸売業
7	化学	27 小売業
8	医薬品	28 銀行業
9	石油石炭	29 証券・商品先物取引業
10	ゴム製品	30 保険業
11	ガラス・土石	31 その他金融業
12	鉄鋼	32 不動産業
13	非鉄金属	33 サービス業
14	金属製品	34 医療・福祉
15	機械	35 教育・学習支援
16	電気機器	36 学術・開発研究
17	輸送機器	37 国家公務員
18	精密機器	38 地方公務員
19	他製造業	39 外国企業
20	電気・ガス・水道	40 その他

月残業時間の希望	
id	name
1	20時間以下
2	20-40時間
3	40-60時間
4	60時間以上
10	ほか

趣味	
id	name
0	なし
1	販売
2	フード・飲食
3	サービス
4	イベント
5	軽作業・物流
6	事務
7	教育
8	モニター・調査
9	夜の仕事
10	ほか

インターンの経験	
id	name
0	なし
1	実践型長期インターンシップ
2	プロジェクト型インターンシップ
3	実体験型インターンシップ
4	職場見学型インターンシップ
5	ほか

図10 一部の各項目の選択肢の設計図

ベクトル間の類似度を計算する際には、まず比較している2つのベクトルの実際の各項目のパラメータを決定する必要がある。数字型データの項目のパラメータをベクトル化する場合には、推薦をしてほしい就活生のベクトルのパラメータは常に「1」（標準ベクトルと呼ばれる）に設定され、比較対象の就活生のベクトルのパラメータを比較対象の就活生の実際値と推薦をしてほしい就活生の実際値の「偏差の程度」に設定する。実際値の「偏差の程度」とは、推薦をしてほしい就活生と比較対象の就活生のデータの実際の値の比較に基づいて、小さい方が大きい方に占める割合の値を示したものである。これにより、比較対象ベクトルのパラメータの値は「0~1」の区間に分布することとなる。「偏差の程度」が「1」なら双方の項目の実際の値は同じであり、「0」なら双方の項目の実際の値の差が最も大きいことを表す。最後に、項目毎重要度の重みを掛けて計算する。

単一選択肢型データの項目のパラメータをベクトル化する場合には、まず、比較している2つのデータの類似度を表示するため、選択肢の番号の値を比較する。次に、比較結果が同じなら、標準ベクトルと比較対象ベクトル両方の項目のパラメータを「1」に設定し、比較結果が異なる場合は、前項の標準ベクトルのパラメータを「1」に設定すると同時に、後項の比較対象ベクトルのパラメータを「0」に変換する。最後に、この項目のパラメータに該当する項目の重要度の重みを掛けて計算する。

複数選択肢型データを項目のパラメータをベクトル化する場合には、まず、比較している2つのデータの類似度を表示するため、実際に選択された複数の選択肢の番号の値を比較する。ここでは、2つの集合の類似度計算手法であるジャカード指数を用いる。以下にジャカード指数計算数式を式3に示す。

$$Jacc = \frac{|A \cap B|}{|A \cup B|} \dots \dots \dots (式3)$$

重複番号のない2つの列のデータAとBの類似度を検索する場合には、Aに含まれた項目IDの個数とBに含まれた項目IDの個数の比較を行い、両方共に含まれた同じ項目IDの個数で両方に含まれた異なる項目IDの個数を割っ

た結果を 2 つの列のデータの類似度と定義する。ジャッカード指数の値は「0~1」区間に分布し、値が大きいほど、A と B の類似度は高くなっている。

ここでは、就活生 1 と就活生 3 の間の類似度計算の例を図 11 に示す。

学生1の基本情報は:	学生1との比較対象学生3の基本情報は:
SPI成績 52.5	SPI成績 55.0
英語成績 530.0	英語成績 420.0
PR点 2,4,8	PR点 5,7
研究科 情報科学専攻	研究科 情報科学専攻
企業地域 1,13,14	企業地域 11,12,13,14,27,40
希望職種 3,4	希望職種 4,5,6
インターンシップ経験 2	インターンシップ経験 0
個人ニーズ 30,20,10,10,20,10	個人ニーズ 20,30,5,15,15,15
エントリーしたい会社 67,498,1225,482,1082	エントリーしたい会社 1439,1437,1106,1431,497
学歴 博士前期課程	学歴 博士前期課程
アルバイト経験 1	アルバイト経験 2,3
専門領域 ゲーム・エンターテインメント領域	専門領域 環境・エネルギー領域
成績 72.2	成績 72.9
趣味 5,6,11	趣味 5,8,11
資格 2,8,21,29	資格 2,29
業種 16,24,25	業種 15,34,35,36
給料 最低22から25万円	給料 最低22から25万円
休暇 15~20日	休暇 15~20日
残業 20~40時間	残業 20時間以下

図 11 就活生の個人情報を示す画面

全 18 次元の項目に対して重要度を重みとして付与することで、特徴ベクトルパラメータを生成し、コサイン類似度を計算する。推薦を希望する就活生は、他の全ての就活生と比較した上で、自分以外の類似度が最も高い就活生が過去にエントリーした企業を、関連企業として推薦する。推薦画面のイメージを図 12 に示す。最後に、企業間類似計算による企業推薦の内部処理の様子を図 13 に示す。就活生のデータベースから抽出されたエントリー企業 5 社を対象として、一つずつ企業間類似度計算を行い、この 5 社以外の全ての企業から、それぞれの企業と最も類似した企業を関連企業として推薦する。

学生1と一番状況が似ている学生5の類似度は: 0.9476405366583261	
学生1に推薦する会社は:	
会社ID:10771	
四季報コード	599
会社名	トヨタ自動車
地域	愛知

会社ID:10938	
四季報コード	602
会社名	三菱自動車
地域	東京

会社ID:3408	
四季報コード	176
会社名	日清紡ホールディングス
地域	東京

会社ID:7459	
四季報コード	595
会社名	IHI
地域	東京

会社ID:15022	
四季報コード	1066
会社名	NTTデータ
地域	東京

図 12 就活生間類似度計算による企業推薦の画面

企業1082との比較対象企業2569のベクトル積は: 330.72901523863294	0.5665209783294795
企業1082との比較対象企業2569のベクトルのルート積は: Math.sqrt(2535.6322)=50.35586131462854 Math.sqrt(262.2270414433112)=16.19342587111545 815.420952639407	0.39315111345278264
企業1082との比較対象企業2569のCosSimilarityは: 0.4055929813556265	0.6531818705663717
企業間類似度計算結果によって、学生1に推薦する企業は: 会社ID:31861 四季報コード 96 会社名 NECフィールディング 地域 東京 東洋経済コード 102034200NECフィールディング	0.4955383038059602
会社ID:10007 四季報コード 497 会社名 日立製作所 地域 東京 東洋経済コード 100265731日立製作所	0.6064289751677169
会社ID:10090 四季報コード 500 会社名 富士電機 地域 東京 東洋経済コード 100052702富士電機	0.614503816244981
会社ID:17800 四季報コード 320 会社名 トrendマイクro 地域 東京 東洋経済コード 100424553トrendマイクro	0.6398722314820913
会社ID:4857 四季報コード 742 会社名 ニフコ 地域 神奈川県 東洋経済コード 100003704ニフコ	0.7084661961540702
	0.927092347392988
	0.7446512430587706
	0.6005453790507178
	0.48584973052035335
	0.5163608540797706
	0.6271456938217745
	0.5693470277608281
	0.7287967833368721
	0.4907565609318617
	企業1082と一番状況が似ている企業は企業742類似度は: 0.9736111011372652

図 13 企業間類似計算による企業推薦の内部処理

6. ケーススタディ

6.1 目的と方法

本学の就職活動支援システムである JAIST Career Support System は学内で開発されており、機能の追加および実験が容易である。そのため、開発した機能を被験者に利用してもらうことで、推薦結果が有用であるかどうかをアンケートによって調査した。20 名の日本での就職を検討・希望している大学院生を被験者として募集し、企業のデータベースから抽出された五つの評価基準に基づいたケーススタディを行った。また、アンケートの結果を分析して、推薦機能の妥当性や信頼性、使いやすさなどを総合的に評価した。加えて、被験者からのフィードバックに基づいて提案機能における不十分な点を発見し、さらなる改良に向けた検討を行った。

具体的な手順は以下の通りである。まず、20 名の被験者はそれぞれ就職活動に対する希望などの個人情報をシステムに入力した。次に被験者は、就職活動の際にエントリーしたい 5 社を JAIST Career Support System を利用して自由に検索した。次に、提案手法の妥当性を検討するため、本研究で提案した重み付けを行った推薦と重み付けを行わない推薦を、就活生間類似度計算と企業間類似度計算により行った。被験者は、それぞれの手法により推薦された各企業に対して満足度評価を実施した。最後に推薦機能に関するアンケート調査を行った。

6.2 有効性の判定

被験者は、前述した 4 つの推薦方法(方法 1:重みなし就活生間類似度計算,方法 2:重み付け就活生間類似度計算,方法 3:重みなし企業間類似度計算,方法 4:重み付け企業間類似度計算)それぞれに対して、推薦された企業を一つずつ、「仕事内容の希望への一致」、「給料の希望への一致」、「勤務地の希望への一致」、「休暇の希望への一致」、「残業の希望への一致」及び、「志望程度」の観点からそれぞれ推薦結果の満足度として評価した。加えて、企業名を知っていたかどうかについても確認した。4 つの方法に基づき推薦されたそれぞれの 5 社の各評価基準に対する満足度を集計して分析することで、提案手法の有効性と適用可能性について議論する。

6.3 結果の分析

まず、20名の被験者の「勤務地」、「職種」、「初任給」、「有給年休」、「業種」および「月残業時間」について、重要度の分析を表4に示す。

表4 入力された重要度の統計表

被験者	分散	最大配分の項目と割合
1	66.6667	勤務地 30
2	76.6667	初任給 30
3	66.6667	初任給 30
4	26.6667	業種・職種 20
5	16.6667	初任給・残業 20
6	26.6667	業種・残業 20
7	86.6667	職種 25
8	366.6667	初任給 50
9	693.8667	勤務地 70
10	136.6667	職種 30
11	456.6667	勤務地 60
12	306.6667	初任給 50
13	114.2667	勤務地・初任給 30
14	136.6667	初任給・業種 30
15	136.6667	初任給・職種 30
16	276.6667	勤務地 50
17	56.6667	職種 30
18	194.2667	職種 30
19	486.6667	勤務地 60
20	290.2667	勤務地 50

表4に示す通り、被験者ごとの項目に対する重要度には大きなばらつきがあることがわかる。具体的には、就職活動における、「勤務地」に対する重要度について、「70%」の非常に高い点数を付ける就活生もいた。「5%」の非常に低い点数を付ける就活生もいた。このように、項目に対する重要度の個人差は非常に大きい。したがって、重要度上位の項目について、個人差を反映した重要度が類似計算にどのような影響を与えるかは分析すべき内容であったと言える。

図14に、4つの手法によって推薦された企業に対する全被験者の各観点の平均満足度を示す。これらの平均満足度に関してANOVA[11]による分散分析を行った。有意水準を5%に設定すると、分析結果のP値が0.05より小さければ、比較している4つの方法における各観点の平均値は有意に異なると言える。その結果、「仕事内容の満足度」(P値=0.000)、「給料の満足度」(P値=0.017)、「年休の満足度」(P値=0.000)、「残業時間の満足度」(P値=0.000)、「志望程度」(P値=0.000)はそれぞれ方法間で有意差が認められた。一方、「勤務地の満足度」(P=0.296)については方法間の有意差が認められなかった。ただし、勤務地の満足度に関しては、4つの方法全てで平均が4.5以上であったため、どの方法であっても勤務地に関する満足度は高かったためであると言える。なお、実際に多くの被験者は[首都圏]を勤務地の第一または第二希望としていたが、企業データベースに含まれる多くの企業が[首都圏]であったため、[勤務地の満足度]が低くなりにくかったことが考えられる。

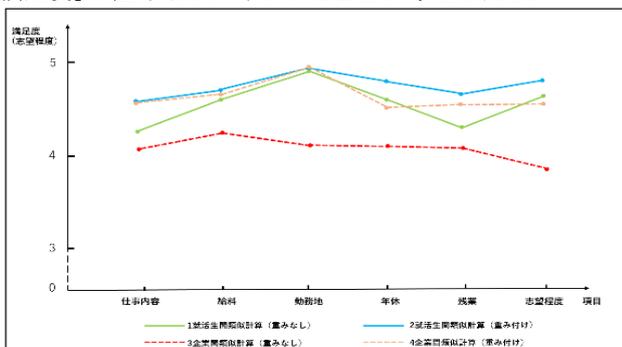


図14 推薦された企業の平均満足度統計結果

次に平均値に有意差があった観点に対して、どの方法間に有意差があったのかを示すために、多重比較を行った結果を図15に示す。二つの方法間の平均の差の有意差を判定するため、Bonferroniの方法の手法を適用した。具体的には、検定で得られた有意確率P値をBonferroniの補正によって修正した。ここで、4つの方法を相互に6回を比較するため、P値が0.008(0.05/6)より小さければ、比較している2つの方法の平均値の間に有意差があると判断できる。この結果、方法1と2の間には大きな差はみられないが、方法2と方法3、方法3と方法4の間には一部で有意差が見られた。したがって、提案した企業が重視する採用の観点と就活生のアピールのギャップに注目した上で、就活生の個人差を反映して重み付けを行う提案手法は一定の効果があったと言える。

仕事内容	方法2	方法3	方法4
方法1	0.020655	0.258376	0.015578
方法2		1.98E-04	0.920176
方法3			1.17E-04
給料	方法2	方法3	方法4
方法1	0.54461	0.026562	0.782442
方法2		3.09E-03	0.319835
方法3			2.30E-02
年休	方法2	方法3	方法4
方法1	0.078302	0.001744	0.510639
方法2		3.96E-06	0.008266
方法3			4.25E-03
残業時間	方法2	方法3	方法4
方法1	0.096057	0.038647	0.168642
方法2		6.72E-05	0.655358
方法3			6.55E-01
志望の程度	方法2	方法3	方法4
方法1	0.013252	0.000388	0.09092
方法2		1.21E-07	0.318811
方法3			1.06E-06

図15 Bonferroniの方法による検定結果

総合的に分析すると、方法2の重みつき就活生間類似度計算が全体として高い評価となっている。ただし、方法1の重みなし就活生間類似度計算との有意差は見られない。また、就活生間類似度計算と比較すると企業間類似度計算の平均値は低い傾向にあり、企業間類似度計算の中では仕事内容、年休、志望の程度において方法4の重みつき企業間類似度計算が方法3の重みなしの手法より高い満足度が得られたことがわかった。

次に、各手法で推薦された企業のうち異業種の企業と企業名を知らない企業とがどの程度含まれていたかについて図16にまとめる。

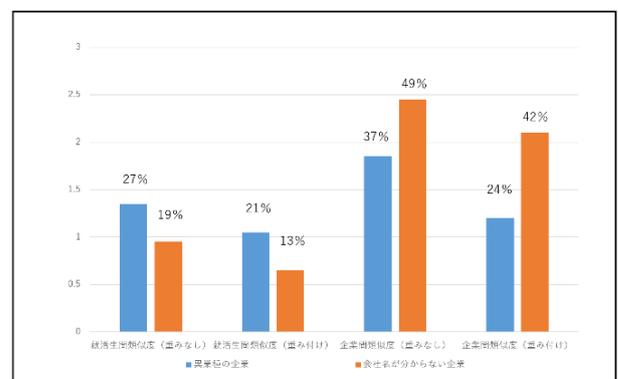


図16 異業種企業と企業名を知らない企業の割合

異業種企業と企業名を知らない企業の割合は提案手法によって対照的な結果となった。まず個人差の重みを付け

ない就活生間類似度計算と企業間類似度計算方法のベースラインと個人差の重みを付ける提案手法を比較すると、全てのケースでベースラインの方が提案手法よりも異業種・知らない企業の件数が多い結果となった。これは、被験者の特徴の重みを利用しないことによって推薦内容の範囲は広がったものの、志望度に関する満足度が低くなったと考えられる。また、就活生間類似度と企業間類似度を比較すると、いずれも企業間類似度の推薦割合が上回った。これは、就活生が知名度の高い大手企業をエントリーする傾向があり、就活生間類似度による企業推薦では、比較的知名度の高い企業が多く含まれていたためであると考えられる。今回のケーススタディは被験者数および各被験者がエントリーした企業数が十分ではないため、類似度計算結果で重複したケースもしばしば見られた。システムを利用する就活生数の増加とエントリー履歴の活用により、提案手法の精度のさらなる向上が期待できる。

7. おわりに

7.1 まとめ

本研究では、まず、就職活動の現状を調査した上で、企業の採用時に重視する観点と就活生のアピールポイントのギャップに注目した。さらに、本学の JAIST Career Support System を例として、従来の就職活動支援システムがますます多様化する就活生のニーズを満たすことができなくなっている課題について議論した。こうした課題を解決するために本研究では、キーワード検索に基づいた企業情報の検索機能に加えて、就活生のニーズに合わせた新たな就職支援に関連する機能の開発を行うこととした。

就活生が実際に企業を検索する場合、自分の既有知識に基づいて検索キーワードを決めることで、企業全体に大きく割合を占めている異業種や企業名を知らない企業の検索が難しくなる。したがって、これらの課題を解決するために、異業種や企業名を知らない企業を推薦できる企業推薦システムが有効な解決手段と考えられる。本研究では、個人情報や検索履歴、企業データベースを活用した協調フィルタリングおよび内容ベースフィルタリングの手法を組み合わせた関連企業の推薦機能を設計・開発した。主に、就活生と同様の興味を持っている他の就活生の検索履歴を活用したフィルタリング手法および、企業データベースから各企業の特徴ベクトルを生成して企業間の類似度を計算するフィルタリング手法により推薦システムを構築した。また、企業側の採用時の観点と就活生のアピールポイントから抽出された項目を類似度計算の重み付けに用いた。

さらに、近年の就活生のエントリー時期が3月頃に集中することから、就職活動の初期において就活生の蓄積されたデータが不十分なケースが考えられる。そこで、就活生の個人差を反映した重みをあらかじめ与えることでコサイン類似度アルゴリズムを改良し、企業推薦機能のスタートアップ時の精度の向上を目指した。

最後に、20名の大学院生を被験者として、本研究で開発した企業推薦機能のケーススタディを行った。重み付けのありなしと就活生間・企業間の類似度計算手法の4つの方法によって推薦された20社の企業に関する被験者の満足度を評価した。これらの方法間の平均値の差をANOVAで分析したところ、就活生間の類似度計算手法は企業間の類似度計算手法と比較して平均値が高い傾向にあり、企業間の類似度計算の中では個人差を考慮した重み付け類似度計算が、個人差を考慮しないベースラインの類似度計算手法よりも満足度が高い傾向となった。さらに、推薦された企業のうち、異業種や企業名を知らない企業の割合を分析し

た結果、一定数の就活生の検索履歴とは方向性の異なる企業の検索を行うことができた。

まとめると、就活生の就活時に重視する観点を採り入れた特徴ベクトルを構成することで、日本の就職活動における「エントリー期間の集中」「企業検索の時間コストの大きさ」、「異業種や企業名を知らない企業の情報検索」などの問題解決の第一歩となったと言える。

7.2 今後の課題

本研究における今後の課題としては、まず、現在の機能は既存のシステムの外側で開発しているため、JAIST Career Support System の「企業推薦機能」として取り込む必要がある。また、今回提案した手法は、就活生間の類似度計算による推薦手法と企業間の類似度計算による推薦手法を別々に行っているため、推薦企業の類似度を統一し、双方の手法を同時に活用して類似度が最も高い企業を優先に推薦することも考えられる。さらに、企業間類似度計算には処理時間がかかるため、実用化を考慮する場合には、システムのプログラムの最適化やデータベースのテーブルの前処理を考慮する必要がある。企業情報データベースは全てのカラムに情報が格納されているとは限らないため、欠損値に対するデータ補完方法についても検討していく必要もある。

謝辞 ケーススタディにご協力いただいた20名の被験者の皆様には、複数回にわたる実験にご理解いただき、心から感謝いたします。

参考文献

- [1] “観察者”.
https://www.guancha.cn/society/2013_05_30_147983.shtml
- [2] “マイナビ2017年卒の就活生のエントリー数”.
<http://kenjasyukatsu.com/archives/1150>
- [3] 鮑若愚, 白井康之, 湊真一. 就職活動就職者向けの企業情報推薦システムの提案. 情報処理学会第76回全国大会. 4N-7.
- [4] 赤木里騎, 福島亮, 徐海燕. 学年を超えた就職支援システムの構築. 平成28年度電気・情報関係学会九州支部連合大会. 11-2A-08
- [5] 三鬼安加里, 南野謙一, 後藤裕介, 渡邊慶和. 研究室における就職支援システムの開発. 情報処理学会第73回全国大会. 3ZD-5.
- [6] “JAIST Career Support System”. <https://careersupport.jaist.ac.jp/>.
- [7] “就職活動の流れ”.
<https://www.thu.ac.jp/career/career/flow.html>.
- [8] “2018就職白書”.
https://www.recruitcareer.co.jp/news/20180215_01.pdf
- [9] 2018年度経団連の“採用選考に関する指針”.
http://www.keidanren.or.jp/policy/2016/081_shishin.pdf
- [10] 推薦システムのアルゴリズム. 神鷹敏弘. 人工知能学会誌2007年11月号[神鷹07], 2008年1月号[神鷹08a]
- [11] 岩井貴充, 中尾彰宏. ジニ係数とコサイン類似度を用いたモバイルトラフィックにおける高精度アプリケーション判定手法(ネットワークシステム). 信学技報 116(484).199-204. 2017-03-02
- [12] ANOVA. 分散分析.
<http://www.shiga-med.ac.jp/~koyama/stat/com-anova.html>.