

就活生に役立つツイートの自動抽出スクリプトの研究開発 Study on Automatic Tweet Extraction Script Useful for Job Hunting Students

田中綾華[†] 大槻明[‡]
 日本大学 経済学部

1 はじめに

近年、新型コロナウイルスの影響で新卒の就職活動が厳しい状況に置かれており[1]。新卒採用市場ではリーマンショック時の就職氷河期の再来という表現をするケースも多くなってきた。このような状況下においては、あらゆる面で新卒の就活生をサポートすることが重要であると考えられる。ゆえに本研究は、学部就活生に役立つツイートの自動抽出スクリプトのための基礎モデルについて研究開発した。

2 提案モデル

2.1 就活に役立つ単語の抽出

大槻ら[2]が実施した卒業生アンケートの一つに、後輩(就活生)のための自由記述アドバイスが存在するため、この自由記述テキストをテキストマイニングすることで、就活生に役立つキーワードを抽出した。そして、これらのキーワードをツイートしているビジネスインフルエンサーを検索することで就活生に役立つツイートをしているアカウントを抽出する。

大槻らの卒業生アンケートの総回答数は143件あり、このうちの105人が自由記述に回答していたため、この自由記述データをテキストマイニング(名詞の頻度分析)することで単語の頻度表を作成した(表1 a)。次にこの頻度表から第3四分位点を求め(表1 b)、この第3四分位点を超える頻度の単語から、就活生に役立つ単語を抽出した。

表1. 名詞の頻度表(表 a)及び四分位数表(表 b)

Term	Info1	Info2	Freq	相対度数
こと	名詞	非自立	69	
自分	名詞	一般	32	
の	名詞	非自立	29	
*	名詞	サ変接続	28	
企業	名詞	一般	21	0.3
参加	名詞	サ変接続	20	
面接	名詞	サ変接続	18	0.26
業界	名詞	一般	17	0.25
人	名詞	一般	17	
方	名詞	非自立	17	
就活	名詞	一般	16	0.23
活動	名詞	サ変接続	13	
就職	名詞	サ変接続	13	0.19
インターン	名詞	一般	13	0.19

四分位数	ave
第1四分位点	2.7
第2四分位点	1.0
第3四分位点	2.0
第4四分位点	69.0

頻度2以上の単語を対象にできる。

2.2 就活生に役立つビジネスインフルエンサーの定義

Influence¹サイトによると、フォロワー数1000人以上でナノインフルエンサー、1万以上でマイクロインフルエンサー、10万以上でミドルインフルエンサーとされているので、本研究ではフォロワー数が1万人以上でかつ、前節で抽出した就活生に役立つ単語を含むツイートを多くしているアカウントのことを就活生に役立つビジネスインフルエンサーと定義した。この「就活生に役立つ単語を含むツイートを多くしている」の定義は次に述べるとおりである。

フォロワー数が1万人以上いるアカウントから「就活」という単語がヒットしたアカウントのうち、直近のツイート内容を確認して仕事で得た経験や就活についての話題を多く取り上げていたアカウント10人を取り上げ、さらにこの10アカウントからそれぞれ、Otsukiら[3]のツイート取得スクリプトを用いて50ツイート(今回は2020年11月7日から直近の50ツイート)を抽出してテキストマイニング(形態素解析及び形態素の頻度分析)を行う。この結果が表2であるが、アカウント毎に単語の頻度総数に差があるため、アカウント毎の単語の頻度総数でアカウント毎の各単語の頻度を割ることで相対度数に変換したうえで比較分析を行う(表3)。

表2. 選定単語の頻度分析の結果

	企業	面接	業界	就活	インターン	仕事	経験	新卒	内定	選考
@marco_biz_man	9	3	3	3	0	18	7	6	2	0
@newsalaryman_21	5	1	5	5	0	5	2	4	1	0
@glocaleducate	2	3	3	7	1	11	4	7	0	0
@Loveable180220	2	1	0	0	0	7	2	0	0	0
@maamorun	1	0	0	3	0	22	4	9	0	0
@Shukatsu_0707	12	18	3	19	4	2	1	1	10	7
@YutaroAkiyama	6	4	0	1	3	16	1	20	3	0
@kamoshika_en	12	4	3	4	1	12	6	12	2	0
@endoayumu	11	4	4	1	0	15	5	3	0	0
@AmaCareer	9	5	1	2	1	13	6	0	1	0

注)行がアカウント、列が前節の選定単語、値が選定単語の頻度をそれぞれ表す。

表2の頻度を相対度数に変換したものが表3である。表3の一番右の列(線形結合値)に示すように、行アカウント毎に、列の各選定単語の相対頻度スコアを足し合わせることで各アカウントの総合スコアを算出し、このスコアの上位(今回は上位5つの)アカウントを本研究では就活生に役立つビジネスインフルエンサーとして抽出した。

表3. 各アカウントの総合スコアの算出

	企業	面接	業界	就活	インターン	仕事	経験	新卒	内定	選考	線形結合値
@marco_biz_man	2.73913	0.782609	0.73913	0.695652	0	1.043478	1.014493	0.26087	0.144928	0	7.4202899
@newsalaryman_21	1.521739	0.26087	1.231884	1.15942	0	0.289856	0.289856	0.173913	0.072464	0	5
@glocaleducate	0.608696	0.782609	0.73913	1.623188	0.1884058	0.637681	0.57971	0.304348	0	0	5.4637681
@Loveable180220	0.608696	0.26087	0	0	0	0.405797	0.289856	0	0	0	1.5652174
@maamorun	0.304348	0	0	0.695652	0	1.275362	0.57971	0.391304	0	0	3.2453768
@Shukatsu_0707	3.652174	4.695652	0.73913	4.405797	0.7536232	0.115942	0.144928	0.043478	0.724638	0.405797	15.681159
@YutaroAkiyama	1.826087	1.043478	0	0.231884	0.5652174	0.927536	0.144928	0.869565	0.217391	0	5.826087
@kamoshika_en	3.652174	1.043478	0.73913	0.927536	0.1884058	0.695652	0.869565	0.521739	0.144928	0	8.7826087
@endoayumu	3.347826	1.043478	0.985507	0.231884	0	0.869565	0.724638	0.130435	0	0	7.3333333
@AmaCareer	2.73913	1.304348	0.246377	0.463768	0.1884058	0.753623	0.869565	0	0.072464	0	6.6376812

2.3 リツイート経路木

前節で抽出された就活生に役立つビジネスインフルエンサーの全ツイートを取得し、太田[4]の手法を参考にリツイート経路木を構築する。ノードはツイートIDであり、リツイート経路木はリツイートを経路木としてつなげていくものであるため、どの投稿をきっかけにしてツイートが拡散していったのかを推定することが可能である。ゆえに本研究では、このリツイート経路木を用いて就活に役立つリツイートサブグラフの抽出を目指す。なお、このリツイートサブグラフの抽出には次節で述べる中心性分析を用いる。

¹ <https://influencer-jp.com/number-of-followers/>

2.3 中心性分析

前節のリツイート経路木をグラフデータに変換し、このグラフデータを対象に次数中心性分析[5]を行う。中心性分析には、近接中心性[6]や媒介中心性[7]などが存在するが、リツイート経路木においては、多くのリツイート(次数)を持っているノードほど重要であると考えられるため、本研究では次数中心性を用いることで、多くのリツイートを持つノード(ツイート ID)を抽出することができる。そして、そのノードに端を発するリツイートサブグラフを抽出することで、就活生に役立つ知見が導出できると考えられる。なお、次章の評価実験では、リツイートサブグラフの抽出までは行っておらず、次数中心性分析で抽出されたツイートの有効性について評価実験を行った。

3 評価実験及び考察

2.2 節の 10 アカウントにおいて、中心性スコア (2,3 節) の top3 のツイートをそれぞれ抽出した (表 4)。そして、日本大学経済学部で実際に就活中の 4 年生 12 名を対象に、令和 2 年 4 月 30 日～5 月 31 日の期間でアンケートを実施し、これらのツイートを実際に見てもらい、就活に役立つか? 役立つかないか? また役立つ場合は何に対して役立つのか? ということを複数回答してもらった。この結果を表 4 の 4 列目 (1 面接対策) 以降に示している。「1 面接対策」から「10 その他」までが就活の何に

役立つかという回答結果、「11 役に立たない」が就活に役立つなかったという回答結果をそれぞれ表している。

まず、「11 役に立たない」の回答総数 30 に対して、1～10 の役立つと回答された総数は 184 であったため、本提案アプローチで抽出したツイートが就活中の学生に役立っていたことが確認できた。

具体的にツイートごとに最も回答数が多かった項目を抽出した結果、次に示す通りキャリア形成に役立つという回答が最も多かった。

- 面接対策に役立つ : 最も回答数の多かったツイート数 1 (以下も同様)
- 自己分析 : ツイート数 2
- ES 作成 : ツイート数 1
- キャリア形成 : ツイート数 10

逆に、最も回答数が多かった項目で役立つなかったと回答されていたツイートが 2 つ (Account4, Score112 と Account5, Score30) 存在していたが、これらの理由について自由記述で回答してもらったところ、次のような理由であった。

- 税金の知識や投資など就活以外に役立つような内容であったこと。
- 新卒採用時の就活では無く転職活動に関するアドバイスであったこと

表 4. 評価実験の結果

Account	Score	Tweets	1 面接対策	2 自己分析	3 ES作成	4 キャリア形成	5 業種選定	6 職種選定	7 企業選定	8 インタビュー選	9 面接質問選	10 その他	11 役に立たない
1	106	新卒で入ったビルメーカーで地方営業したのが27歳のときにToecisco点取った。半年後に本社マーケティング部に異動になった。同じ地方にいたニッパンの営業が同時期にToecisco点取った。半年後にアメリカ支店に異動になった。何が違うのかというToeciscoはキャリア形成の観点で職務のフロー。				9	1	1	2			1	1
1	64	大企業の新卒をやってきて魅力的な職歴を2つの特徴。①自信を持って自分の経験を語れる②誰も知らない質問にも堂々としている③一線級で必要な知識がある④面接でアピールしたいのが明確⑤明確な表現が少い⑥考えと行動に一貫性がある⑦強い言葉を使わない	10	5	4								
1	55	就活で周りには優秀な人が溢れてるし、やること多いし、書籍でもオズだと書かれてた。書籍だと相談したら『就活は、自分の20歳年の生き方を立ち止まって考えて、次の40-50年の生き方の方向性を決める。自分らしい生き方を再定義』するだけだよ。』と書かれ、頭もモヤモヤが吹っ飛んだ。ナイス親父。		6	2	5	3	3					1
2	28	30年間で学んだ成長する習慣 ①小さな事を継続する ②挫折しても再度始める ③常に無理を高く掲げる ④快適の外側の事に耽り損じ⑤自分のメンツを守るまで学ばない ⑥評論家からず人の話をよく聞く ⑦短期間は耐えたら強くなる ⑧逆字書、英検2級で動きつつ学ぶ		2		9	1						1
2	15	英語習得に才能は必要ありません。真事中は英語ニュースを見て、podcast聞いてシャドーイングしながら聞き、通勤電車では英字新聞読む。休日の朝起きて英語ニュースを聞きながら、風呂に入り読む時もラジオを聞く。こんな生活を2年続ければ、単語すら聞き取れず取られた凡人でもToecisco点を取れるんです				7							2
2	9	お説で仕事観で海外就職を目指すのは本当に良かった。悔しくて全く諦め切れず、行きたい、でも失敗したら、と数年間何度も繰り返した。最後は人に背中を押されて決心したが、今となってはあそこ挑戦してなかったらと思う方が多い。やりたい事をやろう、後悔は減っていき人生の質を上げてあげたい		2		9	1	1					
3	171	一見お金がかかるように見えて、経歴上リターンが大きいこと・気がなってきた本は買う・諦められた本は買う・買える時は買う・種類高格の勉強(取柄)・買いたい物ばかり近き・健康には悪い時から投資・短期で将来asで考える・道具は質の良いものを買う・高層階級層クラスで生活		1		6							3
3	62	若いうちに書きたいビジネスプラン・ワークシート①mes.ただし最終的に一番大事なのは「人間力」②英語と会計③積み重ねの知識④体力をつける⑤高い当事者意識⑥学び続ける姿勢⑦投資は自分自身⑧期待を超える意識⑨上り前提を理解する⑩常に考える癖をつける⑪目の前の仕事は全力でやる	1	4	1	10							
3	10	優秀なマナーマンに所属していること①稼ぐ意識がしっかりあるか②だと覚悟がある。短期でのmaxProの最適化という意味ではなく、中長期視点でもそのままでの稼をちゃんと蓄けていくか、コミットがあるかどうか。立場が事業会社/代理店問わず。稼ぐ意識の蓄積が結果にも直結する				6	1	2					2
4	177	何度も言うけど月30万の給料なら稼ぐことは①貯金②月2万③積み立て投資④月2万⑤自己投資⑥月2万⑦勉強⑧毎日2時間半この金額で時間のcon.を最大化する。最終目標の資産を身につけて稼げるスキルを身につける。これら全てをいって稼げるだけ稼げるのにはやるだけの価値はある!				8							2
4	112	社会人になったの頃から知っておきたかったこと・税金の知識・お金の貯め方・お金を働かせる・投資と稼働の違い・会社員に合った副業・副業への時間投資方法・個人の発能力の伸ばし方・資産形成と資産運用の違い これらを理解するまでに30歳まで時間がかかった。		2		3							9
4	55	人に伝わる文章のポイント①選②選③選④選⑤選⑥選⑦選⑧選⑨選⑩選⑪選⑫選⑬選⑭選⑮選⑯選⑰選⑱選⑲選⑳選㉑選㉒選㉓選㉔選㉕選㉖選㉗選㉘選㉙選㉚選㉛選㉜選㉝選㉞選㉟選㊱選㊲選㊳選㊴選㊵選㊶選㊷選㊸選㊹選㊺選㊻選㊼選㊽選㊾選㊿	1			11							
5	40	人より早く出世する為には①100のこと②活躍できる会社を選ぶ③職務経験を積む④評価プロセスと運用実態の把握⑤上司と肩並みになる⑥個人目標は3段階⑦設定⑧戦略マネジメント⑨アポーターを増やす⑩人事権のある人に仕事で名前を売る⑪常にパフォーマンスを上げる⑫自分を磨く				9			3				
5	38	就活生の時は得意な評価される理由が良かった。頭の自とは関係なく、経験労働のホワイトカラーには無関係だとすら思っていた。社会に出て10年経つ今、諦めない「精神力」、繁忙期を乗り切る「体力」、理不尽と付き合う「忍耐力」を兼ね備えた体力者が重宝される理由がよくわかる。		2	4	1	4						3
5	30	①初めて転職活動する人で「オズ」「オズ」知らない人います! 転職のボーナスの権利は即日退職日より後だとボーナス貰えないなるので! ②大企業で働いてきたり、転職する必要がある場合は引継ぎ用紙出したり、海外転職の場合は一時金出たりするし転職もあるあるなので詳しくは聞きましょう					5	1		3			6

4 おわりに

本研究では、就活生に役立つツイートの自動抽出スクリプトのための基礎モデルについて研究開発し、評価実験を通じて提案モデルの有効性を確認した。

今後の課題として、現状のモデルでは就活時と就職後の両方に役立つ単語が抽出されていたため、ここから就職後に役立つ単語を除去する方法について検討したうえでスクリプトとして実装(自動化)していきたい。また、将来的にはリツイートサブグラフの自動抽出スクリプトにまで発展させていきたい。

参考文献

[1] 日本経済新聞 <https://www.nikkei.com/article/DGXZQOUA225KW0S1A920C200000/> (2021/12/27 閲覧済み)。

[2] 大槻明, 森林美咲: 日本大学経済学部卒業生 DB を用いた学修ポートフォリオと就職活動の関連性分析, FIT2021.

[3] Akira Otsuki, Masayoshi Kawamura: The Study about the Analysis of Responsiveness Pair Clustering of Social Network Bipartite Graph, Advanced Computing: An International Journal (ACIJ), AIRCC, Volume 4, Number 6, pp.1-14, 2013

[4] 太田侑介: Twitterにおけるリツイート経路の可視化とユーザ発見支援, 電気通信大学情報通信工学科卒業論文, 2010.

[5] Proctor CH, Loomis CP: Analysis of sociometric data. Res Methods Soc Relat 2:561-85, 1951.

[6] M. A. Beauchamp. An improved index of centrality. Behavioral Science, 10:161-163, 1965.

[7] 津川翔: 情報メディア実験 M28: ソーシャルネットワーク分析, <http://www.mibel.cs.tsukuba.ac.jp/~s-tugawa/jikken/centrality.pdf> (2021/3/20 閲覧済み)

† 日本大学経済学部 Nihon University, College of Economics