

「関西学院大学」の計量テキスト分析*

中野 康 人**

1 はじめに

本稿の目的は、SNS 上で蓄積される関西学院大学に関する話題を時系列的かつ記述的に分析することにある。

SNS、またはインターネットを媒介するソーシャル・メディアは、今日の日本社会では情報収集手段の中で大きな位置を占める。総務省情報通信政策研究所の『情報通信メディアの利用時間と情報行動に関する調査報告書』によれば、10代20代の回答者は「いち早く世の中のできごとや動きを知る」ためにはテレビラジオや紙媒体のメディアよりもインターネットを重視しており、平日のネット利用時間は10代で191.5分、20代で275.0分となっており、利用時間も長い。同調査によれば、数あるソーシャル・メディアの中でもTwitterは全世代で46.2%の回答者が利用しており、10代では67.4%、20代では78.6%と若年層では過半数が利用しているサービスである。Twitterをはじめとするソーシャル・メディアに流れる情報は、そのユーザーの意識や関心を表象する貴重なデータといえる。もちろん、ソーシャル・メディアに流れる情報が人々の意識や行動を正確に反映するものだと主張している訳ではない。しかし、ある時点でソーシャル・メディアに流れる情報は、それを閲覧する人々に影響を与えうる。ソーシャル・メディアにおけるコミュニケーションの中で人々の意見が増幅・強化される可能性が指摘されている (Sunstein, 2001 [2002])。

Salganik (2018 [2019]) は、ソーシャル・メディアを含むいわゆる「ビッグデータ」を観察デー

タして捉える利点と欠点を整理した。利点としては、巨大、常時オン、非反応的の三つの側面があり、欠点としては、不完全、アクセス不能、非代表的、ドリフト、アルゴリズムによる交絡、汚染、センシティブの7つの側面が挙げられている。そしてその上で、ビッグデータを社会科学の研究に活用する際の研究戦略を挙げている。その一つが「事物を数える」である。重要または興味深い研究上の間を見つけ、その理解を深めるために数え上げをおこなうというのである。

本稿では、ソーシャル・メディアの中でもTwitterに注目し、そこで関西学院大学がどのような語られ方をしているのか、を数え上げていく。

2 Twitter データ

2.1 Twitter の概要

ソーシャル・メディアとしてのTwitterは、2006年に誕生した。そのサービスは2006年7月15日にアメリカでスタートした。2022年1月段階で、日本を拠点とするユーザー数は58,950,000ユーザーを超えている (Statista, 2023)。利用登録したユーザーがつぶやき (tweet) を投稿し、公開範囲の対象となるユーザー同士でコミュニケーションがおこなわれる。つぶやきは文字情報だけでなく、画像や映像、URLなども添付可能である。一回のつぶやきに含めることができる文字数には制約があり、現時点でいわゆる全角140文字 (いわゆる半角280文字) 以内に収める必要がある。他ユーザーのオリジナル tweet をそのまま再度 tweet して拡散する retweet (日本語版で

*キーワード：計量テキスト分析、ソーシャル・メディア、ビッグデータ

**関西学院大学社会学部教授

は2010年1月から実装)、他ユーザーの tweet を引用して自らのコメントを付けて tweet する quote (2015年4月から実装)、公開される tweet ではなくユーザー同士に閉じたコミュニケーションを提供する DM 機能などもある。

2.2 データの概要

今回分析対象となるデータは、Twitter API v2 を利用して「関西学院大学」「関西学院」「関学」というキーワードで tweet を検索して収集したものである。検索結果には、オリジナル tweet にくわえて、retweet と quote も含まれる。キーワード検索の結果であるため、関西学院のことを意図していないデータが含まれる可能性がある。例えば、「関東学院」を意図した「関学」や名前の一部に「関学」を含む「一関学院」などが検索対象に引っかかる。後者については後処理でデータから除外できるが、前者についてはデータに含まれたままになっている。検索は、2022年11月から2023年4月にかけて複数回にわけて実施した。獲得されたデータは、検索時点で存在しているユーザーの tweet で、鍵がかかっていない一般に公開されている tweet である。既に削除されたユーザーの tweet、存在しているユーザーの tweet だが個別に削除された tweet、鍵アカウントから発信された tweet は獲得対象とならない。それらを除けば、検索時点で「関西学院大学」「関西学院」「関学」を明示してつぶやいたすべての tweet が収集できている。得られたデータの概要は表1の通りである。

2.3 tweet の頻度

2022年11月時点で確認できる最古の関西学院に関する tweet は、2007年6月2日につぶやかれた

関西にもはしかが本格上陸：関学、立命はしかで休講のニュース。

という内容のつぶやきである。その後、一日に数件程度の頻度から徐々につぶやきが増えていき、2013年以降は日々500件前後の頻度である(表2、図1)。ただし、つぶやき頻度には周期性がある。曜日ごとに頻度を集計すると、平日に比べて土日の tweet が顕著に多い。月ごとに頻度を集計すると、5月と11月をピークとする波があり、7月と1月は頻度的には谷になっている。なお、一日あたりの tweet 件数の最大値は2018年05月22日の35,121件で、一日に一万件をこえる tweet が飛び交ったのはこの時期のみである。それ以外は、学祭や学生団体の試合結果などで数千件の頻

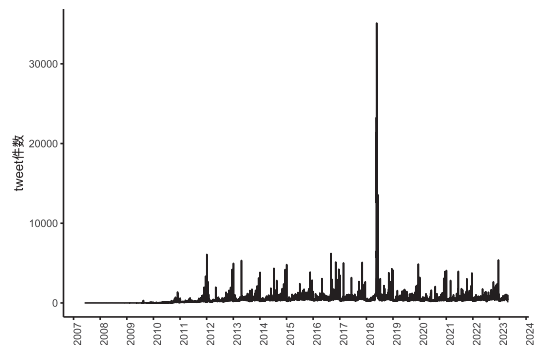


図1 一日あたりの tweet 件数の推移

表1 データの概要

概要	説明
データソース	Twitter API v2
検索語	「関西学院大学」「関西学院」「関学」
獲得データ期間	2007年06月02日から2023年04月24日まで
獲得データ件数	2,675,132件
獲得データユーザー数	544,821ユーザー

表2 一日あたりの tweet 件数

総計量	値
最小値	1
平均値	509.1
中央値	390
最大値	35,121
標準偏差	1,122.5

表3 ユーザー数と tweet 頻度

	ユーザー数	tweet 数平均値	tweet 数中央値	tweet 数最大値	tweet 数標準偏差
全 tweet	544,821	2,420.087	27	44,517	7,811.403
オリジナル tweet のみ	294,078	3,647.041	50	44,517	9,530.688

度に伸びるのが頻度の山となる程度である。

2.4 tweetの種類

また、つぶやきの種類別に頻度をみると、全体的には6:4でオリジナルなtweetの方が多（表5）。各オリジナルtweetがretweetされる回数は多くなく、オリジナルtweetの86.3%は被retweet回数0回である（表6）。被retweet回数の平均値は0.9回、中央値は0回、最大値は15,497回である（表7）。もっともretweetされたtweetの内容は

世の中に蔓延する関西トラップ。「関西大学」→「かんさい」だいがく。「関西学院大学」→「かんせい（くわんせい）」がくいんだいがく。「関西高校」→「かんぜい」こうこう。「関西学院高校」→間違い。「関西学院高等部」。「関西電気保安協会」→「かんさいーでんきはーあんきょーかい」。

で、フォロワー数10,715ユーザーのフリーアナウンサーが2018年05月19日に発信したものである。Twitter Japan社の#拡散の科学プロジェクトによれば、日本語tweetで「バズる」と投稿者が感じるretweet数の中央値は1,300回であるという。このデータで1,300回以上retweetされたtweetは41件で、そのおおよそ半数は2018年5月近辺に分布しており、それ以外は2011年から2023年にかけて年に数件あるかないかという程度である。

2.5 ユーザー

関西学院についてtweetしたユーザー数は544,821ユーザー、オリジナルtweetのみに限定

すれば294,078ユーザーである。年ごとの新規参入ユーザー数（初めて関西学院を含むtweetを投稿したユーザー数）は、2012年以降は毎年3~4万程度であるが、2018年のみ12万ユーザーをこえている（図3）。月毎に同じ数値をみると、学生団体の試合やイベントがあるタイミングと、新入生が入学する四月に新規ユーザーが比較的多くなる傾向があったが、入学前に「#春から関学」というハッシュタグをつけてtweetする傾向が出てからは、四月よりも前に新規ユーザーの増加がシフトしている。

1ユーザーあたりの総tweet数の平均値は2,420件、中央値は27件、最大値は44,517件である（表3）。オリジナルtweetのみに限定すれば、1ユーザーあたりの総tweet数の平均値は3,647件、中央値は50件、最大値は44,517件である。これらのユーザーの58.8%は総tweet数が1件であり、総tweet数が10件未満のユーザーで93.6%を占める（表4）。総tweet数が1,000件をこえるユーザー数は141ユーザーで、このうち総tweet数が10,000件をこえるユーザー数は6ユーザーである。総tweet数が10,000件をこえる上位3ユーザーは、いずれも受験関連のbotである。そして残りの3ユーザーは学内学生団体の公式アカウントである。総tweet数が1,000件をこえる141ユーザーのうち、おおよそ1/3は学生団体や競技関連団体で、残りのユーザーの多くは受験関連の情報を繰り返しtweetしているアカウントである。source（tweetの投稿に使ったアプリ）が確認できる期間のデータから、こうした高頻度ユーザーの使用アプリを確認すると、特に学生団体以外のユーザーはtwittbot.netなどのbotサービスを利用して、機械的にtweetを生成していることがわかる。オリジナルtweetに限定すると、その141ユーザーのtweetが全オリジナルtweetの29.6%をしめている。オリジナルtweetにそれがretweetされたものも含めると、その141ユーザーのtweetが全tweetの23.4%をしめている。

新規ユーザーがそのユーザーアカウントを作成してから初めて関西学院に関するtweetを投稿すまでの時間は、中央値で2.03年後である。ユーザーアカウントを作成してから一ヶ月未満で関西

表4 1ユーザーあたりのtweet件数度数分布表

総tweet 件数範囲	ユーザー数	累積 ユーザー数	累積比率
1件	320,466	320,466	0.588
2~4件	147,707	468,173	0.859
5~9件	41,927	510,100	0.936
10~99件	32,617	542,717	0.996
100~999件	1,963	544,680	1
1000~9999件	135	544,815	1
10000件以上	6	844,821	1

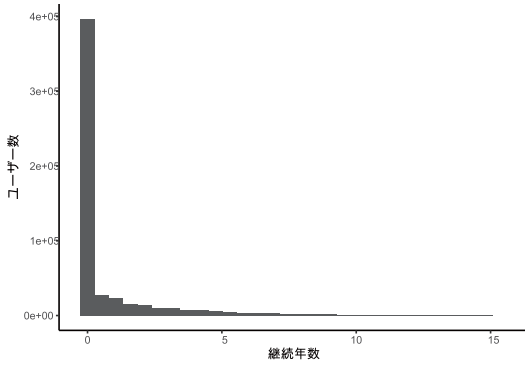


図2 ユーザーごとの関学に関する tweet の継続年数

表5 tweet 種類の比率

tweet 種別	件数	比率
オリジナル	1,652,583	0.618
引用	23,318	0.009
リツイート	999,231	0.374

表6 被 retweet 回数の度数分布表

最大被 retweet 回数範囲	件数	累積件数	累積比率
0 回	1,448,068	1,448,068	0.863
1 回	100,950	1,549,018	0.923
2~4 回	71,788	1,620,806	0.966
5~9 回	29,755	1,650,561	0.984
10~99 回	25,402	1,675,963	0.999
100~999 回	1,309	1,677,272	1
1000~9999 回	61	1,677,333	1
10000 回以上	2	1,677,335	1

表7 被 retweet 回数

総計量	値
被 retweet 回数最小値	0
被 retweet 回数平均値	0.9
被 retweet 回数中央値	0
被 retweet 回数最大値	15479
被 retweet 回数標準偏差	22.25

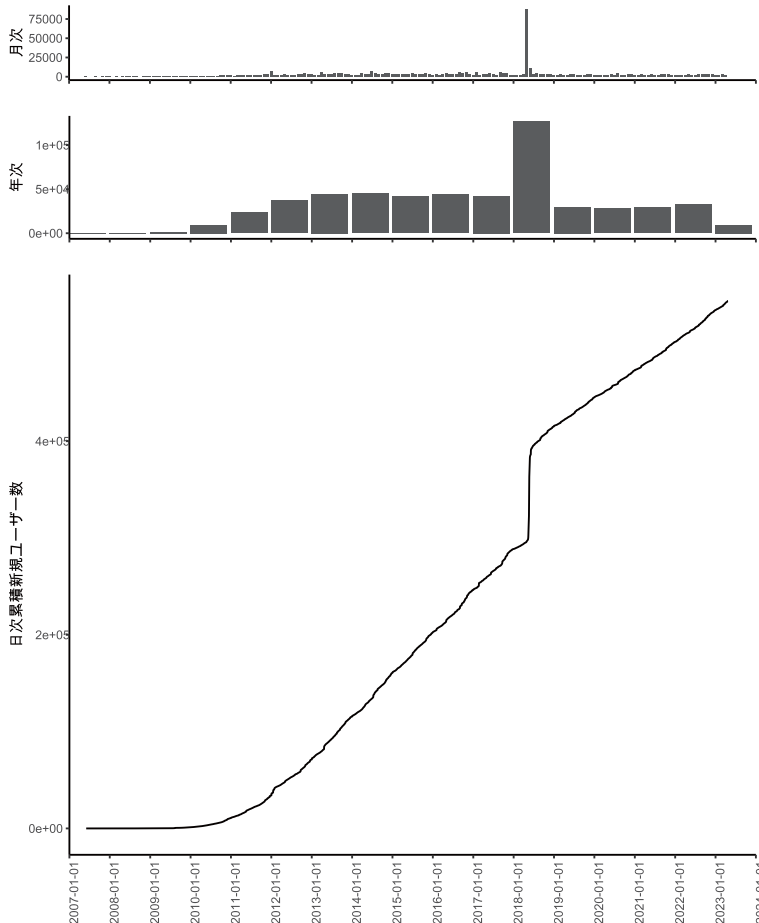


図3 新規ユーザー数の推移

学院に関する tweet をしたユーザーは、すべてのユーザーのうち 5.26% である。初めて関西学院に関する tweet を投稿してから最後に投稿するまでの期間は、中央値で 0.82 年後である (図 2)。ただし、tweet が一回限りのユーザーを除外すると、初めて関西学院に関する tweet を投稿してから最後に投稿するまでの期間は、中央値で 1.99 年後である。4 年を超えて tweet しているのはすべてのユーザーのうち 7.40% である。

2.6 tweet の内容

tweet の内容を形態素解析した結果が表 8 である。形態素解析には、解析器に Mecab (0.996)、辞書に mecab-unidic-NEologd (20220603) を使用した。以下では、形態素解析の結果で名詞として抽出された単語について分析を進める。表 9 は、名詞の出現頻度上位 30 である。ただし、データ取得時にキーワードとした「関西学院」「関西学院大学」「関学」や「人」「事」といった極めて一般的な名詞と数詞は除外している。大学名もしくは地名と思しき単語の他は、「アメフト」「応援」「偏差値」「試験」などがある。学生団体の活動に関する tweet や入試に関する tweet が頻出しているものと推察される。

tweet 内容の概要を把握するため、tweet に含まれる名詞のデータからトピックの抽出を試みる。ここでは、トピック数 30 で tweet が発信された年を共変量とする Structural Topic Model (STM) を適用した。STM は、大量のテキストデータから文書に含まれるトピックとトピックに関連する

言葉を推測して抽出するピックモデルの一種である (Roberts et. al., 2016)。結果として抽出されたトピックの内容が、表 10 である。期待値は、各トピックが一つの tweet に含まれる確率、キーワードはそのトピックを構成する主要な単語となっている。トピックの内容を一つずつ解釈すると次のようになる。

Topic_01 は、関西地区の私立大学をフルネームで記述するようなトピックで、学生団体の活動入試などいずれについても関連する内容である。Topic_02 は、学生がかかわるトピックで、リツイートや拡散を意図した内容になっている。Topic_03 は、URL を含むトピックで、youtube などへのリンクが内容となる。Topic_04 は、関西地区の私立大学を省略したかたちで記述するようなトピックで、学生団体の活動や入試などいずれについても関連する内容である。Topic_05 は、関西四私大・関関同立のトピックで、学部や授業

表 9 名詞頻度

gen	頻度
大学	514,381
関西	336,326
選手	294,915
日大	272,908
学生	172,797
試合	161,425
大阪	159,101
アメフト	158,304
同志社	157,609
関大	149,171
京都	148,565
監督	147,467
女子	120,061
立命館	119,110
龍谷	101,517
日本	98,494
神戸	95,418
西宮	87,296
今日	87,199
近大	83,909
応援	81,395
笑	80,324
sports	78,263
近畿	77,524
偏差値	71,167
結果	69,250
試験	68,373
立命館大学	68,214
大会	67,765
問題	67,696

表 8 形態素解析結果

集計値	値
総語彙数	130,716,277.00
異なり語彙数	181,552.00
出現頻度最小値	1.00
出現頻度中央値	5.00
出現頻度平均値	719.99
出現頻度最大値	3,865,900.00
出現頻度標準偏差	26,589.71
総語彙数	53,572,441.00
異なり語彙数	171,992.00
名詞出現頻度最小値	1.00
名詞出現頻度中央値	5.00
名詞出現頻度平均値	311.48
名詞出現頻度最大値	1,808,576.00
名詞出現頻度標準偏差	10,162.95

表 10 抽出されたトピック

トピック	期待値	キーワード
Topic_01	0.02	立命館大学, 同志社大学, 関西大学, 近畿大学, 京都産業大学, kansai, 甲南大学, 天理大学, a リーグ
Topic_02	0.02	学生, チーム, 名, retweet, 現在, なう, 警戒, 拡散希望, www
Topic_03	0.02	http://, 更新, 動画, 手, blog, 社会学部, twitter, youtube, club
Topic_04	0.03	関大, 近大, 同志社, 立命館, 龍谷, 大経, 南, 関西, 甲
Topic_05	0.02	学部, 授業, 人間, 関関同立, 法学部, 狙い目, 文, ctb, 上位
Topic_06	0.04	関西学院, 兵庫, 県, 入試, 自分, 大学受験, 時計台, e9, 未来
Topic_07	0.04	笑, 気, 笑い, から, 楽しみ, 学校, 子, 先輩, 奴
Topic_08	0.03	人, 今日, 明日, 所, 出身, 卒業, 全員, ありがとうございます, 勉強
Topic_09	0.02	ラグビー, 西宮, 甲東園, 以上, 自習, 神戸女学院, 3 位, 市, 1 位
Topic_10	0.04	試合, - 1, - 2, 結果, 終了, 勝利, 試合終了, soccer, - 3
Topic_11	0.07	関学, ミス, 投票, mi, ハート, ss, in, live, 撮影
Topic_12	0.02	今, 市, 関, 学院, the, 大阪, 兵庫県, 下, 時代
Topic_13	0.05	日, @キャンパス, 神戸, 西宮, 開催, 月, 土, 三田, 祭
Topic_14	0.04	circle, 部, 参加, 紹介, 新歓, イベント, 春から関学, よろしく願います, 生
Topic_15	0.04	関西学院大学, 年, 記事, 神戸新聞, news, 度, 確認, 建築, 初
Topic_16	0.03	関学, 立命, 成功, 分, qb, - 1, 攻撃, 相手, asahi
Topic_17	0.03	大学, 関西, 高校, こう, 合格, だいがく, 関西学院高等部, かん, 完成
Topic_18	0.03	回, リーグ戦, 点, 関, 試合, 春季, 野球, 秋季, 実況
Topic_19	0.05	関学大, 関西, 学生, 対, 男子, 選手権, league, 部, 回
Topic_20	0.04	日本, fighters, 甲子園ボウル, kg, 応援, 一, アメフト, アメリカンフットボール, 日
Topic_21	0.03	事, 前, 時, 後, 話, 目, ob, 公開, lv
Topic_22	0.03	関西学院, kg, mc, ミスキャンパス, 中, 注目, 間, 報告, n・o
Topic_23	0.03	京都, 大阪, 女子, 関西, 近畿, 関西学院, 大学, 産業, 偏差値
Topic_24	0.03	皆, 笑顔, 今年, 本当, ありがとう, 写真, 最後, 時間, 会
Topic_25	0.08	選手, 日大, 監督, 関学大, アメフト, 会見, 問題, タックル, 負傷
Topic_26	0.02	方, より, 御覧, 興味, 不安, 連絡, また, アカウント, a
Topic_27	0.03	教授, 先生, 物, 英語, 学, 問題, 研究, 元, 為
Topic_28	0.03	明治, 大, 早稲田, gt, 東京, 立教, 上智, 法, 中央
Topic_29	0.03	関学, 質問, 皆, 募集, みな, 事, 練習, こと, 友達
Topic_30	0.01	関学, 発表, 他, 期待, 共, 場合, 体, 以降, 確定

に関わる内容になっている。Topic_06 は、関西学院の入試に関するトピックで、大学受験やキャンパスに関わる内容である。Topic_07 は、楽しみや笑いを含むトピックで、感情を表現する内容になっている。Topic_08 は、出身や卒業などを含むトピックで、卒業生に関する内容になっている。Topic_09 は、地名や競技名や順位など含むトピックで、学生団体の活動に関する内容になっている。Topic_10 は、点数や試合などを含むトピックで、サッカーの試合に関する内容になって

いる。Topic_11 は、ミスや投票などを含むトピックで、ミスコンに関する内容になっている。Topic_12 は、「美濃関学院」を含むトピックで、関西学院のデータからは除外すべき内容である。Topic_13 は、@キャンパスや地名を含むトピックで、新聞社の企画記事などに関する内容である。Topic_14 は、春から関学を含むトピックで、新入生に関する内容である。Topic_15 は、神戸新聞や記事を含むトピックで、新聞社のニュース記事に関する内容である。Topic_16 は、QB や攻

撃などを含むトピックで、アメリカンフットボールの試合経過に関する内容である。Topic_17は、高等部や合格などを含むトピックで、関西学院高等部に関する内容である。Topic_18は、リーグ戦や試合などを含むトピックで、野球に関する内容である。Topic_19は、選手権や大会を含むトピックで、学生団体の活動に関する内容である。Topic_20は、Fighters や甲子園ボウルを含むトピックで、アメリカンフットボールの学生日本一を決める試合に関する内容である。Topic_21は、OB や試合などを含むトピックで、学生団体の活動に関する内容である。Topic_22は、ミスキャンパスなどを含むトピックで、ミスキャンパス関西学院に関する内容である。Topic_23は、地名もしくは大学名や偏差値を含むトピックで、入試偏差値に関する内容である。Topic_24は、笑顔やありがとうを含むトピックで、感謝を伝える内容である。Topic_25は、選手やアメフトを含むトピックで、アメリカンフットボールのプレーに関する内容である。Topic_26は、興味や不安を含むトピックで、連絡を請う内容である。Topic_27は、教授や先生を含むトピックで、教員に関する内容である。Topic_28は、関東地区の大学名を含むトピックで、学生団体の活動や入試などいずれについても関連する内容である。Topic_29は、質問や募集を含むトピックで、質問箱に関する内容である。Topic_30は、発表や期待を含むトピックで、入試結果などについての反応に関する内容である。

各トピックの期待値を比較すると、Topic_25とTopic_11が顕著に期待値が高い。全体として、学生団体の活動に関する内容と入試や学生生活を含むそれ以外の内容がおおよそ半々の期待値になっている。図4は、各トピックの期待値を月毎に集計してその推移をプロットしたものである。2012年以前においては、Topic_06(大学受験やキャンパス)、Topic_16(アメリカンフットボールの試合経過)、Topic_18(野球)などが高い期待値を示している。ユーザー数が増えた2012年以降は、全体的に時間の流れの中で細かな増減を繰り返しているが、大きなトレンドとしてはいずれのトピックも期待値は安定しており、いずれかのトピックが特に増加または減少しているわけで

はない。甲子園ボウルのような大きなイベントや、入試や入学などの時間のサイクルが決まっている内容に関するトピックは、その内容に応じて増減がある。

全体的なトピックの傾向は掴めたので、関西学院大学の記述のされ方という視点で tweet 内容をみてみよう。次の tweet は、関西学院大学に関する tweet の中で2022年に多く retweet されたものである。

特殊な読み方をする漢字・立命館（よびこう）・法政（なんかん）・駒澤（ちゅうとはんば）・創価（ゆうめい）・日本（ほんじん）・関西学院（すいせん）・近畿（まぐる）

その retweet 回数は1,084回で極端に多くはないが、他ユーザーからの高評価を意味するいいね（like）回数は10,797回ある。内容はいくつかの大学に関してその名前を揶揄するような内容になっている。このように、ある組織を「ディスる」記述はソーシャル・メディアではよく目にするものである。この tweet を投稿したユーザーアカウントは本稿執筆時点には既に削除されているが、2022年のある時期に万単位のユーザーの目にこの内容が晒されたことは事実である。

関西学院大学については、ここ数年入学者における一般入試入学者比率の低さを指摘する声がある。2022年に多く retweet された tweet にも「すいせん」という言葉が関西学院にあてられている。その是非や当否はここでは論じないが、そうした指摘の極まった形が「推薦学院」という呼称であろう。関西学院大学に関する tweet の中で、「推薦学院」という語を含む tweet の数は1,779件ある。その最古の tweet は、2019年07月14日に投稿されたものである。ただし、tweet の中には関西学院や関学という単語を使わずに推薦学院のみを記述するものもある。そこで、新たに推薦学院を検索語としたデータを追加で収集してその概要を見てみよう。

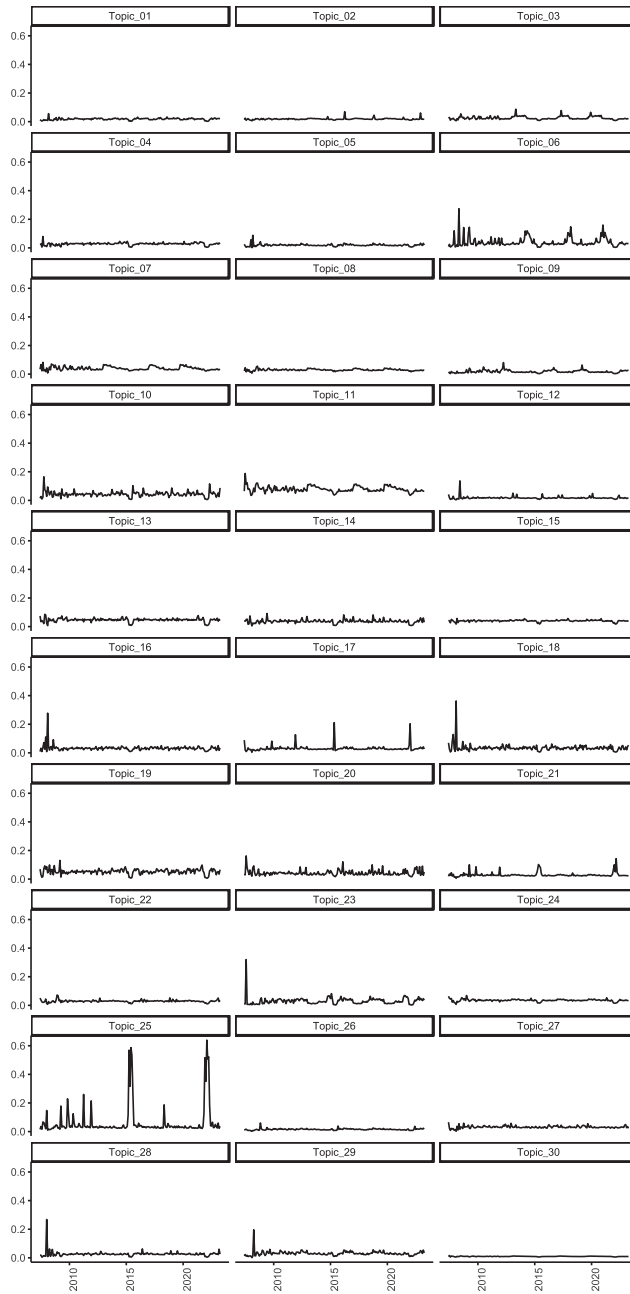


図4 月ごとのトピック期待値の推移

表11 追加データの概要

概要	説明
データソース	Twitter API v2
検索語	「推薦学院」「推薦学院大学」
獲得データ期間	2011年11月08日から2022年12月31日まで
獲得データ件数	1,312件
獲得データユーザー数	887人

3 追加のデータ

3.1 追加データの概要

追加データは、Twitter API v2を利用して「推薦学院」「推薦学院大学」というキーワードで tweet を検索して収集したものである。検索は、2023年1月に実施した。データの制約は、関西学院大学に関するデータと同様である。

3.2 tweet の頻度

2023年1月時点で確認できる最古の推薦学院に関する tweet は、2011年11月08日の tweet であるが、内容からすると関西学院大学に関連するものであるかは判然としない。その後も数年間何件かの tweet が続くが、2016年08月31日の以下の tweet が関西学院と推薦学院を関連づけた最初の tweet である。

関西推薦学院大学てなんですか？ 知恵袋で
 いろんな人が関西推薦学院大学とコメント... <https://t.co/BpEp6qPhuf>

この tweet は、2014年5月にユーザーアカウントが登録されてから2016年11月にかけて658,912件の定型的な形式の tweet (Yahoo! 知恵袋の質問の内容から一定の文字数を切り出し、必ず URL が添付される) を繰り返し投稿しているユーザーによって発信されたものであることから、bot と考えて差し支えないだろう。その後の tweet 頻度は図5の通りで、しばらくは年に数件程度であったものが2019年後半に二桁に、そし

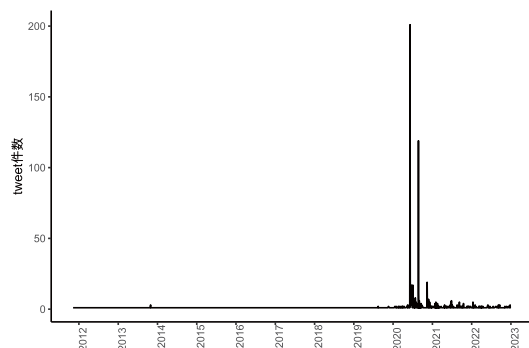


図5 一日あたりの tweet 件数の推移 (推薦学院)

て2020年に三桁の頻度になり、それ以降は月々数十件程度の頻度になっている。いわゆる「バズる」ような頻度ではないが、like の頻度は数千件になっておりこの頃には人目に触れるような内容になっている。その中で2019年から頻繁に tweet された内容が、

【関西学院大学】関関同立の一角で日本一おしゃれなキャンパスとも言われる。通称「関学」。難易度は十分に難関大学だが、一方で推薦率が私大の中で突出して高く「推薦学院大学」とも言われている。しかし全国から学生が集まるのでブランド力はやはり高い大学である。

である。この tweet は、大学入試偏差値の序列や大学を揶揄するようなラベリングを絡めて様々な大学に関する内容を tweet するユーザーによるもので、2019年7月にユーザーアカウントが登録されてから2022年8月にかけて12,103件の tweet を発信している。ユーザー名に bot と明記されており、まったく同内容の tweet を繰り返し流していることから、このユーザーも bot と考えて間違い無いだろう。上記の tweet は、オリジナル tweet の形で一言一句違わずに74回発信されている。また、このデータの中でもっとも衆目に晒されたのが2020年06月07日に発信された、

大学紹介その9「推薦学院大学」関西きってのお洒落大学である。金持ちの子息令嬢が多く、キャンパスではブランド品の話題が飛び交う。「一般入学者ゼロ」を公約として掲げており、一般入学者根絶運動を着々と進めている (2019年入学者の一般入学比率は36

という tweet で、retweet 数725回、like 数4,945回である。この tweet は、2019年6月にユーザーアカウントが登録されてから3,606件以上の tweet を発信し続けている受験産業ユーザーによるものである。当該ユーザーは、YouTube チャンネルでも同種のコンテンツを発信している。

こうした大学序列に関する bot や受験産業などから繰り返される tweet が目立つようになった、

2020年夏頃から単発のオリジナル tweet でも推薦学院という単語が使われるようになる。例えば、学生団体の大会や試合の結果を評する際に推薦学院という単語を使用する tweet が出てきた。

4 まとめ

ここまで、関西学院に関する内容を含む tweet データを収集し、その内容の概要を記述的に紹介した。関西学院に関する tweet は2007年6月にはじめて発信され、今では一日に500件前後が安定的に tweet され、学生団体のイベントなどがある際は一日数千件程度に膨れ上がる。この間、一日に35,000件を超える tweet が発信されたのは2018年5月である。今回のデータで様々な「最大値」を測定するとその多くがこの時期の tweet になる。アメリカンフットボール部の試合中に危険行為があり、ニュース番組で報道されるなど世間的な耳目を集めた時期である。

tweet の内容は、学生団体の活動に関するものとそれ以外の入試や学生生活に関するものが半々ある。毎年繰り返されるイベントや入試などの学事暦に応じて内容の量は上下するものの、大きなトレンドとしては内容の比率は安定している。ただし、その内容の質には変化がありうる。本稿では、一般入試入学者比率の低さを揶揄して使われる「推薦学院」という言葉の使われ方を時系列的に記述した。この呼称は、2019年から2020年にかけて大学序列に関する bot や受験産業などが機械的に繰り返し使用したものであるが、それが2020年以降その他のユーザーにも広まっていったことが確認できた。この呼称に関する最初の tweet で言及されていた Yahoo! 知恵袋についてデータを収集して確認すると、2023年4月段階で17,633件の関西学院関連の質問があり、71,713件の回答がある。この中で「推薦学院」という言葉が使われ始めるのが2016年8月である。特定できる範囲内でみる限り、いくつかの同じユーザー ID から執拗にこの言葉を使った書き込みがあり、その後2020年以降は加速度的に件数が増加している。はじまりは一般入試の対策を生業とする受験産業であったり一般入試における序列に拘泥するユーザーであったとしても、それがソー

シャル・メディアで繰り返し発信されることによって、その他の多くのユーザーに影響を与えたものと言える。各種の推薦入試制度はそれぞれ大学に入学するに値する者を受け入れる制度であるし、入学方式別の入学後の成績分布を知る立場からすると、推薦制度や推薦入学者を蔑むようなラベリングには大きく違和感を覚える。しかし一度定着してしまった呼称は、特にその内容が入試とダイレクトに関わるものだけに、払拭するのは容易ではないだろう。

冒頭で紹介したように Twitter をはじめとするソーシャル・メディアは、情報獲得の媒体として今日の社会では広く浸透している。しかし、今回のデータで確認したように、そこに流れる情報には偏りがある。量の視点でみると、極めて少数のユーザーから発信される情報が多くのユーザーのもとに届けられている。今回は記述的な分析にとどめたが、今後は記述内容の伝播のされ方を詳細に分析していきたい。ただし、今後の展開には大きな懸念がある。今回のデータは、表1、表11で紹介したように Twitter API v2 を利用した。ところが、2023年春にこの API の仕様と利用規約が変更となり、実質的に研究用のデータ収集が困難な状態になった。Salganik (2018 [2019]) が言うところの「アクセス不能」である。Twitter というサービスそのものもゆらいでいる印象がある。今後、社会調査のためのビッグデータとしては Twitter とは異なるソーシャル・メディアの利用を検討する必要がある。

引用文献

- Kudo., Taku, 2013, MeCab <https://taku910.github.io/mecab/> (最終閲覧 2023年6月24日)。
- Roberts., Margaret E., Stewart., Brandon M. and Airoidi., Edoardo M., 2016, 'A Model of Text for Experimentation in the Social Sciences,' *Journal of the American Statistical Association*, 111 (515) : 988-1003.
- Sato., Toshinori, 2015, 'Neologism dictionary based on the language resources on the Web for unidic-mecab,' <https://github.com/neologd/mecab-unidic-neologd> (最終閲覧 2023年6月24日)。
- 総務省情報通信政策研究所, 2022, 『令和3年度情報通信メディアの利用時間と情報行動に関する調査』, https://www.soumu.go.jp/main_content/000831290.pdf.

Statista, 2023, <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>
(最終閲覧 2023 年 6 月 11 日).

Sunstein, Cass R., 2001, Republic.com, Princeton: Princeton University Press. (石川幸憲訳, 2002, 『インターネ

ットは民主主義の敵か』毎日新聞社).

Twitter Japan #拡散の科学プロジェクト, 2023, 「#拡散の科学」, <https://marketing.twitter.com/content/dam/marketing-twitter/apac/ja/insights/kakusan/kakusan.pdf>
(最終閲覧 2023 年 6 月 24 日).

Quantitative Text Analysis of Kwansei Gakuin University on Social Media

Yasuto NAKANO

ABSTRACT

This article analyzes topics related to Kwansei Gakuin University in the contexts where it appears on social media, in a chronological and descriptive manner. Social media has a significant position among means of information dissemination in contemporary Japanese society. Contents discussed in the media could influence users' perspectives. Using Twitter API v2, a total of 2,675,132 tweets mentioning Kwansei Gakuin University were collected to extract how the university attracted attention and what is being discussed. The first tweet related to Kwansei Gakuin University was posted in June 2007, and at present, it appears in about 500 tweets per day, with a significant increase to several thousand tweets per day during events organized by student groups. In all, 544,821 users have tweeted about the university, and 141 individuals are responsible for 29.6% of all tweets. The contents of the tweets are divided equally between those related to student activities and other topics, such as entrance examinations and student life. The negative labeling of Kwansei Gakuin University began with the posts of a small number of users mechanically repeating tweets in large numbers, gradually influencing the language used by general users.

Key Words: quantitative text analysis, social media, big data