

SNSにおける有名人の発信力に関する要因分析

大東文化大学大学院経営学研究科

李詩媛

概要

昨今ではSNSの普及によって、広告主は有名人を用いた情報や広告を消費者へ届けられるようになった。本研究では、中国のSNSにおける有名人のデータを集め、SNSにおける有名人の発信力により大きな影響を与える属性の組み合わせを探し、決定木分析と類似したペアの比較に基づくリコメンデーションを通して、発信力と属性あるいは実績との関連を明らかにした。本研究では限定されたデータならびに主観的評価に基づく属性をもとに分析したものであり、したがってその結果は直ちに应用可能とはいえないものの、将来的に芸能人あるいは芸能人と広告主に対する有用な情報となり得る可能性がある。将来的には、データ収集の適正性を注意し、ファンの視点から考えること（受容性）やパラソーシャル関係を含めて検討していく必要があると考えた。

1. 研究背景と動機

近年、芸能人のイメージキャラクターを利用したブランディングは枚挙に暇がない。短期的に知名度・信頼度を上げることで、知名度の高さにより消費者の印象に残りやすくなること、また、当該タレントのファンが購買してくれるなど多くのメリットが存在している。昨今ではSNSの普及によって、芸能人はSNSを通して自身の様々な面を広く伝え、ファン層を広げながら情報や広告をファンへ届けられるようになった。

ここで発信力とは、SNSを使っての発信や、ブログで発信するという、伝える手段のことだけではなく、「自分の意見や考えを、相手や周囲にわかりやすく伝えるための能力」のことをまとめていう。例えば、ファンの数を考慮して、コメント・リツイート・いいね！の数が多いほど、発信力が強いと考えることができる。一方、SNSにおいては有名人の発信力の差が存在している。今、二人の芸能人の活動内容や世の中に対する影響力が類似しているとして、もし二人の間にSNSによる発信力に大きな差が存在するとすれば、その原因は何か？例えば、代表作の数や活動歴、あるいはプロフィールが近い二人の俳優の中で、1人は発信力が強いが、もう1人が低い場合には、発信力が強い人のSNS上でのアクティビティは大いに参考になるに違いない。その差を明確にすることにより、芸能人の発信力を強めるための参考にすることができる。

一方、広告主においては、発信力とSNS広告との関係は深く、売上高に直接的に影響を与えている。芸能人を起用して広告を出稿することは一般にはコストがかかるため慎重に行わなければならない。本研究では、何らかの属性が発信力に高い影響を与えるのであれば、それらの属性を持つ芸能人を利用することがより有効である可能性があるということを明確にしたいと考えている。

2. 研究目的と意義

過去2年間、人気のタレントショーはたくさんあり（例えば、『PRODUCE 101 JAPAN』）、一部の番組はオンラインメディアでのみ放送されている。その中で人気を博している多くは芸能歴の短い新人である。経験はないが、番組のおかげでインターネットで人気を博し、SNSで視聴者との交流を続けて、多くの広告やパフォーマンスのリソースをもたらしている。逆に、インターネットの力を無視して自分の能力を伸ばすことに力を入れている芸能人もいる。このような芸能人に対する観客の理解はその作品のみにとどまっている。あるいは、実績があってもSNSの運営を軽視する年配の芸能人もたくさんいる。SNSを使う人は常に新しいことを追求しているから、優れた広告リソースをこれらの芸能人にもたすことができず、結果としてSNSを軽視する芸能人にとっては意図せず不利益を被ることとなる。以上の結果は、個々の芸能人に対しては、自身の発信力を伸ばす方法を提示することとなり、一方、広告主においては広告を出稿するための参考にすることができる。

以上のような昨今の芸能人をめぐるSNSの利用状況と発信力への影響力を鑑み、本研究では、SNSにおける有名人の発信力により大きな影響を与える属性の組み合わせを探し、発信力と属性あるいは実績との関連を明らかにする。また、個性（属性）の近さに基づき、発信力の高い人と低い人の中で類似しているペアを発見し、発信力の低い人に対して発信力をより向上させるためのリコメンデーションを提示する。

3. 研究方法

3.1 データ収集

中国において、もっともポピュラーであるWeiboにおいて、フォロワーが多い上位50名の芸能人を対象として選択した。また、これに加えて中国のデータベースKOLRANKによって、2020年4月でSina WeiboというプラットフォームでKBI¹指数が最も高い100人を選択した。重複した部分があるため、合計128人の中国の芸能人を対象に、ファンの数、コメント・リツイート・いいねの数、及び属性のデータを集めた。

表1にKOLRANK2020年4月ウェイボー芸能人KBI指数ランキングの例を挙げる。このランキングで左から右までのデータは、ランキング順番・Weiboアカウント・ファンの数・一ヶ月以内オリジナルウェイボーの数・投稿一つあたりウェイボーの平均コメント数・投稿一つあたりウェイボーの平均リツイート数・投稿一つあたりウェイボーの平均「いいね！」数・KBI指数が示されている。KBI指数を算出する方法は明らかにされていないが、KBI指数は広告代理店やブランド運営部門などの職員が参照し、広告戦略やブランド構築のための信頼できる参考情報として利用しているといわれている。

¹ KBIは、Weibo インフルエンサー指数を指す。これは、Weibo アカウントの影響力、普及力、およびユーザー価値を評価するために KolRank プラットフォームで使用される重要な指標である。インデックスが高いほど、アカウントの品質は高くなる。

表 1 KOLRANK 2020年4月ウェイボー芸能人KBI指数ランキング

排名	微博账号	粉丝数	原创微博数	平均评论数	平均转发数	平均点赞数	KBI指数 ^②	操作
1	 蔡徐坤 1776448504	3342万+	6	100万+	100万+	1926120	2319	☆ (+) ③
2	X玖少年团肖... 1792951112	2679万+	1	100万+	100万+	6692040	2307	☆ (+) ③
3	努力努力再努... 27068969...	4994万+	9	76万+	100万+	823534	2289	☆ (+) ③
4	Dear-迪丽热巴 1669879400	7267万+	8	60万+	79万+	681286	2261	☆ (+) ③

3.2 決定木分析

決定木とは、ツリーによってデータを分析する手法で、ある事項に対する観察結果から、その事項の目標値に関する結論を導くことができる。内部の節点の変数に対応し、子である節点への枝はその変数の取り得る値を示す。葉（端点）は、根(root)からの経路によって表される変数値に対して、目的変数の予測値を表す。データから決定木を作る機械学習の手法のことを決定木学習（decision tree learning）、または略して単に決定木と呼ぶ。

決定木による分類モデルはその分類にいたる過程が容易に解釈できるので、決定木はデータマイニングでよく用いられる。その場合、決定木は、葉が分類を表し、枝がその分類に至るまでの特徴の集まりを表す木構造を示す。

決定木は、データの集合を表現したり分類や法則化を助ける数学的手法、計算手法であるともいえる。データは説明変数のベクトル X と目的変数 y からなる以下のような形式のレコードである。

$$(X, y) = (X_1, X_2, X_3, \dots, X_n, y)$$

目的変数 y は、分類や法則化を行う対象であり、残りの変数 $X_1, X_2, X_3, \dots, X_n$ は分類や法則化を行うための説明変数である。

本研究では、発信力という目的変数により強い影響を与える簡潔な属性の組み合わせを探し出すために、Weka（参考文献追加）を用いて、決定木分析を応用することとした。

Wekaは、ニュージーランドのワイカト大学（University of Waikato）のIan H. Witten、Eibe Frankを中心とした機械学習の研究者によって開発され続けている、グラフィカルユーザーインターフェイス、標準のターミナルアプリケーション、またはJavaAPIを介してアクセスできる検証されたオープンソースの機械学習ソフトウェアである。Wekaは、データの前処理、分類と予測、クラスタリング、相関ルール、視覚化に関するアルゴリズムの集合体である。Wekaでは、データセットの中の列(変数)を属性(attribute)、行(個体)をインスタンス(instance)、特定のタスクを実行するアルゴリズムの集まりをスキーム(scheme)、判別・分類を行うスキームを分類器(classifier)、樹木モデルを決定木(decision tree)と呼ぶ。

3.3 類似したペアの比較に基づくリコメンデーション

次に、属性が類似しており、かつ発信力に大きな差があるペアを抽出した。抽出されたペアにおいて、発信力が弱い人が発信力を強めるためには何をすべきかについて、両者の差分に着目してリコメンデーションを行うこととした。

4. 研究内容と結果

4.1 データの収集

まずはじめに、Weiboでフォロワーが多い上位50名の芸能人を抽出した（表2）。

表 2 ファン数が最も多い50人の芸能人の最新の10件記事のウェイボーデータ

名前	謝娜	何炅	杨幂	Angelababy	陈坤	赵丽颖	赵薇	易烊千玺	姚晨	邓超
リツイート	145,750	26,543	2,013,377	762,588	99,339	3,603,435	1,090,864	10,702,794	4,133	144,025
コメント	83,217	30,623	1,286,737	447,259	101,617	1,605,227	91,443	3,575,136	できない	143,404
賛成数	811,383	885,840	2,314,478	1,864,507	432,299	5,263,969	494,170	13,038,666	85,426	2,337,761
ファンの数	125,617,949	119,872,535	107,513,800	102,137,568	93,474,390	86,714,018	85,624,291	84,499,838	83,831,865	80,908,522
発信力(%)	0.83	0.79	5.22	3.01	0.68	12.08	1.96	32.33	0.11	3.24
名前	王源	张杰	王俊凯	林心如	唐嫣	林志颖	郭德纲	胡歌	陈乔恩	陈赫
リツイート	8,593,112	2,006,452	7,116,501	87,480	542,948	2,831	193,164	745,455	46,401	63,791
コメント	7,292,171	270,936	4,032,994	できない	851,534	6,745	できない	431,802	64,538	79,192
賛成数	19,552,648	1,700,209	11,417,883	202,501	1,135,038	81,162	3,169,971	2,301,625	709,930	1,469,363
ファンの数	80,493,491	78,314,369	77,686,279	76,797,182	76,428,656	73,539,621	72,007,882	71,720,116	71,121,593	68,824,189
発信力(%)	44.03	5.08	29.05	0.38	3.31	0.12	4.67	4.85	1.15	2.34
名前	迪丽热巴	王力宏	刘亦菲	罗志祥	贾乃亮	范冰冰	鹿晗	黄晓明	李易峰	文章
リツイート	9,049,117	1,711,173	1,090,098	768,253	183,938	160,257	4,134,926	50,769	1,287,148	2,971
コメント	4,250,137	196,837	955,251	3,010,599	149,384	95,233	890,399	55,606	791,476	できない
賛成数	7,071,626	391,937	4,428,829	44,555,196	321,579	1,049,437	5,738,762	251,553	2,427,099	382,818
ファンの数	68,417,028	67,976,012	67,341,216	65,959,613	6,490,187	63,081,005	61,997,276	61,009,657	58,368,120	56,941,132
発信力	29.77	1.12	9.61	73.28	10.09	2.07	17.36	0.59	7.72	0.68
名前	李晨	薛之谦	杨紫	黄子韬	林俊杰	韩庚	林更新	刘烨	杨洋	吴亦凡
リツイート	45,562	1,015,835	6,630,828	1,675,702	223,334	146,381	1,292,251	42,771	2,857,930	2,436,533
コメント	26,326	680,658	946,317	1,255,754	89,153	132,103	182,522	56,922	1,613,309	697,070
賛成数	314,327	2,727,310	7,382,905	3,188,903	576,436	198,297	2,918,884	261,671	4,309,846	3,662,940
ファンの数	55,966,243	55,756,679	55,523,601	55,155,450	53,794,184	53,795,062	53,202,887	53,081,673	51,381,393	50,734,258
発信力	0.69	7.93	26.94	11.1	1.65	0.89	8.26	0.68	17.09	13.4
名前	张艺兴	高圆圆	范玮琪	戚薇	韩寒	高晓松	李小璐	郑恺	马伊琍	徐璐
リツイート	10,000,000	32,751	3,503	235,867	107,650	11,712	7,414	329,866	7,852	57,527
コメント	10,000,000	17,526	できない	209,219	121,493	できない	できない	272,135	できない	20,249
賛成数	9,546,483	251,985	149,664	589,911	1,113,185	31,081	617,386	6,431,335	85,482	192,736
ファンの数	49,231,340	49,047,080	47,350,597	46,584,570	45,795,985	45,206,685	45,045,853	45,351,595	44,818,250	44,542,743
発信力	60.02	0.62	0.32	2.22	2.93	0.09	1.39	15.51	0.21	0.61

各芸能人に関しては、以下のような属性を用意した。

- ジャンル：対象とする芸能人の活動する分野。歌手、役者、アイドル、ダンサー、“？”の5つに分類した。司会者、漫才師、作家は数が少なかったため、その他という意味で“？”として分類した。
- 演技：演技の熟練度合いの評価。百度と知乎という2つプラットフォームで調べて総合的な口コミに基づき、著者らが判断したものである。演技が高評価を得ている場合であれば‘1’、そうでなければ‘0’とする。
- ルックス：ルックスがこの人のキャリアに対して及ぼす効果。外部の評判情報などをもとにした主観的な推測である。ルックスが明らかにその芸能人の評判にポジティブな影響を与えていると判断できる場合は‘1’になる。そうでない場合は‘0’とする。
- 徳望：対象者の徳望に関する評価。これは一般には年齢とも関係がある。徳望があったら、‘1’に、そうでない場合は‘0’とする。ネット等の評判をもとに、著者が主観的に設定したものである。

- エンターテイメント：バラエティー番組での表現力やエンターテイメント力を推測したもの。当該芸能人の実績や外部の評判情報等をもとにした主観的な評価である。表現力やエンターテイメント力が強いなら、'1'になる。そうでない場合は、'0'とする。
- 代表作の多さ：映画、テレビ番組、歌等の代表作が多いか少ないか。主観的に見れば、多いのは'1'になる。逆に、少ない場合は'0'とする。過去の実績をもとに、著者が主観的に設定したものである。
- ゴシップ：過去に悪いゴシップがあった場合は、'1'とする。そうでない場合は'0'とする。ネット等の評判をもとに、著者が主観的に設定したものである。
- 親和力：ファンに対して距離を置かず活動しているかどうかということ指している。普段の表現と発信した内容によって主観的な推測である。低いから高いまで5段階で評価している。
- 話題の豊富さ：当人の仕事や活動範囲を超えて、広範囲な話題に知見があると思われる場合は'1'になる。逆に、そうでなければ'0'になる。ネット等の評判をもとに、著者が主観的に設定したものである。
- 芸歴：デビューから計算した数値データ（年数）。
- ファンの性別：。女性のファンが多いか、男性のファンが多いか、あるいは両方とも多いかの3種類に分けられる。主観的な評価である。
- 呟くタイプ：Weibo個人ページで発信する主なタイプを指し、テキスト・画像・ビデオ・総合の4種類に分けられる。過去の投稿内容をもとに主観的に判断した結果である。

なお、目的変数である発信力(L)は以下の式により評価する。

$$\text{発信力} = \log(1 + (\text{リツイート} \cdot \text{コメント} \cdot \text{賛成数}) / \text{ファンの数})$$

ここで、リツイート・コメント・賛成数は最近の10件ウェイボー発信の総合である（4月に限らず）。なお、リツイート、コメント数、賛成数は個人による差が極めて大きいため、上記のように対数をとって発信力を評価することとした。

以上の属性に対して、抽出された芸能人の評価を表3に示す。

表 3 芸能人の属性データ

名前	性別	年齢	ジャンル	演技	顔	徳望	エンターティ代表作の多	ゴシップ	親和力	話題の豊富さ	芸歴	多数なファンの性別	味くタイプ	LOG10(1+L)	発信力	
遊郷	女	39	?	0	0	0	1	1	0	5 (最高)	1	24	同方	3ビデオ	0.26202033	0.82818579
何畏	男	46	?	0	0	1	1	1	0	5 (最高)	1	25	同方	4総合	0.2920453	0.78667394
杨幂	女	34	役者	0	1	0	0	1	0	2	0	14	同方	4総合	0.79394438	5.22220589
Angelababy	女	31	役者	0	1	0	1	0	0	4	1	17	同方	4総合	0.60314577	3.01001293
陈坤	男	45	役者	1	1	1	0	1	0	3	0	21	女	4総合	0.22465311	0.67746363
赵丽颖	女	33	役者	1	1	0	0	1	0	3	0	14	同方	4総合	1.11651498	12.0772065
赵薇	女	44	役者	1	1	1	0	1	0	3	0	17	同方	4総合	0.47099018	1.95794556
葛荟婕	女	20	アイドル	1	1	0	1	0	0	4	1	10	女	4総合	1.52280136	32.3273945
姚晨	女	41	役者	1	1	0	0	1	0	4	0	14	同方	1テキスト	0.04408159	0.10683169
邓超	男	41	役者	1	0	0	1	1	0	4	1	19	同方	1テキスト	0.62784083	3.24463967
王源	男	20	アイドル	0	1	0	0	0	0	5 (最高)	0	7	女	3ビデオ	1.65346177	44.0258343
张杰	男	38	歌手	0	0	0	0	1	0	3	0	16	同方	3ビデオ	0.78383307	5.07901302
王俊凯	男	21	アイドル	0	1	0	0	0	0	3	0	8	女	4総合	1.47783543	29.0495743
林心如	女	44	役者	1	1	0	0	1	0	3	0	27	同方	2画像	0.13912102	0.37759328
唐嫣	女	37	役者	1	1	0	0	1	0	3	0	11	同方	2画像	0.63444187	3.30964867
林志颖	男	46	役者	1	1	0	0	1	0	3	0	28	女	2画像	0.05052922	0.12338655
郭德纲	男	47	?	0	0	1	1	1	0	5 (最高)	1	25	同方	2画像	0.75362209	4.67050954
胡歌	男	38	役者	1	1	1	0	1	0	4	1	15	同方	4総合	0.7672031	4.85063633
陈乔恩	女	41	役者	1	1	0	0	1	0	3	0	19	同方	4総合	0.33328137	1.1541769
林赫	男	35	役者	1	0	0	1	1	1	4	1	11	同方	2画像	0.52409773	2.34270251
迪丽热巴	女	28	役者	0	1	0	0	0	0	3	1	7	同方	4総合	1.48819208	29.774576
王力宏	男	44	歌手	0	1	1	0	1	0	3	0	25	同方	4総合	0.32591847	1.1196395
刘亦菲	女	33	役者	1	1	1	0	1	0	3	0	18	同方	2画像	1.02587869	9.61399034
罗志祥	男	41	歌手	1	1	0	1	1	1	5 (最高)	1	25	同方	2画像	1.87086164	73.2782468
贾乃亮	男	36	役者	0	0	0	0	0	0	3	1	18	同方	4総合	1.04495636	10.0906338
范冰冰	女	39	役者	1	1	0	0	1	0	3	1	24	同方	2画像	0.48894776	2.06865284
鹿晗	男	30	アイドル	0	1	0	1	0	0	4	1	8	女	4総合	1.26392456	17.3621935
黄晓明	男	43	役者	1	1	0	0	1	0	3	1	22	同方	2画像	0.2004878	0.58667434
李易峰	男	33	アイドル	0	1	0	0	0	0	4	0	13	女	3ビデオ	0.94049124	7.71949311
文卓	男	36	役者	1	0	0	0	1	1	2	0	16	同方	4総合	0.22466636	0.67752253
李晨	男	42	役者	1	1	0	0	1	0	4	0	23	同方	2画像	0.22790671	0.69008563
薛之谦	男	37	歌手	0	1	0	1	1	1	5 (最高)	1	15	同方	2画像	0.95105188	7.93412212
杨幂	女	28	役者	1	1	0	0	1	0	5 (最高)	0	18	同方	4総合	1.44628216	26.9435675
黄子韬	男	27	アイドル	0	1	0	1	0	0	5 (最高)	1	8	女	4総合	1.08266197	11.0966625
林俊杰	男	39	歌手	0	0	1	0	1	0	5 (最高)	0	17	同方	4総合	0.42364751	1.65245187
韩庚	男	36	アイドル	0	1	0	0	0	0	3	0	15	同方	4総合	0.27560678	0.88629138
林更新	男	32	役者	1	1	0	0	1	0	3	0	9	同方	2画像	0.96653151	8.25830561
刘烨	男	42	役者	1	1	1	0	1	0	3	0	21	同方	4総合	0.22550624	0.6807698

4.2 決定木分析

4.2.1 Wekaによる分析

本研究では Weka に組み込まれているアルゴリズムのうち、J48 を使用した。J48 (C4.5) は属性で分類する決定木において最もポピュラーな分類法である。

C4.5 では、ゲイン比 (gain ratio) を分岐基準としている。ゲイン比は、エントロピーとゲイン値から求められる。

データの例を挙げると、芸能人の属性に関するデータがあるとする。ここで、目的変数は発信力とする。全ての属性 (説明変数) についてゲイン比を求め、その中で最も大きいゲイン比を持つ「変数」を決定木の分岐に用いる変数の第1候補とする。変数が連続の量的データの場合は、群間の平方和が最大になる変数を第1候補とする。分割されたデータに対して、この操作を繰り返して決定木が生成される。

次に、手動で枝刈りする。全部の属性を説明変数として使用すると、正しく分類されたインスタンス (Correctly Classified Instances) が 43.3071% しかない、分析データに依存したノイズとなりうるからであるから、さらに枝刈りする必要がある。前処理 (preprocess) でノイズになっていると思われる属性を一つずつ削除して、この指標 (Correctly Classified Instances) が高くなるような説明変数のみを採用することとした。

4.2.2 決定木分析の結果

最終的に得られた決定木、ならびに Confusion Matrix を図1に示す。Confusion Matrix とは、もとのデータの正しいクラス分類と、決定木の結果得られた分類結果をマトリクスとしてまとめたものであり、分類学習モデルの性能を測る指標として使われる。

```

年齢 <= 34
| ジャンル = 役者: 1~1.5 (33.51/15.13)
| ジャンル = アイドル
|   性別 = 男
|   | 話題の豊富さ = 1
|   |   年齢 <= 21: 1.5~2 (2.0)
|   |   年齢 > 21
|   |     年齢 <= 25: 2~3 (2.0)
|   |     年齢 > 25: 1~1.5 (5.0/2.0)
|   |   話題の豊富さ = 0
|   |     年齢 <= 18: 1.5~2 (3.0/1.0)
|   |     年齢 > 18
|   |       年齢 <= 27: 1~1.5 (15.48/5.48)
|   |       年齢 > 27: 0.5~1 (2.45/0.96)
|   |   性別 = 女
|   |     話題の豊富さ = 1
|   |       年齢 <= 23: 1.5~2 (3.0)
|   |       年齢 > 23: 1~1.5 (4.0)
|   |     話題の豊富さ = 0: 1.5~2 (6.0)
|   ジャンル = 歌手: 1.5~2 (12.56/5.28)
|   ジャンル = ?: 1~1.5 (0.0)
年齢 > 34
| 年齢 <= 38: 0.5~1 (13.0/6.0)
| 年齢 > 38: 0~0.5 (25.0/5.0)

Number of Leaves :    14

Size of the tree :    25

```

	a	b	c	d	e	<-- classified as
20	2	0	0	0	0	a = 0~0.5
4	6	9	2	0	0	b = 0.5~1
0	4	29	8	2	0	c = 1~1.5
1	0	10	18	3	0	d = 1.5~2
0	1	4	2	2	0	e = 2~3

図1 決定木分析の結果

図1より、性別、年齢、ジャンル、話題の豊富さという4つの属性が発信力に与える影響が高いことがわかる。芸能人のこの4つの属性によって、発信力を「A:0~0.5」「B:0.5~1」「C:1~1.5」

「D:1.5~2」「E:2~3」という5つの分類に正しく分ける確率は59.06%である。目的変数であるクラスの数5つであるため、仮にランダムで推測した場合の正解率が20%とすれば大きな識別力があるといえる。また、ACDという三つの分類を推測するの正確率がBEという二つの分類より高い。そして、この4つの属性の中では、主観的な属性は「話題の豊富さ」のみであり、その他（性別、年齢、ジャンル）は客観的な指標である。以上のことから、本結果においては、主観的な評価よりは客観的な評価の方が発信力に対する説明力が高いといえる。

4.2.3 課題

データ数が少ないこと、また選んだ属性が限定的であったため、精度と実効の問題が存在していると考えられる。また、属性は主観的に判定したものが多い。無論、話題の豊富さなどの変数は、もとより客観的な評価自体が存在しないため、こうした変数を仮に主観的であっても組み入れる意義は大きいと考える。しかしながら一方で、主観の度合いによって結果に大きく影響している可能性はある。

4.3 類似したペアの比較に基づくリコメンデーション

4.3.1 ペアの作成

各芸能人二人に対してそれぞれ類似度を算出し、類似しているペアを算出する。ただし、性別・ジャンルは一致して分類する方が、参考にする価値が大きいと考えられるため、性別、あるいはジャンルが異なるペアは除外する。

芸能人 i, j の類似度を、 R_{ij} で表す。両者が類似しているほど、 R が小さくなる。全部で14の属性があり、 X_{ij}^k は属性 k における芸能人 i と j の類似度を示す（性別・年齢・…・呟くタイプ）。ただし、性別とジャンルのいずれかが異なる場合には、リコメンデーションは実質的な意味を持たないと考え、ここでは、性別とジャンルが同じもののみを対象として比較する。また、各項目は最大値と最小値の差分により正規化されているものとする。

ペアの年齢 (N) の類似度 (式1)

$$N_{ij} = \frac{|N_i - N_j|}{N_{max} - N_{min}}$$

総類似度 R の計算式 (式2)

$$R_{ij} = \sum_{k=1}^{12} |X_{ij}^k|$$

4.3.2 類似度の高いペアの算出

前の過程において R が小さいほど類似度が高い。また、全般的な結果を参照し、 R が5より小さい場合には、属性としてかなり近いということが直観的にも納得できると思われたため、閾値として5を設定した。したがって、例えば二人のペアの R が5より小さい場合は、二人の差異が大きくないといえるが、もし発信力における差が大きいなら、改善の余地があり、比較の価値が高い。このため、類似度が高いペアの中で、発信力の差が大きくなるケースを抽出する。

4.3.3 抽出したデータの観察

以上のような方法により、類似しており、かつ発信力に差があるペア100組を抽出した。ほとんどのペアの二人は10歳以上の年齢差が存在している。若い人が活躍している方が多く、SNSもよく使われているので、年齢が発信力に高い影響を与えていることは明らかである。次に、より類似しておりかつ発

信力の異なるペアに焦点を当てるために、年齢差が8歳以下のグループを観察する。このようなペアは10組がある。表4に抽出したペアと、想定されるレコメンデーションを示す。

表 4 年齢差8歳以下の類似のペア

ユーザ1	ユーザ2	発信力1	発信力2	発信力の差	年齢1	年齢2	年齢の差	レコメンデーション
53	28	2.6642	1.2639	1.4002	29	30	1	28の代表作とゴシップを多くなり、つぶやくタイプは「画像」に変える方がよい
61	38	2.0039	0.9665	1.0374	29	32	3	38は女性のファンを増加し、呟くタイプは「総合」に変えるほうがよい
67	21	1.7509	0.5241	1.2268	31	35	4	21はゴシップと話題の豊富さを減って、呟くタイプは「総合」に変えるほうがよい
69	4	2.0721	0.7939	1.2782	29	34	5	4は話題の豊富さを増加し、呟くタイプを「画像」に変えるほうがよい
69	24	2.0721	1.0259	1.0462	29	33	4	24は話題の豊富さを増加するほうがよい
85	44	2.0117	0.1217	1.8900	37	44	7	44はゴシップを控える方がよい
107	38	2.1189	0.9665	1.1524	30	32	2	38は女性のファンを増加し、呟くタイプは「総合」に変えるほうがよい
114	69	0.9916	2.0721	1.0805	30	29	1	114は話題の豊富さを増加するほうがよい
115	107	0.8319	2.1189	1.2870	23	30	7	115は演技をうまくなって、女性のファンを増加し、呟くタイプは「総合」に変えるほうがよい
124	85	0.8849	2.0117	1.1269	42	37	5	124は話題の豊富さを増加し、呟くタイプは「総合」に変えるほうがよい

4.3.5 レコメンデーションに関する考察

一般的には、同じ属性のタイプであっても呟くタイプが「画像」と「総合」になっている方が発信力が高いということがわかった。また、女性のファンを増加し、ゴシップを減らすことが発信力に良い影響を与える（No. 53のゴシップは芸能人自分の悪いではなく、ファンの大規模な不正行為のせいでNo. 53の芸能人に悪い影響を与えた）。さらに、芸能人の能力（演技、代表作）を高めると発信力も強くなるのがわかる。

5 分析のまとめ

5.1 決定木による分析

分類の正確率が一番高い組み合わせは、性別・年齢・ジャンル・話題の豊富さという4つの属性による分類である。結果として、分類正確率は59.06%である。これにより、「もしジャンルはアイドルだったら」「もし話題の豊富さが高いだったら」という異なる条件から、最終的に発信力はどの範囲にあるかその確率を割り出すことができる。選択した属性が有限であり、データもかなりの程度主観的であるから、精度が限られている。

性別とジャンルが発信力に強く影響していることは仮説通りであるが、第三部分でデータ処理したところでも性別とジャンルが同じ二人対象を比較していた。しかし、年齢と話題の豊富さの重要性については分析により明らかになった点である。特に年齢は重要な属性であり、活躍の程度と関連していて、発信力にも影響している。年齢と同じ、テレビで出番の頻度や、最近のドラマや映画に出演しているかどうか、重要な属性として加える方が良いと考えられる。

5.2 類似性に基づくリコメンデーション

本分析の結果から、呟くタイプは一般に「画像」と「総合」の方が発信力が高い。また、女性のファンを増加し、ゴシップを減らすことが発信力に良い影響を与える。さらに、芸能人の能力（演技、代表作）を高めると発信力が強くなることがわかる。

「ペアを作る」部分は、年齢が発信力に与える影響が非常に大きいということを意識しなかったため、類似度の計算においては、各項目をすべて同じ尺度で正規化するべきかどうかは検討の余地がある。例えば、年齢はレコメンデーションにおいては極めて重要な要素であり、その意味においてはよりウェイトを掛けてみるべきであったかもしれない。要するに、第一部分と同じ結果が出て、実効性が非常に重要であるとわかる。

また、データは128人の14個の属性から成る。サンプル数が少ないので、規則性がはっきりでないという問題がある。特に、データ数が少ないため、本来であれば、より参考になる人がペアとして導き出されていない可能性が高い。

5.3 意義

芸能人にとって、自分の発信力を高めるために、自分と類似している他の芸能人と比べて、データを分析して、改善できる部分を探し出すことが重要である。ここでは、芸能人がSNSにおける呟くタイプ、ファンの性別、ゴシップ、個人能力なども発信力に影響を与えている。

広告主にとっては、すでに発信力が高い芸能人を利用する場合は、出稿するための費用もより多くかかる。しかし、もし将来的な発信力が増加することが一定の確率以上で見込めるのであれば、当面の発信力が弱いかもかもしれないが、将来性を見込むことも可能である。実際、CMや広告などではじめてブレイクする芸能人も多数おり、将来的な発信力に期待できるため、広告主としてはどのような属性を持った芸能人が発信力のポテンシャルを持っているかを見極めることは重要である。

6 関連研究

以下、本研究に関連して、有名人のソーシャルメディアにおける情報発信力に関する研究をまとめる。

6.1 有名人に関する研究

Mannらの研究[Mann et al., 2020]は、有名人の画像の多次元性を測定し、さまざまな有名人の職業にわたって一般化できる、信頼性が高く有効なスケールを開発した。EFA（探索的因子分析）、CFA（確認的因子分析）、およびSEMは、信頼性が高く有効なスケールの有名人の画像を開発するために使用されてきた。スケールには、有名人のイメージの7つの側面が含まれる。それは、社会的および倫理的責任、生活とスタイル、専門的能力、有名人の職業に対する視点（perspective towards celebrity's profession）、ファンに対する方向性、魅力的な体格と公共のイメージである。これにより、研究者や実務家は、ブランドの推奨の分野で有名人の画像研究に標準化をもたらすことができると指摘した。

ChungとCho[Chung・Cho, 2014]は、視聴者がリアリティTVやソーシャルメディアを介してメディアキャラクターとのパラソーシャル関係を構築する方法と、有名人の支持や購入意向への影響を調査した。結果は、メディアキャラクターと対話するためのリアリティTV視聴とSNSの使用が、メディアキャラクターと視聴者の間のパラソーシャル関係と積極的に関連していることを示した。次に、パラソーシャル関係は、視聴者の支持者の認識とブランドの信頼性、およびお気に入りのメディアキャラクターによって支持されたブランドの購入意向と積極的に関連していた。結果はまた、自己開示（self-disclosure）がパラソーシャル関係を形成し、有名人の支持の有効性を仲介する上で重要な役割を果たしたことを示した。

上記で言及した「パラソーシャル関係」とは、メディアを通じて、視聴者は有名人との親密なつながりを感じ、有名人を個人的に知っていると感じている状態であり、いわば一方的な関係である。漫画やアニメのキャラとの間にこの関係を築いている人もいるほどである。また、「個人的および個人的に」視聴者と会話しているメディアキャラクターは、視聴者の反応をより期待させ、したがって、パラソーシャル関係を容易に作り出すことができる。この「戦略的に管理された自己開示」は「デジタル親密さ」を生み出す可能性があり、実際には、視聴者が直接会うことができない有名人との実質的で意味のある相互作用の可能性を視聴者に提供する可能性がある。

Mulayousefの研究[Mulayousef, 2018]では、SNSでの有名人の自己開示により、ファンとのパラソーシャルインタラクション（PI）が向上することを示した。また、有名人が信頼できると見なされると、ファンとのPIが高くなる一方、有名人のジャンルとSNSサイトの選択はPIに影響を与えないと指摘した。このパラソーシャルインタラクションはHorton and Wohl (1956)によって初めて提出され、メディアパーソナリティとメディアユーザーの関係を表示し、視聴者の一方的で錯覚的な個人的関係だと説明している。また、PIが確立されると、メディアパーソナリティの価値観と考え方がメディア利用者に影響を与え始め、メディアパーソナリティを模範にするようになるという。

6.2 ファンの視点からの研究

熊本の研究[熊本, 2015]によると、新たな知識や情報の獲得を重要視している人は年齢層に関係なく多いが、芸能人や有名人とのつながりを重要視している人は若年者層に多いことや、リアルタイムに情報発信できることをメリットと考える人が多い一方で、表示されるツイートが多すぎる点や情報の真偽を判断するのが難しい点をデメリットと考える人が多いことがわかった。

Kimらの研究[Kim・Kim, 2015]は、ソーシャルネットワークサービス（SNS）ユーザーと有名人の間のパラソーシャル関係に影響を与える要因とユーザーの購入意向への影響を調査した。SNSの利用と有名人とのパラソーシャル関係は正の相関関係にある。また、ユーザー/有名人のパラソーシャル関係と有名人の評判も、購入意向との良好な関係を示している。さらに、ユーザーの性別は、有名人の評判、パラソーシャル関係、購入意向の関係に影響を与える。有名人の信頼性はその評判により大きな影響が及ぼされ、女性間の準社会的関係にプラスの影響を及ぼした。社会貢献は、男性間の準社会的関係にプラスの影響を及ぼした。

しかし、有名人の信頼性は、男性と女性の間で購入意向に大きく異なる影響を示した。有名人の評判の重要な要素としての有名人の信頼性は、男性のみの購入意向にプラスの影響を及ぼした。社会貢献は女性の購入意向にプラスの影響を与えたが、男性の購入意向にはマイナスの影響を与えた。

6.3 今後の展望

本研究では、上記のような先行文献を参考にして、有名人の属性を考慮すべきであったと考えられる。まず、属性を選択するとき、Mannら開発したスケールを参考しながら決定すれば、より全面的に有名人のイメージが含まれる可能性がある。

また、視聴者からの視点は不可欠であり、パラソーシャル関係とインタラクションの影響を重視しなければならない。視聴者にとって、有名人との繋がりという一方的な思い込みは最終の購買意欲に強い影響を与えているので、発信力にも影響していると考えられる。アンケートの中で、視聴者は有名人に対する主観的な意識も考えられなければならない。

さらに、有名人の自己開示の程度はパラソーシャル関係を作り出すための重要な要素であるので、それも発信力に影響を与えているのだと推測できる。

一方、先行研究は、データに基づく定量的な研究ではなく、少数の例に基づく定性分析がほとんどである。これに対して本研究では、主観的であっても感覚的な指標をできるだけ定量化することで、データに基づく分析を試みた。将来の研究では、関連研究で言及した点を改善する方が有効だと考えられる。

その他、データの収集について、幾つの問題が存在している。

まず、2020年4月のデータを採用したが、属性の選択について実効性を注意しなかった。4月に活躍していたか、出番の頻度はどうであったか、どこで活躍していたかなどの変数を入れる方がいい。

また、データの収集が困難かつかなり主観的なもので、目的変数（発信力）により大きな影響を与える属性を先行研究によって推測し選択すれば、収集したデータの信頼度を高めることができる。

さらに、時間の制限は4月だけではなく、より長い時間、例えば一年のデータを使った方が平均的な水準を分析することができる。また、客観性を持たせるためには、各説明変数の評価においては大規模

なアンケートを取ることも有効な方法である。あるいは、SNSのロコミを大量に一括して分析できれば、より正確な分析が可能になる。しかし、一般にSNS のテキスト情報を大量に取得することは、規約上の問題もあるうえ、技術的にも課題がある。例えば、テキスト情報を解析する自然言語処理技術は、新聞記事や小説などの文法的に正しい文章に対してはかなりの精度で分析することができるものの、SNS 等の書き込みデータに関しては精度が十分とは到底いえない状況である。SNSというプラットフォームで書かれたテキストは人や時代によって異なり、複雑である。特にSNSでは複雑な感情が微妙な表現で表されていると考えられるが、現状の感情分析、極性分析でSNS 上の文章を対象にするのは現状では困難であると考えられる。

参考論文

- Mann, B. J. S., Parmar, Y., & Ghuman, M. K. (2020). A New Scale to Capture the Multidimensionality of Celebrity Image. *Global Business Review*, 0972150920919599.
- 熊本忠彦.(2015).アンケート調査に基づくTwitterユーザとツイート印象の分析. 第7回データ工学と情報マネジメントに関するフォーラム講演論文集,F5-4.
- Chung, S., & Cho, H. (2014, June). Parasocial relationship via reality TV and social media: its implications for celebrity endorsement. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (pp. 47-54).
- Mulayousef, A. (2018). Following Celebrities on Social Networking Sites: The Role Of Parasocial Interaction, Self-Disclosure, Trustworthiness, and Time Spent on Sns.
- Kim, H., Ko, E., & Kim, J. (2015). SNS users' para-social relationships with celebrities: social media effects on purchase intentions. *Journal of Global Scholars of Marketing Science*, 25(3), 279-294.

参考にした URL など

フォロワーが最も多いWeiboアカウント：

<https://zh.wikipedia.org/wiki/%E5%85%B3%E6%B3%A8%E4%BA%BA%E6%95%B0%E6%9C%80%E5%A4%9A%E7%9A%84%E6%96%B0%E6%B5%AA%E5%BE%AE%E5%8D%9A%E5%B8%90%E5%8F%B7>

KOLRANK：

<https://www.kolrank.com/>

決定木分析（ディジジョンツリー）とは？概要や活用方法、ランダムフォレストも解説：

<https://cacao.com/ja/blog/what-is-decision-tree/#i>

WEKAと樹木モデル：

https://www1.doshisha.ac.jp/~mjjin/R/Chap_20/20.html

WEKA3 – The workbench for machine learning:

<https://www.cs.waikato.ac.nz/ml/weka/>

決定木の最初の

百度ホームページ：

<https://www.baidu.com/>

知乎ホームページ :

<https://www.zhihu.com/>

微博ホームページ :

<https://www.weibo.com/>