

ロジスティック回帰を用いた MC バトルにおける MC 推定

吉實隼稀 但馬康宏

岡山県立大学大学院 情報系工学研究科システム工学専攻

sk621062@cse.oka-pu.ac.jp

概要

本研究ではロジスティック回帰を用いて MC バトルの歌詞を解析し、MC 推定、各 MC の個性を捉えることを目的とした。異なるスタイルを持った7人の MC を研究対象とし、検証を行った。地名を発した数の対戦相手との差やアンサー単語の平均回数の対戦相手との差が MC の個性を推し量るに有効な説明変数であることが示され、各 MC の回帰係数ベクトル w がなす角度を比較した際にそれぞれ差異が見られ、MC ごとの個性を捉えられた。

1 はじめに

MC バトルは元々ラッパーの生活の延長上に存在していた。ラッパーの余興、テクニクを魅せる手段の一つであり、HIPHOP 文化に含まれる。HIPHOP は 1970 年代のアメリカで生まれた文化であり、ラップ、ブレイクダンス、グラフィティ、DJ プレイなどの要素が存在している。ラップは歌唱法の一つであり、似た言葉や語尾が同じ言葉を繰り返す韻を踏むのが特徴的で口語に近い抑揚をつけて発声する。ラップの要素には「内容」、「フロウ(リズム、韻)」、「話し方」などが含まれる。ラップの手法にフリースタイルラップというものがある。ラッパーが即興で韻を踏みながらラップをする表現技法である。やがて、MC バトルの規模が大きくなり、バトルの舞台とヒップホップシーンが分離し始める。しかし、MC バトルは依然として広く浸透しておらず、MC バトルの魅力は HIPHOP に馴染みがないと理解し難い側面がある。そこで本研究では MC バトルの歌詞を解析し、各 MC の特徴を抽出することで初見の人でもどこに注目すればいいのかが明確になると考えた。そこで、本研究ではロジスティック回帰を用いて MC バトルの歌詞を解析し、MC 推定、各 MC の個性を捉えることを目的とした。

2 関連研究

歌詞やテキストデータの分析、分類を行った研究

は多数存在している。

秋元ら[1]による研究では Word2Vec や fastText を用いて特定作詞家による楽曲分類、TF-IDF を用いて作詞家間の対比を行っている。各作詞家の語句選択における特徴を掴む手がかりを得た。山下・萩野ら[2]による研究では品詞に着目し、ユーザの心理的状況に適する歌詞の選曲を支援する方法の一つとして、歌詞から特徴語と名付けた動詞と形容詞を抽出している。抽出された語を代表語とし、その代表語とユーザ気分の関係をモデル化する方法が提案されている。篠井ら[3]による研究では Wikipedia 記事と Web の感想記事のデータを用いて、各アーティストの特徴量を算出、可視化している。得られた特徴量に対しジャンル分類、評価結果が示されている。浅野、山下、松林ら[4]による研究ではロジスティック回帰やランダムフォレストを用いて、Twitter のツイート情報から高専生か否かの分類が行われた。中村、山口ら[5]による研究では SVM 及び DNN を用いて主観文書を学習及び分類して分類結果の有効性について考察されている。山田、樫山[6]による研究ではアプリケーションレビューのデータセットを使用し、教師あり学習によるアプリケーションレビューの分類が行われた。分類器の構築に用いる特徴量として Bag of Words、感情分析の値が利用されていた。坂井、谷岡、松浦、後藤、和田ら[7]による研究ではロジスティック回帰分析を用いて野球の打撃能力評価指標の提案が行われている。グラウンド内の情報を説明変数、打撃結果を目的変数として回帰モデルが作成された。

自然言語処理や機械学習を駆使して多岐に渡る分野の研究が行われてきた。本研究でも上記の関連研究を参考に研究を行った。

3 MC バトルの詳細

本研究の主題である MC バトルについて説明する。MC バトルに関わる人は主に、MC、DJ、司会、観客である。司会はバトルの進行役であり観客の反応を伺ってバトルの勝敗を決定する。DJ はバトル中に音源を流す役割を担っている。バトル前にじゃんけん で先攻、後攻を決める。バトルは後攻の方が有利で

あるため先攻はビートと小節の選択権がある。選択権を持つ MC はその場にいる複数人の DJ が用意してきたビートを視聴し、ビートを選択する。小節数やターン数は大会によるが、8小節2ターン、8小節3ターン、8小節4ターン、16小節2ターンが主である。バトル中は決められた小節内でラップする。相手のターン中にラップをしてはならない。また、言葉だけで闘うため、基本的に相手の体に触れてはいけない。バトル終了後、司会が観客に勝利したと思う方に手と声を上げるように促す。拮抗した場合は司会者の判断で先攻、後攻入れ替わって再度バトルを行う。

3.1 審査基準

バトルには明確な審査基準は存在していない。どちらが格好良いか、面白いかな、ラップが上手いか、印象に残ったか、即興性が高いかなど、注目している点は人それぞれである。観客を盛り上げるために主に以下の表現技法がある。

- ・アンサー
- ・パンチライン
- ・ヴァイブス
- ・ライム
- ・フロウ

これらの表現技法を本稿ではスキルと呼ぶことにする。フロウは歌詞のみでは検出できないため、研究対象外とした。

4 スキル

3.1 節の表現技法について説明する。

「アンサー」は相手のブレた論点にツッコミを入れ、ディスられた内容を否定して自分自身の考えを言うことである。記憶に残るのが「パンチライン」であり、誰も使用したことがない独特なキーワードセンスが求められる。「ヴァイブス」は観客を盛り上げて観客を味方につける一つのスキルであり、声のトーンや言葉、声の張り上げ等で観客に情熱や気合を伝える。語尾が名詞ではなくなる特徴がある。また、早口になって1小節あたりに含まれる言葉数が多くなる。「ライム」は単語の母音を重ね合わせたもので韻を踏むとも言う。「フロウ」はビートに対するアプローチの種類、歌いまわし、抑揚の付け方、ラップの展開など、いろんな解釈がある。[\[8\]](#)

各 MC が特化しているスキルは三者三様であり、各 MC には個性（スタイル）を感じられる。

5 使用したデータ

本研究で用いたデータは[\[9\]\[10\]\[11\]\[12\]](#)等がYouTubeにアップロードしているバトル動画とした。

基本的にバトルは MC 同士、1 vs 1 だが、3 vs 3 の団体戦も存在する。本研究では団体戦は対象としない。バトルの歌詞データは合計 300 個用意した。対象とした MC は「DOTAMA」、「R 指定」、「NAIKA MC」、「ニガリ」、「MU-TON」、「晋平太」、「呂布カルマ」の7名。それぞれ違うスタイルを持ち合わせた MC である。データの内訳を付録の[表 1](#)に示す。歌詞情報は動画の字幕や[\[13\]\[14\]](#)を参考に作成した。歌詞データの形式は付録の[図 1](#)に示す。

6 分析方法

分析には python ライブラリである scikit-learn のロジスティック回帰分析を用いた。ロジスティック関数に含まれる z は総入力である。

$$\varphi(z) = \frac{1}{1 - e^{-z}}$$

z は回帰係数とサンプルの特徴量の線形結合であり、以下のように計算ができる。

$$z = \mathbf{w}^T \mathbf{x} = w_0 + w_1 x_1 + \dots + w_m x_m$$

学習された回帰係数ベクトル \mathbf{w} と説明変数 x_i の値が与えられた時、目的変数のクラス分類を行うことが可能である。[\[15\]](#) 本研究では目的変数を MC とし、説明変数は4節のスキルと MC の個性を表すと考えられる要素に基づいて以下の11個とした。

- ・ライムスコア: x_1
- ・ライムスコアの対戦相手との差: x_2
- ・アンサー単語の平均数の対戦相手との差: x_3
- ・アンサーした回数の対戦相手との差: x_4
- ・パンチラインスコア: x_5
- ・パンチラインスコアの対戦相手との差: x_6
- ・ヴァイブス単語数の対戦相手との差: x_7
- ・地名を発言した回数の対戦相手との差: x_8
- ・名詞の割合: x_9
- ・総単語数の対戦相手との差: x_{10}
- ・総発音数の対戦相手との差: x_{11}

本研究では二値分類器を多クラス問題に拡張するための手法である一対他 (One-vs-Rest:OvR) [\[15\]](#) を使って MC ごとに分類器の性能を検証した。また、性能の評価には K 分割交差検証法を使用した。本研究では $K=5$ とし、スコアの平均値を推定結果とした。

6.1 ライム: x_1, x_2

[図 2](#) はライムスコアの算出方法であり、本研究では脚韻のみを考慮した。

	隙間風(uiiae)	踏み開かせ(uiase)	スコア	合計スコア
①	u	uiiae	1	1
	ui	uiiae	2	3
	uia	uiiae	3	6
	uiaa	uiiae	4	10
	uiaae	uiiae	5	15
②	i	uiaae	1	16
	ia	uiaae	2	18
	iaa	uiaae	3	21
	iaae	uiaae	4	25
③	a	uiaae	1	26
	aa	uiaae	2	28
	aac	uiaae	3	31
④	a	uiaae	1	32
	ae	uiaae	2	34
⑤	e	uiaae	1	36

図 2 ライムスコアの算出例

6.2 アンサー : x_3, x_4

アンサー単語の平均数の対戦相手との差とアンサーした回数の対戦相手との差の算出方法を図 3 に示す。本研究では対戦相手が前のターンで使用した単語と同じ単語を使用することをアンサーとした。対象の単語は名詞のみとした。

[晋平太]		[KBD]	
単語	アンサー回数	単語	アンサー回数
ゼログレバンナ	1	俺	1
俺	1	4番	1
時代	1	ゴリラ	1
ゲノム	1	王者	1
お前	2	突破	1
こと	1	千羽鶴	1
何	1	アンストッパブル	1
KBD	1	何	1
合計 9		合計 8	
アンサー数の差 = $9 - 8 = 1$			
1単語あたりのアンサー回数の平均 = $1.125 - 1 = 0.125$			

図 3 アンサー回数の差, 平均の算出例

6.3 パンチライン : x_5, x_6

パンチラインスコアは TF-IDF を用いて算出している。TF 値 $tf(t, d)$ は各バトル d の各単語 t として算

出。IDF 値 $idf(t)$ は全バトルの各単語 t として算出。また、単語 t は名詞のみを対象としている。バトルにおける TF-IDF 値が上位 10 個の単語を合計したものをパンチラインスコアとした。

6.4 ヴァイブス : x_7

本研究で対象としたヴァイブス単語を示す。
「だ」「ぜ」「たい」「だぜ」「ねえ」「だろ」
「だぞ」「だよ」「じゃねえか」「ねえよ」
地名の抽出は形態素解析 janome を使用し、「地域」カテゴリに該当した単語とした。

7 結果

本研究では全バトルデータ 300 個の内、240 個を学習データ、60 個をテストデータとして検証を行った。また性能評価の指標は F1 値とした。F1 値を評価指標にした理由は、一体他手法において、One (該当 MC) の割合に比べて Rest (該当 MC 以外) の割合が非常に大きく、全ての予測を Rest にすることで、スコアが高くなってしまいう評価指標である Accuracy よりも良好なモデルを選択できると考えたからである。7 人それぞれの MC の性能の結果を付録の表 2 に示す。

次に、F1 値が最大時の各 MC の回帰係数を降順に並べたものを表 3~9 に示す。これは数値が高い係数ほど目的変数を決定づける影響が大きいと解釈できる。また、F1 値が最大時の各 MC の回帰係数ベクトル w がなす角度を付録の表 10 に示す。

8 考察

表 2 より「DOTAMA」「R 指定」「ムートン」の F1 値が 0.6 以上となり、対象 MC の中で比較的良好な結果を示した。表 3~9 から各 MC の考察をする。

「DOTAMA」はアンサー単語の平均回数の対戦相手との差が最も高い値を示しており、著者の印象通りであった。「R 指定」は多くのスキルを持ち合わせている MC であり、ヴァイブス、ライムに関する項目が上位であるため、F1 値が高くなったと考えられる。また、実際にバトルで地名を発することも多く、それが結果に現れていた。一方、パンチラインスコアの値が最も低い結果となっており、実際のバトルでの印象と異なるため、パンチラインスコアの算出方法を改善する必要がある。「ムートン」はフロウとパンチラインを得意とする MC であり、地名を発することも多く、結果にも現れていた。「ニガリ」

「晋平太」「呂布カルマ」はF1値が0.5以下を示す結果となった。「ニガリ」はフロウを得意とするタイプであり、フロウを表現する説明変数がなかったためF1値が低くなったと考えられる。「晋平太」はライムを得意とするMCだが、ライムスコアとライムスコアの対戦相手との差の値が小さい結果を示した。これは、本研究では脚韻のみを対象としており、頭韻や文中の韻を考慮できていないからと考えられる。「呂布カルマ」はパンチラインやアンサーを得意とするMCだが、パンチラインスコアの差が負の値を示していることからF1値が低い結果になったと思われる。また、呂布カルマは強烈なディスリスペクトを武器にしているため、ディスリスペクトを表現する説明変数を加える検討が必要である。

表10より、各MCの重みベクトル w がなす角度が最低値26.3°、最高値61.1°を示しており、本研究で対象としたMC7人それぞれ個性があることが示された。

9 まとめ

本研究ではロジスティック回帰を用いてMCバトルの歌詞からMC推定、MCごとの個性を捉えることを行った。地名を発した数の対戦相手との差やアンサー単語の平均回数の対戦相手との差がMCの個性を押し量るに有効な説明変数であることが示された。今後は推定の精度が向上するよう、各種スコアの算出方法の改善、ディスリスペクト等の有効な説明変数を新たに加えることやロジスティック回帰分析以外の分析方法による検証が必要である。

表3 DOTAMA

w_3	0.210225
w_1	0.020439
w_2	0.001485
w_9	-0.00017
w_{11}	-0.01266
w_{10}	-0.02183
w_7	-0.02464
w_4	-0.02632
w_5	-0.10839
w_6	-0.20259
w_8	-0.23187

表4 NAIKA MC

w_{11}	0.004866
w_2	-0.00672
w_{10}	-0.01005
w_1	-0.02214
w_4	-0.08291
w_9	-0.09516
w_5	-0.1053
w_3	-0.11205
w_7	-0.12504
w_6	-0.18351
w_8	-0.27695

表5 R指定

w_8	0.038483
w_7	0.026391
w_1	0.025577
w_{11}	0.018532
w_2	-0.00449
w_9	-0.00567
w_3	-0.02606
w_4	-0.02855
w_{10}	-0.05764
w_4	-0.05904
w_5	-0.39409

表6 ニガリ

w_4	0.07665
w_9	0.072778
w_6	0.060022
w_3	0.041122
w_{10}	0.007714
w_2	0.007588
w_{11}	-0.00301
w_7	-0.00978
w_1	-0.01644
w_8	-0.02151
w_5	-0.32709

表7 ムートン

w_6	0.409044
w_8	0.310527
w_5	0.172197
w_9	0.123374
w_7	0.051614
w_1	0.007915
w_2	-0.00174
w_{10}	-0.01101
w_{11}	-0.0154
w_4	-0.03477
w_3	-0.54254

表8 晋平太

w_3	0.378351
w_5	0.274882
w_8	0.082565
w_4	0.033094
w_{10}	0.024409
w_7	0.006827
w_2	0.001677
w_1	0.000894
w_9	-0.01216
w_{11}	-0.01412
w_6	-0.18754

表9 呂布カルマ

w_3	0.317477
w_5	0.123233
w_{10}	0.033993
w_2	0.010452
w_{11}	-0.01269
w_4	-0.02179
w_7	-0.04845
w_1	-0.05453
w_9	-0.07704
w_8	-0.09048
w_6	-0.32453

参考文献

1. 秋岡明香, 意味ベクトルを用いた J-POP 歌詞の文体分析, 情報処理学会, 第 82 回全国大会講演論文集, 2020 巻 1 号, p313-314, 2020 年
2. 山下佑子 萩野晃大, 楽曲のサビを構成する動詞・形容詞の代表語への集約と気分語の関係の分析-歌詞全体とサビ部分の比較を通して, 情報処理学会, 研究報告音楽情報科学 (MUS), 2015-MUS-107, 57 号, p1-6, 2015 年
3. 篠井暖, ArtistVector : Web 文書分散表現によるアーティスト特徴量獲得, 情報処理学会, 研究報告自然言語処理 (NL), 2018-NL-236, 3 号, p1-7, 2018 年
4. 浅妻佑弥 山下晃弘 松林勝志, SNS 上への発言の特徴分析に基づくユーザの属性推定, 情報処理学会, 第 80 回全国大会講演論文集, 2018 巻 1 号, p541-542, 2018 年
5. 中村鴻介 山口実靖, 機械学習による主観文書分類結果の解釈性の付与に関する-考察, Web とデータベースに関するフォーラム, 情報処理学会, WebDB Forum 2019 論文集, p17-20, 2019 年
6. 山田侑樹 樋山淳雄, fastText を用いたアプリケーションレビューの分類と局所的な説明の付与による考察, 情報処理学会, ソフトウェアエンジニアリングシンポジウム 2020, p219-225, 2020 年
7. 坂井衛 谷岡広樹 松浦健二 後藤田中 和田智仁, ロジスティック回帰分析を用いた打撃能力評価指標の提案, 情報処理学会, 研究報告数理モデル化と問題解決 (MPS), 2020-MPS-127, 8 号, p1-6, 2020 年
8. 晋平太, フリースタイル・ラップの教科書 MC バトルははじめの一步, イースト・プレス, 2016 年
9. 戦国 MC battle, <https://youtube.com/senritumc>
10. ikftv , <https://www.youtube.com/ikftv>
11. 凱旋 MCbattle, https://youtube.com/channel/UCe_EvY8GrvYgx8PbwRBc75g
12. UMB, <https://www.youtube.com/umbofficial>
13. MC バトルリリック <https://ameblo.jp/tempest156/>
14. MC battle リリックまとめ, <http://mc-battle.info/>
15. Sebastian Raschka, Python 機械学習プログラミング 達人データサイエンティストによる理論と実践, インプレス, 2016 年

A 付録

表 1 データの内訳

MC	データ数	割合
DOTAMA	30	10.0%
NAIKAMC	30	10.0%
R指定	30	10.0%
ニガリ	39	13.0%
ムートン	46	15.3%
晋平太	57	19.0%
呂布カルマ	68	22.6%
合計	300	100%

<p>[R指定]</p> <p>1 ビートでは逃げない</p> <p>2 しっかりと魂のぶつかり合いがしたい</p> <p>3 なあ本戦で何度だって</p> <p>4 戦うチャンスはあったのに今日が初めてだぜ</p> <p>5 なあ輸入さんいろいろあって</p> <p>6 戦えなかったが今日せっかくのチャンス</p> <p>7 魂のぶつかり合いとやら</p> <p>8 あんたの言葉の重みをぶつけてくれよ</p> <p>[輸入道]</p> <p>まずは絶対お前はネタ仕込んで</p> <p>来ると思ってたけど意外と意気込んでる</p> <p>オッケーいいねそのバイブスのままやろうぜ</p> <p>俺だってお前のことぶっ殺したかったぜ</p> <p>yeah ライブやった後なのに</p> <p>こいつ全然声枯れてねえ バイブス足らねえ</p> <p>マイク掴んで何が言いてえ</p> <p>それをちゃんと最後に伝えるのはどっちか見極めてくれ</p>	<p>1小節一行区切り 8小節or16小節</p>	<p>1ターン×2 or 3 or 4 大会によって異なる</p>
--	-------------------------------	---------------------------------------

図 1 テキストデータの形式

表 2 K 分割交差検証の結果

評価指標	DOTAMA	NAIKAMC	R指定	ニガリ	ムートン	晋平太	呂布カルマ
F1値	0.659	0.537	0.616	0.465	0.628	0.454	0.491

表 10 各 MC の回帰係数ベクトル w がなす角度

	DOTAMA	NAIKAMC	R指定	ニガリ	ムートン	晋平太	呂布カルマ
DOTAMA		47.5	39.1	34.7	45.7	26.3	43.4
NAIKAMC			58.2	51.2	61.1	45.0	43.7
R指定				46.6	50.7	49.9	67.0
ニガリ					39.3	33.1	41.7
ムートン						37.0	59.9
晋平太							30.0
呂布カルマ							