

平成 29 年 12 月 28 日

各位

会社名 株式会社ソケット
代表者名 代表取締役社長 浦部 浩司
(証券コード：3634)

**ソケット、昨年予測的中した
正解率 79%の紅白勝敗予測モデルで今年も勝敗を予測**

株式会社ソケット（本社：東京都渋谷区、代表取締役社長：浦部浩司、以下「ソケット」）は、昨年的中した紅白歌合戦の勝敗予測において、今年も紅白勝敗を予測、そのレポートを公開いたしました。

本予測モデルは、オリジナルの「感性メタ（※）」を教師データとして、2 部制がはじまった第 40 回 1989（平成元年）以降から、本年度第 68 回 2017（平成 29 年）の紅白歌合戦で、紅白それぞれに分かれて披露された楽曲を対象に歌詞データから“感情”を分析、予測しています。

（※）感性メタとは、特に日本の繊細な感性表現にこだわり、たとえば音楽の歌詞においては全てを読み、内容の真意を理解してから付与するなど機械には出来ない膨大な作業によって生まれた国内唯一のエンターテインメントメタ(ゲノム)データです。

紅白歌合戦の曲目の歌詞データから感情を分析、勝敗を予測 2 年目

さて、昨年行った感情分析からの紅白勝敗予測ですが、序盤からの白組優勢の経過を覆し、最終的に見事紅組勝利ということで予測が的中しました。

紅白といえば、数時間にも及ぶ対抗形式で衣装はじめ、あらゆる演出を含めて歌や演奏を披露、そのパフォーマンスに対する勝敗を決めるものですが、ソケットの予測分析モデルは、楽曲の歌詞データから算出した感情スコアと過去の勝敗を元に、他の相関、因子となりうる変数を入れない非常にシンプルなものです。それでも昨年の予測モデルでは正解率 96%という精度で、かつ昨年の初回予測も正解しました。シンプルなモデルでこれだけの予測ができる感情スコアのさまざまな可能性を検証していく上で、今年も昨年同様、この方針を変えずに挑戦してみようと思います。

詳しくは、昨年のレポート『紅白歌合戦の曲目の歌詞データから感情を分析、勝敗を予測してみる』をご参照ください。

http://www.sockets.co.jp/kansei/kansei_report01.html

紅白歌合戦の歌詞データの感情をスコアリング

では早速、昨年のおさらいも交えてご紹介していきたいと思います。

まず予測分析の対象データは、昨年同様、2部制がはじまった第40回1989（平成元年）以降から、今年2017年第68回2017（平成29年）の紅白歌合戦としています。

対象期間：第40回1989（平成元年）～第68回2017（平成29年）

※紅組・白組の垣根を越えて披露された楽曲は対象外とする

※『紅白エディション』はオリジナル楽曲の歌詞を解析対象とする

※メドレーは当日まで曲内容が分からないため対象外とする

全29回、全対象楽曲1,396曲　うち紅組712曲、白組684曲

メドレーなどの除外を含む対象期間全披露楽曲1,509　分析対象率93%

上記対象となる各楽曲の歌詞データを、ソケット感情分析エンジン（特許出願中）で10種類の感情スコアを算出、それを年代・紅白の組ごとに数値を集計、個々の感情スコア値ごとに正規化します。そして、ロジスティック回帰分析から構築したソケット紅白勝敗予測モデルで予測する（※1）という流れになります。

（※1）ロジスティック回帰分析とは、ある現象の発生する確率を、その現象の起因を説明するために観測された変数群によって説明をするための分析手法。今回の目的では、勝ち（1.0）か負け（0.0）かを、その現象の規制を説明するために観測された変数群としてソケット独自の感情スコアを利用して予測。このとき 勝ち（1）／負け（0）という値で得られている変数は1もしくは0という2値のいずれかで示され、この1もしくは0という値を、他の説明変数が決定付けている、もしくは影響を与えているとみなして分析を行ない、数式を構築します。

概ね、昨年とやり方は変わりませんが、この1年間で、ソケット感情分析エンジンはあらゆる要素技術や研究結果などを元に、日々精度アップに向けたチューニングを行い、進化してきました。どう進化したのか？を簡単に要約すると、歌詞の語彙の学習により、以下ポイントが主に1年前と比べ精度アップしています。

1. 感情コーパスの補正
2. 未知語のスコア推定
3. 連語のスコア推定

ここは難しい説明よりも、実際にどのように変わったかを見ていただいた方がわかりやすいと思いますので、今年の紅白トリを務める、紅組の石川さゆり「津軽海峡・冬景色」と白組大トリ、ゆずの「栄光の架け橋」を例に見ていきたいと思います。両楽曲は、過去にも披露されており、昨年もスコア算出していましたので、昨年算出したスコアと今年算出したスコア結果を照らし合わせてみました。

センテンスごとの感情スコア(1番の歌詞引用)

♪津軽海峡・冬景色 / 石川さゆり

no	sentence	感情分析 エンジン	感情分析										
			幸福・恋愛	好き・喜び	安らぎ	哀しみ	昂ぶり	希望	不安・恐怖	怒り・苛立ち	嫌い・不愉快	もどかしさ	
1	上野発の夜行列車 おりた時から	従来											
		進化版											
2	青森駅は 雪の中	従来											
		進化版											
3	北へ 帰る 人の 群れは 誰も 無口で	従来											
		進化版											
4	海鳴りだけを きいて いる	従来											
		進化版											
5	私も ひとり 連絡船に 乗り	従来											
		進化版											
6	ごえそいな 駒 見つめ 泣いて いました	従来											
		進化版											
7	ああ 津軽海峡 冬 景色	従来											
		進化版											

センテンスごとの感情スコア(1番の歌詞引用)

♪栄光の架け橋 / ゆず

no	sentence	感情分析 エンジン	感情分析										
			幸福・恋愛	好き・喜び	安らぎ	哀しみ	昂ぶり	希望	不安・恐怖	怒り・苛立ち	嫌い・不愉快	もどかしさ	
1	誰にも 見せない 涙が あった 人知れず 流した 涙があった	従来											
		進化版											
2	決して 平らな 道では なかった けれど 確かに 歩んで 来た 道だ	従来											
		進化版											
3	あの時 想い 描いた 夢の 途中で 今も	従来											
		進化版											
4	何度も 何度も あきらめ かけた 夢の 途中	従来											
		進化版											
5	いくつもの 日々を 越えて 辿り着いた 今がある	従来											
		進化版											
6	だから もう 迷わずに 進めばいい	従来											
		進化版											
7	栄光の 架け橋へと	従来											
		進化版											

こちらは、センテンスごとに感情スコアを算出した結果になります。上段が昨年算出したスコア、下段が進化した今年の感情スコアを元に、センテンスごとに感情10種がどのような構成割合となっているのか表したのになります(昨年と今年のスコアリングでは、正規化の処理なども異なりますので、スコア実数値自体での比較はできませんので、省いています)。

赤枠内を見ていただくと分かりやすいかと思いますが、昨年は感情に関するフレーズを抽出し、10種ごとに分類、つまり感情を判別するダイレクトなワードからのスコアリングでしたが、主に先ほどあげた3点の精度アップにより、いわゆる“行間を読む”ような、複雑な“感情”を読み取ったスコアリングに進化しました。その結果、平坦だったスコア値が今年は直接的に感情に関するフレーズがなくても感情スコアを算出するようなエンジンになっています。また、「希望」や「不安」、「哀しみ」や「安らぎ」など、人の入り混じった感情をより詳細に理解することで人に寄り添えるエンジンを目指しています。こうしてセンテンス単位でスコアリングした結果を、楽曲単位、さらに年ごとの紅白別で集計、正規化していきます。ちなみに楽曲単位で比較してみると下記のようになります。

楽曲単位の感情スコア

♪津軽海峡・冬景色 / 石川さゆり

感情分析エンジン	幸福・恋愛	好き・喜び	安らぎ	哀しみ	昂ぶり	希望	不安・恐怖	怒り・苛立ち	嫌い・不愉快	もどかしさ
従来	■	■	■	■	■	■	■	■	■	■
進化版	■	■	■	■	■	■	■	■	■	■

楽曲単位の感情スコア

♪栄光の架け橋 / ゆず

感情分析エンジン	幸福・恋愛	好き・喜び	安らぎ	哀しみ	昂ぶり	希望	不安・恐怖	怒り・苛立ち	嫌い・不愉快	もどかしさ
従来	■	■	■	■	■	■	■	■	■	■
進化版	■	■	■	■	■	■	■	■	■	■

いかがでしょうか？よりきめ細やかに感情を読み取りスコア化することができるようになったソケット感情分析エンジンです。

最新の感情分析エンジンで算出した感情スコアを分析、そして予測

それでは、昨年算出していた過去分の対象楽曲もすべて感情スコアを算出し直す形にして、1から分析、予測していきます。

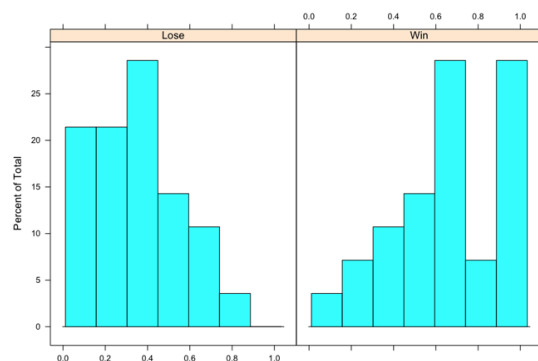
分析手法は昨年と同様です。ロジスティック回帰分析の結果を元にモデリング後、70%のサンプリングをして、残り30%の評価を1000回行い、分布を確認しました。

AUC (※2) は、0.796 でした。

(※2)

ROC 曲線 (検査の性能を2次元のグラフに表したもの) を作成した時に、グラフの曲線よ

参考:
全データを使いモデルを作成→モデルの出力のヒストグラム (正解の勝ち、負け別)



り下の部分の面積を AUC (Area Under the Curve) と言います。AUC は 0 から 1 までの値をとり、値が 1 に近いほど判別能が高いことを示します。判別能がランダムであるときには、AUC = 0.5 となり、0.5 以上であれば、2 値の判別において、精度が出ている、ということになります。

ちなみに昨年は「哀しみ」の値が高いと勝つ要因である影響度が高い、という結果でしたが、今年は、「怒り・苛立ち」次いで「嫌い・不愉快」「もどかしさ」の値が高いと負ける要因になる、という結果がでました。

それでは、今年の予測モデルに当てはめた時の、勝敗実績の正解率を見てみましょう。

2017予測モデルでの予測結果	予測件数	○正解	×はずれ	正解率
紅が勝つと予測	12	8	4	67%
白が勝つと予測	16	14	2	88%
対象期間合計	28	22	1	79%

今回対象とした過去 28 回のうち、予測モデルに当てはめたときに、

- ・ 紅が勝つ予測：12 件
○紅が実際に優勝：8 件 ×白が実際に優勝：4 件 ⇒正解率 67%
- ・ 白が勝つ予測：16 件
○白が実際に優勝：14 件 ×紅が実際に優勝：4 件 ⇒正解率 88%

全 28 回分トータルで、正解率 79%という結果になりました。

最後にいよいよ予測モデルから算出された今年の理論値の発表です。

紅白で比較して高い値の方が「勝利する」となります。

つまり、今年の紅白勝敗予測は、“白組が勝利”と予測します！

果たして昨年に続き、2 連勝なるでしょうか？

2017年	理論値
紅組	0.006
白組	0.021

2017年 白組の勝利と予測！

最後に

ソケットではこの1年、紅白勝敗予測にはじまり、連続ドラマ視聴率予測等々…、ソケットのメタデータや感情分析エンジンを活用したあらゆる分析、予測をご紹介してきましたが、このほかにもご紹介できない事例や分析結果などたくさん取り組みをしてきました。ソケットの感性メタを活用した予測分析は、主に、レコメンド、パーソナライズド、また調達支援や制作、マーケティング支援など多種多様に渡りますが、これらのキーはいずれも“人の感性”であり、そこをいかに理解し、仕組み化できるか。まだまだソケットの感性メタ、感情分析エンジンの活用範囲は広がっていくことと思います。

さらに、感情分析エンジンにとどまらず、ソケットではAI技術における研究開発にも力を入れています。たとえば、今年～来年以降引き続き注目されるであろうAIスピーカーに関連して、「ちょっといい感じの曲お願い」など、人の感情や感性に寄り添う感性AI実用化も目指しています。

また、ブロックチェーンの音楽業界への変革の波を起こすべく、ソケットのテクノロジーで牽引していきたいと思っています。

エンターテインメント×テクノロジー分野の可能性や市場を新たに拡大していけるような有用性のあるデータ・ドリブンな取り組みで、来る2018年もソケットはさらに突き抜けていきたいと思っています。

今後もソケットの技術研究開発にご期待ください。

「ソケットレポート」の記事のご利用、また共同研究などに関するご相談はこちらのお問い合わせフォームよりご連絡ください。

<http://www.sockets.co.jp/inquiry/contact/policy.html>

●公開日

平成29年12月28日（木）

●ソケットレポート

http://www.sockets.co.jp/kansei/kansei_report07.html

株式会社ソケット：<http://www.sockets.co.jp/>

株式会社ソケットは、平成12年6月に設立された「人の気持ちをつなぐ」という事業目的を持ったデータベースサービス会社です。現在、KDDI株式会社、株式会社NTTドコモ、カルチュア・コンビニエンス・クラブ株式会社などへの音楽や映像、書籍などを中心とした検索・レコメンド・ストリーミング・データ提供・アナリティクスなどのデータ

関連サービスを行っております。

本リリースに関するお問い合わせ先：

株式会社ソケット

〒151-0051 東京都渋谷区千駄ヶ谷 4-23-5 JPR 千駄ヶ谷ビル 3F

担当：コーポレート本部 小田嶋

Tel：03-5785-5518 Fax：03-5785-5517

E-mail：ir@sockets.co.jp

以上