

テキストマイニングとネットワーク分析 を用いた映画評価の要因分析

小川 哲司

Factor Analysis of Movie Evaluation by Using Text Mining and Network Analysis

OGAWA, Tetsuji

名古屋経済大学経営学部准教授

要旨： レビュー情報には消費者の潜在的なニーズなどが含まれるため、マーケティング活動などに積極的に活用しようとする動きが始まっている。本研究では、映画レビューの名詞と形容詞に注目したテキストマイニングを通じて、映画評価に影響を及ぼす要因を明らかにすることを目的とした。「鬼滅の刃」と「君の名は。」を取り上げ、頻度分析と共起ネットワークによるネットワーク分析を行った。本研究の成果として、映画作品の評価において中心性の高い要素と構造を明らかにした。さらに、高評価に繋がる共通要因として、映像や音楽による演出と、ストーリーの構成や展開である点を明らかにした。

キーワード： 映画レビュー, 形態素解析, テキストマイニング, 頻度分析, ネットワーク分析

1. はじめに

近年、インターネットの利用拡大や情報通信技術の発展などによって、様々な人・モノ・組織がネットワークにつながり、コンピュータの処理能力向上などの技術革新も相まって、膨大なデジタルデータがネットワーク上で生成、流通、蓄積されるようになった。このような膨大な量のデジタルデータを「ビッグデータ」と呼び、各企業や組織において、新たなサービスやビジネスモデルを生み出したりするために、積極的に活用しようとする動きが始まっている。

デジタルデータは、大きく構造化データと非構造化データに分類することができる。構造化データとは、販売管理や在庫管理などのデータベースにおけるレコードのように、データと属性がセットになったもので、データ形式に一定のルールが存在する。非構造化データは、データの構造にルールが無く、自由に記述されたデータのことであり、テキスト、画像、音声データなどが該当する。特に Web サービスの発展により、SNS やブログなどのデジタルコンテンツから、日々膨大なテキストデ

ータがインターネット上で生成されている。このような状況から、いまやビッグデータの約 8 割を非構造化データが占めているとも言われている。

これまで消費者の動向やニーズを把握しようとする際は、構造化データを収集して、統計的に分析することが一般的な手法であった。しかしながら、構造化データを分析対象とすると、いつ、どこで、誰が、何を購入しているかという表層的な傾向などは捉えることができるが、売れている・売れていない要因などの深層的な部分を捉えることができない。そこで、テキストデータのような非構造化データは、消費者の潜在的なニーズを把握するための情報として、活用に期待が高まっている。

Web サービス上では、オンラインショッピングサイトでのアクセス数や購入履歴などの構造化データのみならず、問い合わせフォームでの問い合わせ文、インターネット掲示板での書き込み、レビューサイトや SNS への投稿など数多くの非構造化データが溢れている。特に、商品やサービスのレビューは、消費者が品質を判断する上で重要な情報源となるため(Zeithaml, Berry &

Parasuraman, 1993)、企業のマーケティング活動における重要性が増している。レビューは、評価者による定量的な評価と、評価値を含む自由記述文で構成されることが多い。評価値とは、その属性の値や記述者の好悪に関する心的な態度を表す表現のことである(小林ら, 2005)。このようなレビューを「消費者の声」として分析することで、そこに含まれる評価値を抽出して、商品やサービスの特性を把握したり、改善に繋げたりすることが可能となる。

一方で、近年自然言語処理(Natural Language Processing: 以下、NLP)技術の進展が著しい。NLPとは、人間が使用する自然言語をコンピュータが処理することである。これまで日本語は、文法構造や単語の区切りが不明確であるため、NLPでの扱いが難しい言語だと言われてきた。しかしながら、インターネット上から大量のテキストデータを入手できるようになったことから、「コーパス」として蓄積をして、そこから規則性や確率を獲得する技術が生まれるようになってきた。このようなNLP技術の進歩によって、実用的な自然言語処理が可能となり、機械翻訳や対話システムなど、高度な分析を伴うサービスが登場するようになった。

例えば、機械翻訳ではある原言語文の単語をどの単語に置き換えるのかという語彙選択の問題と、正しい順序に並び替える並び替えの問題を解決しなければならないが、すでに実用的なレベルの機械翻訳が実現されつつある。また、米Amazon社の「Amazon Echo」などのスマートスピーカーには、精度の高い音声対話エージェントが搭載されるようになった。

このような現状を踏まえて、本研究では、インターネット上のレビューをNLPによる分析を通じて、マーケティング活動につながる新たな知見を獲得することを目指すこととする。

2. 先行研究

非構造化データを分析して、様々な知見を獲得しようとする研究は、これまで数多く行われてきた。安部らは、ホテル予約サイトのレビューを分析して、評価項目別のスコアリングから、ホテルをRecommendする手法を提案している(安部ら, 2020)。佐藤は、商品レビューを分析して、商品の評価表現の抽出を行うことで、商品カテゴリに適した特徴的な評価表現を抽出している(佐藤, 2019)。吉川らは、商品レビューの極性分析をすることで、商品の満足・不満情報を抽出して、不満解決するための代替

品をRecommendするシステムの提案をしている(吉川ら, 2020)。これらの研究は、レビューを分析することで、マーケティングへの活用を目指すものである。

金は、定型化されていないテキストは、単語や文節などの単位に分割し、それらの出現頻度や共起関係などを集計することで、定量的に解析することができるとしている(金, 2021)。松尾らは、語の共起をもとに統計的な指標を用いて、キーワードを抽出する手法を提案している(松尾ら, 2002)。岑らは、映画レビューをワードクラウドにて可視化することで、特徴を捉えやすくなったとしている(岑ら, 2021)。これらの研究より、テキストデータの分析方法としては、共起やワードクラウドのような可視化する手法が有効であることが示唆される。

次に、人が対象物に抱く印象に関する研究として、OsgoodらはSD(Semantic Differential)法を開発して、事物の情緒的意味を定量的に把握しようとした(Osgood, 1957)。SD法とは、形容詞の対を尺度として用いて、被験者に対象の印象がどちらに近いかを質問して、その程度を定量的に把握しようとする調査法のことである。SD法は感性研究において、一般的に用いられる方法として確立されているため、人の対象物に対する印象を測定する有効な方法であると言える。

佐藤は、商品レビューの分析から評価表現の抽出を行った結果、品詞のうち形容詞が消費者の評価と高い適合率があったとしている(佐藤, 2019)。さらに、鈴木らは総理大臣の国会演説における名詞の分布特徴量に注目することで、政治的解釈に有効であることを示した(鈴木ら, 2011)。これらの研究から、人の対象物に対する評価は、名詞と形容詞に注目することで捉えることができると考えられる。

以上の先行研究レビューより、本研究のリサーチクエスションは、映画レビューに含まれる名詞と形容詞に注目することで、映画作品の評価に影響を及ぼす要因を抽出することができるか、とする。これより、映画作品が鑑賞者の評価に及ぼす要因を明らかにして、映画レビューをマーケティング活動に活用するための足掛かりとする。

3. テキストマイニング

3.1. テキストマイニングとは

データは、量が多いほど価値が高いとされる。しかしながら、大量のロコミやレビュー情報などのテキストデータを分析することは容易なことではない。このような

場合、確率・統計学やAI(Artificial Intelligence: 人工知能)などを用いることで、個々のテキストデータからは読み取れない規則性や、相関関係などを観測することが可能となる。このような技術的な手法のことを「テキストマイニング」と呼ぶ。「マイニング」とは「採掘」を意味する単語であることから、テキストマイニングとは、大量のテキストデータから特徴を見出して、埋もれている情報や知識を取り出す手法のことを意味する。

大量のテキストデータに埋もれている情報や知識を取り出す手法として、「検索」や「分類」も挙げることができる。検索とは、何を求めるのかを指定して、それに対して検索結果を得るものである。分類とは、分類した結果を集計することでデータの全体像を把握して、分類された単位でデータを活用しようとするものである。これらは新たな知識や情報を見つけるものではないので、テキストマイニングとは異なるものと解されている。

3.2. テキストマイニングの流れ

テキストマイニングの前処理として、非構造化データとしてのテキストデータを、構造化する必要がある。この処理は、Feldmanらによると、(1)文字列を単語に分ける、(2)単語としての構造を付与する、(3)構造を意味に結び付ける、という順序で行われるとしている(Feldman, 2007)。特に、テキストデータは文を一つの単位として扱うより、単語ごとに分けた方が内容を正確に捉えることができるため、集計可能な単位に分解する必要がある。さらに、テキストデータは、多様性が高く、曖昧性も存在するため、どの単位で情報を抽出するかが重要となる。例えば、「自然言語」という文字列から、「自」「然」「言」「語」と一文字ずつ抽出してもあまり意味を持たないため、意味を持つ単位で抽出をする必要がある。

このような場合、NLP技術を活用して、名詞、動詞、形容詞等の品詞を最小単位として分解するとともに、品詞付与をする「形態素解析」が必要となる。形態素解析は、文法や単語が混在するテキストから、形態素解析ツールの辞書情報と照らし合わせて、テキストを構成している品詞を割り出して分類をしていくものである。

4. 可視化

4.1. 頻度集計

テキストマイニングの最も基本的な手法として、単語の頻度集計が挙げられる。大量のテキストに含まれる表

現として、出現頻度が k 番目に大きい要素は、全体に占める割合 $1/k$ に比例するという「ジップの法則」に従う傾向がある。

テキストマイニングにおける頻度集計では、出現頻度の高い順に並べて、各品詞がどれだけ含まれているかを示すものである。これにより、テキストデータ全体にどのような内容のものが、どの程度含まれているかを確認できるため、データの特徴を把握することが可能となる。

4.2. ワードクラウド

単純な単語の頻度集計だけだと、全体の傾向を直感的に捉えにくいという課題が存在する。そこで、テキストマイニングで抽出した情報の全体像を掴みやすくするためには、可視化することが重要となる。膨大なテキストデータを分析対象とする場合、表示パターンや特徴量の数が多くなるため、複雑なデータ間の関係性などを表現する手段として、可視化ツールを用いたグラフィカルなアプローチが主流となっている。

テキストマイニングの頻度分析の結果を可視化する手法として、ワードクラウドがある。ワードクラウドとは、単語の出現頻度を、フォントサイズに比例させて示したもので、フォントサイズが大きい語は、相対的に出現頻度が高いことを示す。このような特性から、ワードクラウドは、テキストデータの要約を可視化するのに適している。

4.3. 共起ネットワーク

共起とは、ある単語と単語が文中で同時に出現する頻度のことで、単語間の関係性を示すものである。共起関係の強さの尺度として、単語同士のつながりを把握するJaccard係数がある。Jaccard係数は(1)の式で表され、単語の組合せに対する積集合を和集合で除したものである。

$$J_{accard(A,B)} = \frac{n(A \cap B)}{n(A \cup B)} \quad (1)$$

この係数の値が大きいほど、二つの単語が両方使われているテキストの割合が大きいことを示すため、共起性が強いことを示す。目安として、Jaccard係数の値が0.1以上あれば共起関係があり、0.2であれば強い共起関係があると判断できる。

この共起情報を視覚化したものを共起ネットワークと呼ぶ。共起ネットワークは、語と語の段落または文に

おける共起性をもとに、ネットワーク図として可視化したもので、多くの単語の共起関係を一度に分析することを可能にするものである。

共起ネットワークは、テキストの中で用いられた単語をノード(円)と、共起を表すエッジ(線)で構成される。ノード間の距離が近いほど高い共起関係にあり、ノードの大きさが大きいほど出現回数が多いことを示す。

4.4. ネットワーク中心性

共起ネットワークのようなネットワーク図を分析する手法として、社会ネットワーク分析がある。社会ネットワーク分析とは、人間や集団など社会単位としての集合におけるネットワーク構造を分析するものである。社会ネットワークにおける階層構造をモデル化するための一つの指標として、「ネットワーク中心性」がある。

ネットワーク中心性とは、数学のグラフ理論に基づいており、社会ネットワークにおけるノードがどのくらい「中心的」であるかを示すものである。これにより、ネットワーク内における各ノードの重要性を評価したり、比較したりすることができるようになる。

Freeman は、代表的なネットワーク中心モデルとして、次数中心性、近接中心性、媒介中心性を一般化している(Freeman, 1979)。次数中心性は、ある頂点に接続しているエッジの数、すなわち隣接するノードの数に基づく中心性指標である。つまり、多くのノードと隣接しているノードほど重要であると捉えるものである。近接中心性は、ネットワーク内において他のノードに対する近さに注目した指標である。つまり、他頂点への距離の総和が小さいほど重要であると判断する。媒介中心性は、あるノードと他ノードとの間の最短経路上に位置する程度を中心性指標にしたものである。つまり、他のノード同士の関係を媒介しているほど、情報のコントロールが可能になるため、重要なノードであることを示す。よって、媒介中心性が高いノードは、そのノードを経由して情報が伝わることになるため、全体に対する影響力が大きいと言える。

媒介中心性は、ネットワーク全体に与える影響を示すものであり、ネットワーク分析において盛んに用いられている指標であるため、映画作品の評価に与える要因を捉えるのに最適であると考えられる。よって、本研究では媒介中心性を分析指標として用いて、分析を進めることとする。

5. 研究方法

5.1. 研究概要

映画レビューには、投稿者が関心のあるストーリーや出演者に関する情報のみならず、投稿者が映画を視聴した際の印象や感情などの要素が含まれる。本研究では、レビュー内に存在する名詞と形容詞に注目をして、映画作品の評価に影響を及ぼす要因を抽出する要因分析を行う。分析方法として、ワードクラウドを用いた頻度分析と、共起ネットワークを用いたネットワーク分析とする。

分析対象の映画として、「劇場版「鬼滅の刃」無限列車編」(以下、「鬼滅の刃」)と、「君の名は。」の2作品を取り上げる。2作品を取り上げる理由は、比較分析をすることで、映画作品の評価に繋がる共通的な要因を見出せると考えたためである。

「鬼滅の刃」は、2020年公開されたアニメ映画で、国内の興行収入が400億円を超えて、歴代興行収入の1位を記録した作品である。「君の名は。」は、2016年に公開されたアニメ映画で、国内の興行収入が250億円を超えた作品である。2作品とも作品への評価が高く、非常に高い興行収入を記録しているため、映画作品への評価に関する要因分析を行う研究対象として妥当だと考えることができる。

5.2. データセット

レビューは、映画レビューサイトである「Filmmarks 映画」よりスクレイピングによって収集する。「Filmmarks 映画」は、12万以上の作品が登録されており、1億件以上のレビューが投稿されている国内最大級の映画レビューサイトである。

レビューは表1のとおり、0~5までの範囲で付けられた評価スコアと、感想で構成されている。それぞれの作品で、これまで投稿されたレビューを全件取得して、評価スコアが3.5以上のものを分析対象とした。評価スコアが3.5以上とした理由は、肯定的な評価値が多く、映画評価の要因分析に適していると考えたためである。評価スコアが3.5以上のレビュー数は、「鬼滅の刃」では19,505件、「君の名は。」では27,418件である。

表1 レビューの構成

score	review
4.5	映像が綺麗！雪のシーンは本物の雪みたいで引き込まれた。最後のシーンは、自分もさることながら、周囲の大人も泣いていて、みんなが感情移入していた。これまでに観た映画で一番よかったです。
5	テレビより作画、演出が凝ってて良いです。主人公が技を放つときの炎のエフェクトがカッコいいんです！

5.3. 形態素解析

取得したレビューを、「Mecab」を用いて形態素解析を行う。Mecabはオープンソースの形態素解析エンジンであり、名詞、形容詞、助詞などの品詞ごとに分かち書きをすることができる。さらに各品詞において、細分類が可能である。例えば、名詞であれば「固有名詞」や「代名詞」などに細分化することができる。本研究では、名詞においては「固有名詞」、「一般」、「形容動詞語幹」、「サ変接続」、形容詞においては「自立」で細分類する形態素解析をすることとする。

Mecabにはシステム辞書が必要となるが、標準のシステム辞書では正しく分割できないことが多いため、本研究では、新語・固有表現などに強い「mecab-ipadic-neologd」を使用した。さらに、映画の中ではキャラクター名など映画特有の固有名詞などが出現するため、ユーザ辞書を作成して、映画の内容に応じて形態素解析できるようにした。

ワードクラウドは、pythonのライブラリである「nlp」を用いることとして、上位100件までに出現した品詞を表示した。共起ネットワークは、テキストマイニング用ツールである「KH Corder」を用いて作成し

た。設定値として、共起ネットワークのリンク数を「100」、「最小スパニング・ツリーだけ描画」、「中心性(媒介)」とした。最小スパニング・ツリーとは、エッジの重みの合計が最小になるネットワークのことで、重要と見られるエッジのみを選択して描画するものである。

6. 結果

6.1. 頻度分析

それぞれの映画において、出現数が上位10件の単語を抽出したものが表2である。「鬼滅の刃」では、「煉獄」や「炭治郎」のような映画に登場するキャラクターの固有名詞が多く出現している。特に、「煉獄」という固有名詞は他の単語と比較すると突出して出現数が多いため、映画において重要な役割を果たしていることが分かる。

一方で「君の名は。」では、「映像」、「音楽」、「ストーリー」という一般名詞が頻出しており、映画中で登場するキャラクターの固有名詞は出現していない。さらに「綺麗」という形容動詞、「良かった」、「好き」という形容詞が頻出していることが分かる。

表2 各映画作品における頻出単語

鬼滅の刃		君の名は。	
単語	出現数	単語	出現数
煉獄	17,158	作品	11,734
アニメ	10,401	映像	9,446
映画	9,788	音楽	7,206
原作	6,141	綺麗	7,075
炭治郎	5,850	よかった	6,667
漫画	5,583	ストーリー	6,318
シーン	4,773	人	4,593
鬼	4,727	好き	4,465
人	4,625	映画館	4,305

次に、それぞれの映画作品において、頻出上位100件の単語にてワードクラウドを作成した。図1の「鬼滅の刃」のワードクラウドから、「良い」「素晴らしい」「最

高」「面白かった」「かっこいい」などのポジティブな感情を表す形容詞・形容動詞が高い頻度で出現している。さらに、「感動」「号泣」「涙」「笑」などの感情を示

滅の刃」の共起ネットワークから、「煉獄」「映画」「アニメ」「炭治郎」「シーン」ノードが、媒介中心性が高いことが分かる。特に「煉獄」ノードは、最も色が濃く、「映画」「シーン」「炭次郎」ノードを媒介しているため、映画において最も重要な単語であることが示唆される。加えて、「煉獄」ノードと Jaccard 係数が最も高いノードは「炭治郎」ノードである。これより、これらの単語間には強い共起関係があることが分かる。

さらに、「煉獄」「炭治郎」「アニメ」「シーン」ノードが起点となり、大きく4つのネットワークが形成されていることが分かる。これより、本映画の評価は「煉獄」「炭治郎」「アニメ」「シーン」を中心とする要素が大きく影響していると言える。

「煉獄」ノードを中心とするネットワークでは、「煉獄」ノードと「最後」「炎」「カッコいい」「好き」ノードが繋がっている。これらの要素が、「煉獄」を特徴付けるものと考えられる。特に、「最後」ノードとの Jaccard 係数が高いため、「煉獄」ノードと「最後」ノードとの間には高い共起関係があることが分かる。「炭次郎」ノードを中心とするネットワークでは、「炭次郎」ノードと「伊之助」「鬼」「夢」「心」ノードが繋がっている。特に「伊之助」と「夢」ノードとの間に、高い共起関係があることが分かる。「アニメ」ノードを中心とするネットワークでは、「アニメ」ノードと「原作」「鑑賞」「漫画」「劇場」「鬼滅」「作品」ノードが繋がっている。特に、「原作」ノードと「漫画」ノードとの間に、高い共起関係があることが分かる。「シーン」ノードを中心とするネットワークでは、「シーン」ノードと「感動」「戦闘」「猗窩座」ノードが繋がっている。特に、「戦闘」ノードと高い共起関係が見られる。さらに「戦闘」ノードから、「迫力」「映像」「綺麗」「音楽」ノードと高い共起関係で繋がっていることが分かる。

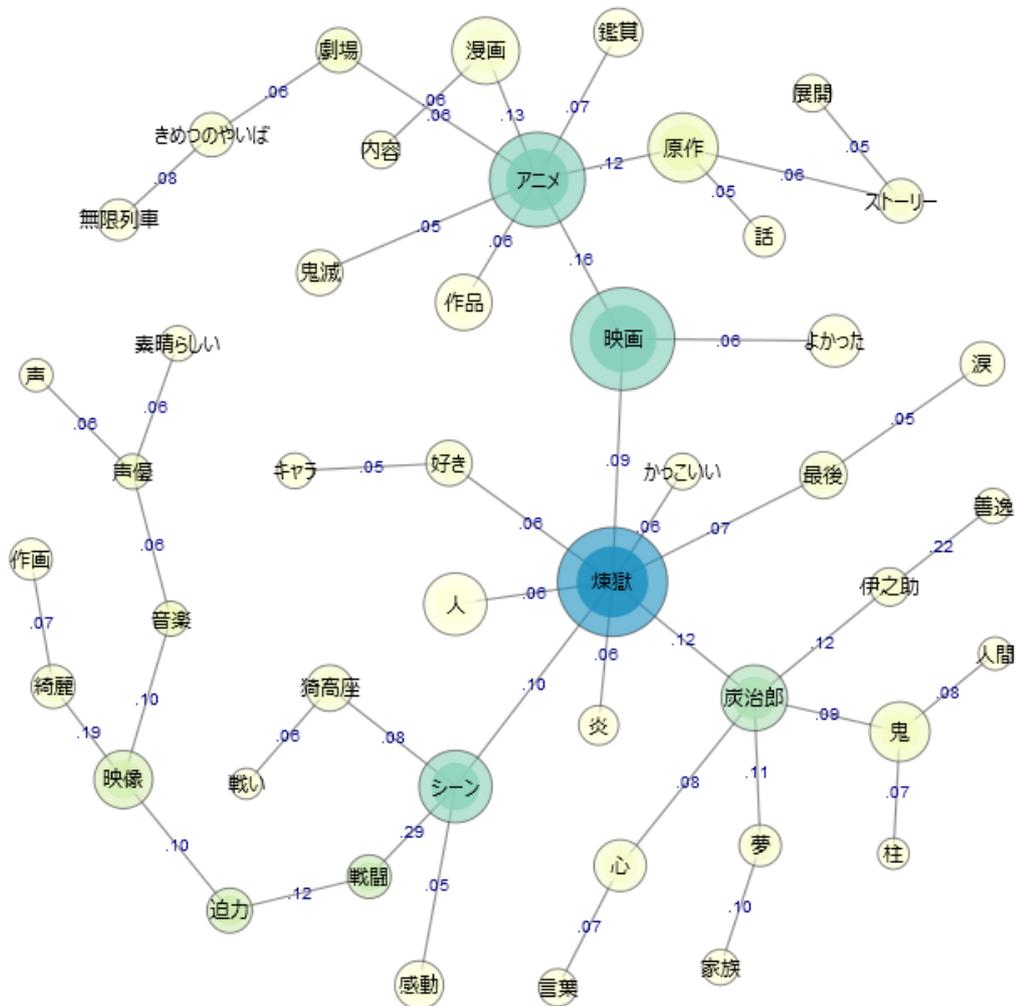


図3 「鬼滅の刃」の共起ネットワーク

6.2.2. 君の名は。

「君の名は。」の共起ネットワークを図4に示す。「君の名は。」の共起ネットワークからは、「映画」「新海誠」「作品」「映像」「シーン」ノードにおいて、媒介中心性が高いことが分かる。特に、「新海誠」ノードは、最も色が濃く、「作品」「映画」「シーン」ノードを媒介しているため、映画において最も重要な単語であることが示唆される。

さらに、「新海誠」「映画」「映像」「シーン」「作品」ノードが起点となり、大きく5つのネットワークが形成されていることが分かる。これより、本映画の評価は、「映画」「新海誠」「シーン」「作品」「映像」を中心とする要素で高くなっていると言える。

「映画」ノードを中心とするネットワークでは、「映画」ノードと「アニメ」「いい」「笑」「感動」「よか

った」ノードが繋がっている。特に「アニメ」と「よかった」ノードとの間に、高い共起関係があることが分かる。「映像」ノードを中心とするネットワークでは、「映像」ノードと「ストーリー」「音楽」「綺麗」ノードが繋がっている。特に「音楽」ノードとの間に、高い共起関係があることが分かる。「新海誠」ノードを中心とするネットワークでは、「新海誠」ノードと「監督」「好き」「名」「曲」ノードが繋がっている。特に「好き」「監督」ノードとの間に、高い共起関係があることが分かる。「シーン」ノードを中心とするネットワークでは、「シーン」ノードと「彗星」「最後」ノードが繋がっており、共に高い共起関係がある。「作品」ノードを中心とするネットワークでは、「鑑賞」「良い」「人」「新海」ノードが繋がっている。特に「新海」「人」ノードとの間に、高い共起関係があることが分かる。

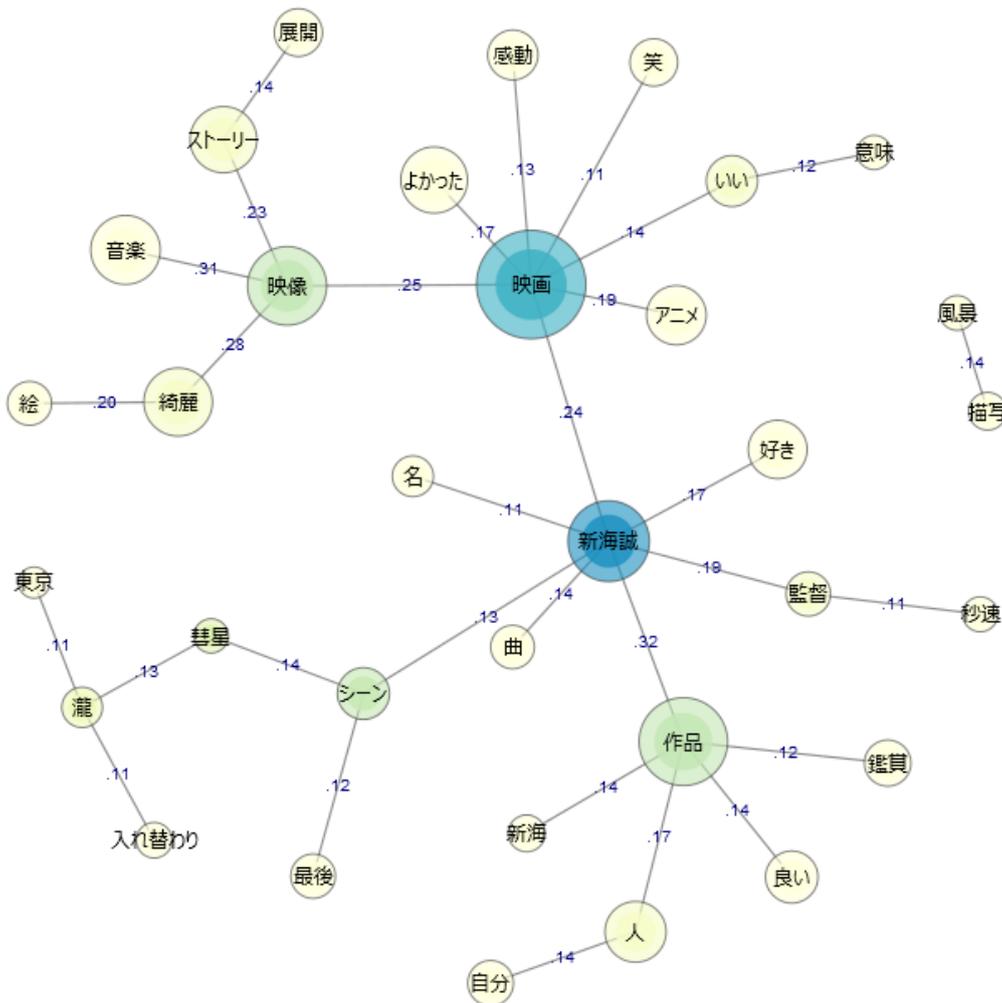


図4 「君の名は。」の共起ネットワーク

6.3. 要因分析

6.3.1. 鬼滅の刃

「鬼滅の刃」の評価に繋がる要因は、頻度分析とネットワーク分析より、「煉獄」というキャラクターの存在が大きいたことが明らかとなった。また「炭治郎」「シーン」「アニメ」も中心性が高く、それらを中心とするネットワークを構成していることから、これらの要素も評価に大きな影響を与える要素であることが分かる。

「煉獄」を中心とするネットワークでは、「カッコいい」「最後」との共起関係が強いことから、「煉獄」のキャラクター性に関わる要素や、最後の展開が評価に繋がっていると考えられる。「最後」には「涙」が繋がっているため、最後の展開は感動を引き起こすものであることが想定できる。これらより、「煉獄」のキャラクター像が確立できたことと、最後の展開が高評価に繋がっていると考えられる。

「炭次郎」を中心とするネットワークでは、「伊之助」「夢」「心」との繋がりが強く、それぞれがさらに「善逸」「家族」「言葉」と繋がっている。これらより、「炭次郎」を中心とするネットワークは、仲間や家族との絆など心理的描写に関わるネットワークで、このキャラクターに共感する部分に評価が集まっていると言える。

「シーン」を中心とするネットワークからは、「戦闘」との共起関係が強く、さらに「迫力」「映像」「綺麗」「音楽」と繋がっていることが分かる。これらより、「シーン」を中心とするネットワークは、表現や演出に関わるネットワークで、映像や音楽による効果に評価が集まっていることが分かる。

「アニメ」を中心とするネットワークでは、「漫画」「原作」との繋がりが強く、「原作」は「ストーリー」と強い繋がりがあある。これより、「アニメ」を中心とするネットワークは、原作漫画をベースとしつつ、映画の構成や展開に関わるネットワークで、ストーリー展開に評価が高まっていることが分かる。

6.3.2. 君の名は。

頻度分析から「作品」が最も出現数が多かったが、ネットワーク分析では「新海誠」が最も中心性が高い結果となった。これは「作品」を中心とするネットワークからは、有益な情報があまり見いだせないため、頻度分析とネットワーク分析の結果に相異が出たと考えられる。よって、「新海誠」が映画製作を監督したことが、高い評価に繋がった要因であると言える。

「新海誠」を中心とするネットワークから、「好き」「秒速」との繋がりが確認できるため、過去の作品を通じて映画監督としての評価が高いことが分かる。また「映画」「シーン」「映像」も中心性が高く、これらを中心とするネットワークが構成されているため、これらの要素も評価に大きな影響を及ぼしていると言える。

「映画」を中心とするネットワークは、「感動」「笑」「いい」との繋がりが強い。よって、「映画」を中心とするネットワークは、感情を喚起するネットワークで、複数の感情を生じさせたことが評価に繋がっていると考えられる。

「シーン」を中心とするネットワークでは、「最後」「彗星」との繋がりが強い。さらに「彗星」は「入れ替わり」と繋がっている。これより、「シーン」を中心とするネットワークは、映画の構成や展開に関わるネットワークで、ストーリー展開が評価に繋がっていると見える。

「映像」を中心とするネットワークでは、「音楽」「綺麗」との繋がりが強い。これより、「映像」を中心とするネットワークは、表現や演出に関するネットワークで、映像や音楽を用いた表現方法が評価に繋がっていることが分かる。

7. 結論

本研究では、映画レビューより映画作品の評価に及ぼす要因を抽出できるのではないかと問題意識から、映画レビューの名詞と形容詞に注目をして、NLPによる分析を行ってきた。本研究の成果は、以下の2点に集約される。

第1に、映画レビューの名詞と形容詞に注目をして、NLPによる分析を行うことで、映画作品の評価に大きな影響を及ぼす要素を抽出したとともに、それらの構造を明らかにした点である。本研究では、映画レビューという非構造化データを、形態素解析を通じて分析を進めてきた。その結果、「鬼滅の刃」では、「煉獄」キャラクター像を中心に、共感をもたらす要素、映像や音楽による演出の要素、ストーリー構成や展開に関する要素が結び付いて、評価に繋がっていることが明らかになった。「君の名は。」では、「新海誠」の作品であることを中心として、感情を喚起する要素、ストーリー構成や展開に関する要素、映像や音楽による演出の要素が結び付いていることが判明した。これにより、大量のレビューデー

タから、映画評価に繋がる要素と構造に縮約することができた。

第2に、2つの映画作品の要因分析より、映画評価が高まる共通的な要因を抽出できた点である。2映画作品の要因分析より、評価に繋がる共通要因として、映像や音楽による演出と、ストーリーの構成や展開であることが明らかとなった。これより異なる作品にも関わらず、評価に繋がる共通的な要素を抽出することができたと言える。

以上より、映画レビューの名詞と形容詞に注目してNLPによる分析を行うことで、映画作品の評価に繋がる要因を明らかにすることができた。これにより、本研究で用いた分析手法は、映画評価の要因分析のみならず、

レビューを活用したマーケティング活動に活用することが期待される。

一方で、本研究の課題としては、映画作品の評価に繋がる要因は明らかにできたものの、評価に繋がるより深層的な部分まで迫ることができなかった点が挙げられる。本研究では、単語の出現頻度を尺度としてきたが、レビューには評価に繋がる深層的な部分まですべて記述されているわけではない。よって、より深層的な要因を抽出するためには、出現頻度は少ないものの特徴的な単語を汲み取る必要があると考えられる。

今後は、特徴的な単語を抽出するアルゴリズムの導入を検討しつつ、他の映画作品にも分析対象を広げて、映画作品の評価に及ぼす要因に関する研究をさらに深化させていきたい。

参考文献

- [1] Feldman, R., and Sanger, J., *The text mining handbook: advanced approaches in analyzing unstructured data*, Cambridge University Press, 2007.
- [2] Freeman L. C., *A Set of Measures of Centrality Based on Betweenness*, *Sociometry*, Vol. 40, No. 1, pp35–41, 1977.
- [3] Freeman, L. C., *Centrality in social networks: Conceptual clarification*, *Social Networks*, Vol. 1, No. 3, pp215–239, 1979.
- [4] Osgood, C. E., Suci, G. J., and Tannenbaum P. H., *The Measurement of Meaning*, University of Illinois Press, 1957.
- [5] Zeithaml, V. A., Berry, L. L., and Parasuraman. A., *The nature and determinants of customer expectations of service*, *Journal of the Academy of Marketing science*, Vol. 21, No. 1, pp1-12, 1993.
- [6] 安部克, 中島伸介「レビュー自動スコアリング方式に基づくホテル推薦システム」, *DEIM Forum 2020*, B4-2, 2020.
- [7] 金明哲「テキストアナリティクスの基礎と実践」, 岩波書店, 2021.
- [8] 佐藤 利紀「品詞共起パターンを用いた評価表現抽出方法 —商品企画の支援を目指して—」, 会津大学短期大学部産業情報学科経営情報コース 2018 年度卒業研究論文要旨集, 2019.
- [9] 西尾寅弥「形容詞の意味・用法の記述的研究」, 秀英出版, 1972.
- [10] 松尾豊, 石塚満「語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム」, *人工知能学会誌*, Vol. 17, No. 3, pp.217-223, 2002.
- [11] 鈴木崇史, 影浦 映「名詞の分布特徴量を用いた政治テキスト分析」, *行動計量学*, Vol. 38, No. 1, pp83-92, 2011.
- [12] 吉川耀敬, 王元元「商品レビューの不満・満足情報抽出に基づく商品推薦システムの提案」, *DEIM Forum 2021*, A5-5, 2021.