

直観的動画検索システムにおける 評価項目別スコアリングの精度向上

笹原 彰斗[†] 山口 史弥^{††} 上田真由美^{†††,††††} 中島 伸介[†]

[†] 京都産業大学 情報理工学部 〒603-8555 京都府京都市北区上賀茂本山

^{††} 京都産業大学 先端情報学研究科 〒603-8555 京都府京都市北区上賀茂本山

^{†††} 流通科学大学 経済学部 〒651-2188 兵庫県神戸市西区学園西町 3-1

^{††††} 大阪大学 サイバーメディアセンター 〒560-0043 大阪府豊中市待兼山町 1 番 32 号

E-mail: [†]{g2054263,nakajima}@cc.kyoto-su.ac.jp, ^{††}i2286211@cc.kyoto-su.ac.jp,

^{†††}Mayumi_Ueda@red.umds.ac.jp

あらまし 近年、動画共有サイトには膨大な動画が存在し、多くのユーザはタイトルやサムネイルから好みの動画を判断して視聴する。しかし、既存の情報では実際に再生するまでユーザが真に求めている動画か判断することが困難である。我々は、動画に対して投稿されたコメントを分析し、その動画に対するユーザの直感的評価である評価項目別スコアの算出に取り組んできた。本稿では、提案手法である評価項目別スコアリング手法の精度向上に取り組んだので報告する。

キーワード コメント分析, 直感的動画検索, 自動スコアリング

1 はじめに

近年、動画共有サイトには膨大な動画が存在し、多くのユーザはキーワード検索をした後に、タイトルやサムネイルから好みの動画を判断して視聴する。しかしながら、ユーザが直感的に観たいと思える動画を検索することは、従来のキーワード検索では限界がある。例えば、比較的新しい表現である「エモい」動画を検索するケースを考える。「エモい」という言葉でキーワード検索した場合、タイトルや説明文に「エモい」というキーワードを含む動画がヒットするが、「エモい」という言葉は含まないものの概念的にエモい動画はヒットしない。特に「エモい」のような新しい表現の場合、コメントに「エモい」という記述がされてない動画も多く存在すると思われるため、従来の単純なキーワード検索は有効に機能しない可能性が高いといえる。タイトルやコメントに「エモい」というキーワードが入ってなくても、「エモい」と類似した意味を持つ「懐かしい」や「切ない」といった言葉が含まれていることが考えられるため、これらのキーワードを適切に分析することで「エモい」動画を効果的に見つけることが可能であると考えた。

我々は先行研究において、ユーザが直感的に観たいと思える直感的動画検索を可能とする、動画に対する評価項目別スコアを用いた直観的動画検索システム [4] を提案した。これは、動画の印象を表す 9 つの各評価項目ごとに fastText で分類器モデルを構築し、動画の特徴をレーダーチャートでユーザに提示するシステムである。このシステムでは、評価項目を表すキーワードをタイトルやコメントに含んでいない動画に対しても、有効な検索を可能にしている。本研究では、学習データ（コメント数）の増加・各種学習方法の検討・ユーザフィードバック

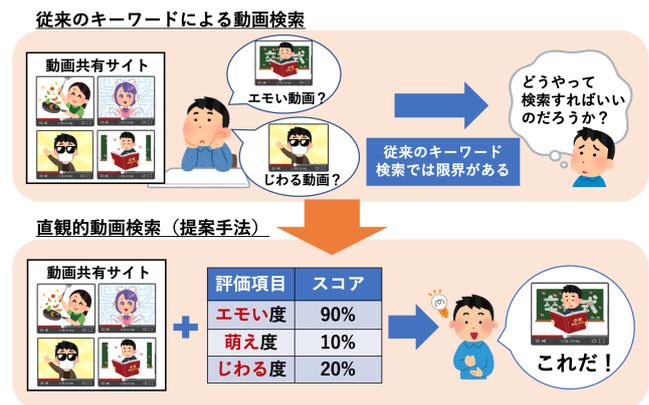


図1 評価項目別スコアリングの手順

を用いた再学習を行うことにより、従来の直観的動画検索システムにおける評価項目別スコアリングの精度向上に取り組んだ。本稿では、評価項目別スコアリングの精度向上手法について説明する。

2 関連研究

以下に、動画の分析およびテキスト分析に関する関連研究をあげ、本研究との差異を示す。近年の動画共有サイトではユーザが自由にコメントすることができる、YouTube [1] やニコニコ動画 [3] などが存在する。また、上記のサイトには膨大な数の動画が日々アップロードされており、タグやキーワードを用いた検索が可能である。

動画分析に基づく関連研究として、村上らは動画に付与されるコメントの「w」という文字に着目し、その動画の「笑い」の度合いを数値化した [5]。中村らは、感性に基づく正規表現辞書

を構築し、ユーザが書き込んだコメントを分析して、positive, negative, happy, sad, surprise といった感性に基づくランキング検索を可能とした [6]。本研究では、これらの感性分析には留まらず、ユーザのより直感的な動画検索を可能とする技術の開発を目指している。

また、レビュー分析に基づく評価表現別自動スコアリング手法に関する従来研究として、松波らは、コスメアイテムに特化した評価表現辞書を構築し、レビュー自動スコアリング手法を開発した [7]。同様に、安部らは、ホテル予約サイトに投稿されたレビュー文を分析することで、ホテルに特化した評価表現辞書を構築し、ホテルレビュー自動スコアリング手法を開発した [8]。

本研究ではこれらの辞書構築手法を参考にしつつも、動画に対するより直感的な評価項目の選定と各評価項目に対する評価表現の効率的な抽出を目指すものである。上述の通り、動画のコメント分析による感性検索の研究や、レビュー解析による評価表現別のスコアリング手法に関する研究は数多く行われているが、本研究にて実施している直感的動画検索のための評価表現別自動スコアリング手法に関する研究開発は行われていない。本研究では、直感的動画検索のための評価表現別自動スコアリング手法に関する研究開発の精度向上に取り組む。

3 評価項目別スコアリング手法

3.1 各評価項目の分類器モデルの構築

本節では、動画に対する評価項目別スコアリングのための分類器モデルの従来の構築方法について述べる。動画共有サイトにおいて動画を視聴するユーザは様々な直感的表現を用いたコメントを投稿している。そこで、実際に使用されている直感的表現を用いたコメントを収集し、学習データとして使用することで評価項目別分類器モデルの構築を行う。

学習データとして使用するコメントデータは、動画共有サイト YouTube が提供する YouTube Data API [2] から取得する。分類器モデルの構築には fastText [9] を用いている。例としてエモい分類器モデル構築についての手順を示す。

エモい分類器構築手順

- (1) 評価項目だと感じる動画を収集し、各評価項目ごとに 5000 コメントずつ収集する。
- (2) fastText を用いて評価項目 9 つに対してマルチラベル分類モデルを作成する。
- (3) エモくないコメントを収集するため、(2) で構築した分類器を用いて 200 本の動画をスコアリングする。
- (4) エモいスコアが 0.05 以下の動画をエモくない動画と定め、そこから 5000 コメントを収集する。
- (5) エモいコメント 5000 とエモくないコメント 5000 を用いて fastText で学習させエモい分類器モデルを構築する。

3.2 評価項目別スコアリングの精度向上手法

3.2.1 学習データ (コメント数) に関する検討

従来のシステムでは、分類器を構築する際、学習データとして

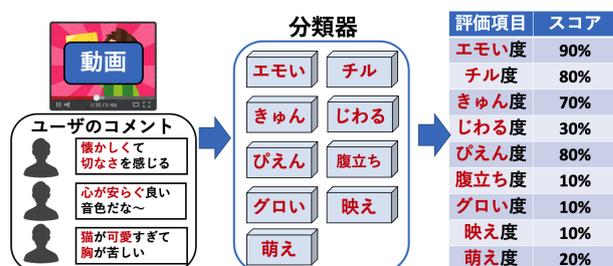


図 2 評価項目別スコアリングの手順

各評価項目ごとに事前に用事した 5000 コメントずつを収集していたが、本研究では収集するコメント数を 10000,15000,20000 に増加させることによって各評価項目に対するスコアリングの精度が向上するか検証する。

3.2.2 各種学習方法に関する検討

従来のシステムでは、単語のベクトル化手法として fastText を用いて学習を行っていたが、本研究では BERT などその他の学習手法を適用することによってスコアリング精度が向上するか検証する。

3.2.3 ユーザフィードバックを用いた再学習

評価項目別スコアリングの精度向上手法の一つとして、ユーザフィードバックに基づく再学習手法の開発を検討している。具体的には、システムを利用してユーザが動画を視聴した際、例えば「エモい」スコアが高いある動画に対して、実際には「エモい」と感じなかった場合は、用意された「エモくない」チェックリストにチェックを入れる。逆にその他の評価項目のスコア、例えば「萌え」スコアが低いもの実際には「萌え」スコアが高い動画であると感じた場合には、そのユーザに「萌え」のチェックリストにチェックを入れてもらう。このように、スコアリングが適切ではない動画に対するフィードバックを蓄積し、再学習を行う際の学習データの追加・削除を行うことで、スコアリング精度の向上を実現しようとしている。

4 おわりに

本稿では、学習データ (コメント数) の増加・各種学習方法の検討・ユーザフィードバックを用いた再学習を行うことにより、従来の直感的動画検索システムにおける評価項目別スコアリングの精度向上に取り組んだ。今後は精度向上の評価実験を実施する。また、評価項目のキーワード収集についても、データ分析に基づき自動収集するアプローチも検討する。

謝 辞

本研究の一部は、JSPS 科研費 (課題番号: 19K12243, 20H04293, 22K12281) および京都産業大学先端科学技術研究所 (ヒューマン・マシン・データ共生科学研究センター) 共同研究プロジェクト (M2001) の助成を受けたものである。ここに記して謝意を表す。

文 献

- [1] YouTube <https://www.youtube.com/>
- [2] YouTube Data API
<https://developers.google.com/youtube/v3/getting-started>
- [3] ニコニコ動画 <https://www.nicovideo.jp/>
- [4] 鴫田 和士ら, 動画に対する評価項目別スコアを用いた直感的動画検索システム, DEIM Forum 2022 C21-5, 2022.
- [5] 村上 直至ら, 動画コンテンツの視聴者コメントに基づくランキングとその評価, DEIM Forum 2012.
- [6] 中村 聡史ら, ソーシャルアノテーションに基づく動画検索手法, DEIM Forum 2009 D6-1, 2009.
- [7] 松波 友稀ら, コスメアイテムに対する評価項目別レビュー自動スコアリング方式の開発, DEIM Forum 2017 B5-3, 2017.
- [8] 安部 克ら, レビュー評価項目別スコアを用いたホテル推薦手法の提案, DEIM Forum 2019 P1-134, 2019.
- [9] fastText <https://fasttext.cc/>