7Y-02

自然言語を利用したライブストリーミングの特徴抽出

本多 充稔 †　　藤田 桂英 ‡
東京農工大学 †　東京農工大学 ‡

1　はじめに

近年，ライブストリーミングの市場規模は急速に拡大している．しかし，その膨大なコンテンツを理解し，処理するための技術基盤は未だ発展途上にある．本研究では，ライブストリーミングの内容を汎用的に表現する特徴抽出手法の提案を目的とする．具体的には，ライブストリーミング固有の自然言語情報を活用し，ジャンル推定タスクを学習させることで，動画の内容を反映した特徴量を得る手法を提案する．また，抽出した特徴量を用いた推薦システムを構築し，その有効性を検証する．

2　学習用データの収集

YouTube Live のアーカイブ動画を学習データとして収集した．アーカイブ動画ではチャットリプレイが時刻と紐づいて保存されているため，リアルタイム配信と等価な時系列情報を取得可能である．2016 年から 2025 年までに配信された，YouTube Live の全ジャンルを対象とする動画リストを作成し，トランスクリプト（字幕データ），チャットデータ，プロフィール情報（再生数，高評価数等）を収集した．

3　チャット特化型の言語モデル

ライブストリーミングのチャットは，短文・スラング・絵文字の多用といった特徴を持つ．そこで本研究では，収集したアーカイブ動画のチャットデータを用いて，BERT[1] へファインチューニングを行った（以後 Chat-BERT と呼称）．有害チャット検出タスク [2] を用いた評価実験を行った結果，TF-IDF 等の古典的手法よりは大幅に高い精度を示したものの，既存の BERT モデルと比較して決定的な性能差は見られなかった．これは，有害チャットに含まれる暴言等が一般的なコーパスにも含まれる語彙であったためと考えられる．

4　ジャンル推定による特徴抽出

ライブストリーミングの内容を表す特徴量を得るためのアプローチとして，動画のジャンル（ゲーム・音楽・スポーツなど）を推定するモデルを構築し，その中間層の出力を特徴量として利用する．ジャンルは動画の内容を大まかに表す指標であり，これを学習したモデルは中間層で配信内容の意味的な情報を保持していると考えられる．

4.1　モデル構成

入力データとして，トランスクリプトとチャットを利用する．動画の配信時間は様々であるため，各動画を N 個の時間区間に分割し，固定長の時系列データとして扱う．

- **トランスクリプト**: 各区間のテキストを結合し，BERT でベクトル化する．
- **チャット**: Chat-BERT を用いて各チャットをベクトル化した後に，区間ごとの平均ベクトルを取得する．

時系列情報の学習には LSTM を用いる．トランスクリプト単体・チャット単体・および両者の LSTM の出力を結合したモデルの 3 種類を

Feature extraction for live streaming using natural language

† Atsumi HONDA, Tokyo University of Agriculture and Technology
‡ Katsuhide FUJITA, Tokyo University of Agriculture and Technology

Copyright ©2026 Information Processing Society of Japan.
All Rights Reserved.

表1　ジャンル推定タスクの実験結果（一部抜粋）

手法	Accuracy
プロフィール情報	0.596
トランスクリプト単体 ($N = 10$)	0.675
チャット単体 ($N = 1$)	0.702
結合モデル (Tr:$N = 10$, Chat:$N = 1$)	0.719

表2　推薦システムに対する被験者評価の統計量

手法	平均値	標準偏差
プロフィール情報	1.98	1.22
トランスクリプト	2.31	1.20
チャット	2.54	1.16
結合モデル	2.99	1.07

構築した．

4.2　ジャンル推定実験と結果

2016年から2019年のデータを用い，YouTube Liveで配信者が設定したジャンルを正解ラベルとして学習・評価を行った．比較対象として，プロフィール情報のみを用いたベースラインを用意した．実験の結果の主要部を表1に示す．自然言語情報を利用したモデルは，プロフィール情報のみの場合よりも高い精度を示した．特に，トランスクリプトとチャットを併用した結合モデルが最も高い精度を達成した．また，時系列の分割数 N によって精度が変動し，トランスクリプトは適度な長さ（$N = 10$），チャットは全体平均（$N = 1$）に近い方が精度が高い傾向が見られた．

5　推薦システムの構築と評価

5.1　推薦システムの流れ

提案手法によって抽出された特徴量の妥当性を検証するため，類似動画の推薦システムを構築した．あるクエリ動画に対し，特徴量ベクトルのコサイン類似度が最も高い動画を推薦する．もし特徴量が内容を適切に表現していれば，人間が「似ている」と感じる動画が推薦される．

5.2　被験者実験

被験者16名に対し，クエリ動画と，4つの手法（プロフィール情報・トランスクリプト・チャット・結合モデル）により推薦された動画を提示し，その類似度を4段階のリッカート尺度（1：全く似ていない 〜 4：非常に似ている）で評価させた．

5.3　実験結果

評価実験の結果を表2に示す．プロフィール情報に基づく推薦は評価が低く，表面的なメタデータでは内容の類似性を捉えられないことが示された．一方で，自然言語情報を用いた手法は高い評価を得ており，特に結合モデルが最も高く評価された．Nemenyi検定の結果，チャットと結合モデルはプロフィール情報に対して有意水準5%で有意差が確認された．これにより，提案手法は配信の雰囲気や文脈を含めた内容の類似性を捉える有効な特徴量を抽出可能であることが示された．

6　おわりに

本研究では，ライブストリーミングの自然言語情報を活用した特徴抽出手法を提案した．ジャンル推定を利用することで，プロフィール情報よりも内容を反映した特徴量が得られることを確認した．トランスクリプトとチャットを統合することで，動画の内容と雰囲気を補完的に捉え，人間の感覚に近い推薦が可能となった．

参考文献

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT Vol.1*, pp. 4171–4186, 2019.

[2] Yasuaki Uechi. Sensai: Toxic chat dataset, https://github.com/holodata/sensai-dataset 2021.

Copyright ©2026 Information Processing Society of Japan.
All Rights Reserved.