

2023 年度 修士論文概要

主査	舟橋 健司	副査	徳田 恵一	研究室	舟橋研究室
入学年度	2022 年度	学籍番号	34414031	氏名	上村 岳斗
論文題目	インターネット配信動画におけるテキストチャット解析によるハイライト動画生成 (Highlight Clip Video Generation with Text Chat Analysis for Internet Streaming Video)				

1 はじめに

近年、情報通信技術の発達や COVID-19 の流行によって、自宅で楽しむ娯楽としてインターネットでの動画視聴の需要が高まっている。動画配信サービスでは、視聴者に対してリアルタイムな臨場感を提供できるライブ配信という仕組みがある。しかし、ライブ配信は数時間行われる場合があり、その内容をすべて視聴することは難しい。そのため、ライブ配信におけるハイライト部分のみを抽出したハイライト動画の需要が高く、実際にハイライト動画を自動生成するための研究が行われている。ライブ配信中に視聴者から書き込まれたテキストチャットを利用してハイライト動画を自動生成する研究 [1][2] があるが、アノテーションの手間や、特定の個人に向けたハイライト動画生成に限定していること、また、テキストチャットの書き込み数が多い箇所がハイライト部分とは限らないという課題がある。そこで、本研究では、人気のあるハイライト動画をテキストチャットのアノテーションに利用することで、同ジャンルの動画に対するハイライト動画作成時のアノテーションを不要にした上で、特定の個人ではなく多くの人がハイライト部分だと感じる箇所を抽出する。

2 提案手法

2.1 概要

提案手法の概要を図 1 に示す。本研究ではまず、ライブ配信中に書き込まれたテキストチャットが、ライブ配信におけるハイライト部分に書き込まれたものかどうかの 2 値分類を行う機械学習モデルを、日本語のデータセットを用いて事前学習済みの BERT[3] をファインチューニングして作成する。次に、10 秒を 1 区間として、モデルがハイライト部分に含まれると判定したテキストチャットが多い区間順に、その区間と前 2 区間と合わせた 3 区間の合計 30 秒をハイライト動画に含める区間として決定する。あらかじめ設定した閾値に決定した区間数が達したところでハイライト動画に含める区間の決定を終え、各区間を元のライブ配信動画と同じ時刻に合うよう並び替えることでハイライト動画を自動生成する。なお、人気のあるハイライト動画作成者は、多くの人が良いと感じるハイライト動画の作成ができていると仮定する。データのアノテーシ

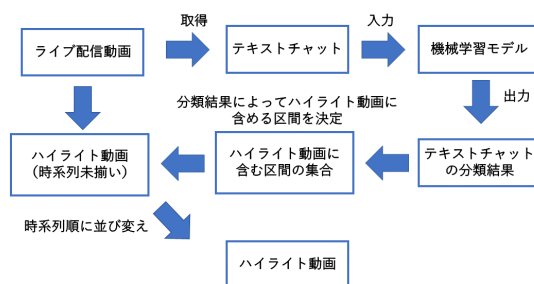


図 1: 提案手法の概要図

ンには、オリジナルのライブ配信のハイライト動画作成を行ったユーザーの内、チャンネル登録者数が 1 万人を超えている 2 名が作成したハイライト動画を利用した。

2.2 学習用データセット

データセットとして、ANYCOLOR 株式会社に所属している「葛葉」が VALORANT をプレイしているライブ配信 A から取得したテキストチャットと、葛葉のライブ配信のハイライト動画作成のみを行っているユーザーの「葛葉切り抜き ch」と「葛葉切り抜き集」が A からそれぞれ作成したハイライト動画を用いる。動画内のハイライト部分に対し視聴者が書き込んだテキストチャットの特徴を学習できるように、以下の手順でテキストチャットのアノテーションを行った。まず、それぞれのハイライト動画には各シーンごとにキャプションがついていたため、それを利用しハイライト動画内の各シーンにおけるハイライト部分のシーンを判断した。入力の手間やタイムラグなどを考慮し、ライブ配信において、そのシーンが始まってから終了後 20 秒の間に送信されたテキストチャットを正解、それ以外を不正解としてアノテーションを行った。また、書き込まれたテキストチャットにおいて、主に正解データ内に多くみられた表現の内、不正解データ内にもみられた、['w', '草', '笑', 'ナイス', '惜しい'] を含む表現と、感嘆表現のテキストチャットを正解とした。その際、例えば「ないす、おいしい」などの意味は同じだが表記が違う表現についても正解とした。また、前処理としてアカウントに対して月額料金を支払うこ

とで利用できるスタンプ、URL を削除した。また、連続文字列を 2 文字までに制限し、完全に同一の表現は一つを残して削除した。その際、正解データと不正解データについて同一の表現があった場合は、正解データを残した。

2.3 モデルと学習結果

BERT の最終層に Linear 層、Pooling 層、CNN 層を追加したモデル [4] を作成し、epoch 数 50、バッチサイズ 8、学習率の初期値は 1×10^{-3} とし、学習を行った。50epoch 中で、検証用データに対する損失が最も低かったモデルのテストデータに対する結果を表 1 に示す。モデルによる精度の違いがみられないことから、ハイライト動画の生成には、BERT+Linear モデルを用いる。

表 1: 学習したモデルの精度

モデル名	Accuracy	Recall	Specificity
BERT+Linear	0.82	0.71	0.93
BERT+Pooling	0.80	0.75	0.84
BERT+CNN	0.81	0.68	0.94

3 評価実験

被験者に自動生成したハイライト動画を含む、4 つの動画を視聴してもらい、提案手法の有用性の評価を行う。対象とする動画は、学習用に用いたライブ配信動画 A の配信者の葛葉が VALORANT をプレイしている別のライブ配信動画 B とする。また、被験者の負担を考慮し、すべての動画を視聴した際の時間が 1 時間程度になるよう B の冒頭 40 分を対象とする。被験者には、以下の 4 つの動画を記述する順番で視聴してもらい、それぞれの評価を最も悪ければ 0、最も良ければ 10 とする 11 段階評価と、任意の自由記述での評価を行ってもらった。その際、どのようなハイライト動画が良いかは個人によるため、特定の質問は設けていない。

1. ライブ配信動画 B の冒頭 40 分
2. 葛葉切り抜き ch が作成したハイライト動画の内 B の冒頭 40 分に含まれる区間 (7 分 58 秒)
3. B の冒頭 40 分を元に提案手法により生成したハイライト動画 (8 分)
4. B の冒頭 40 分を元に区間をランダムに選択し生成したハイライト動画 (8 分)

各被験者の評価結果より作成した箱ひげ図を図 2 に示す。図 2 より、提案手法を用いて自動生成したハイライト動画が人気のハイライト動画作成者が作成した動画よりは劣るが、良い評価を得られている。また、区間をランダムに選択して生成したハイライト動画の評価との比較と、本手法で自動生成したハイライト動画

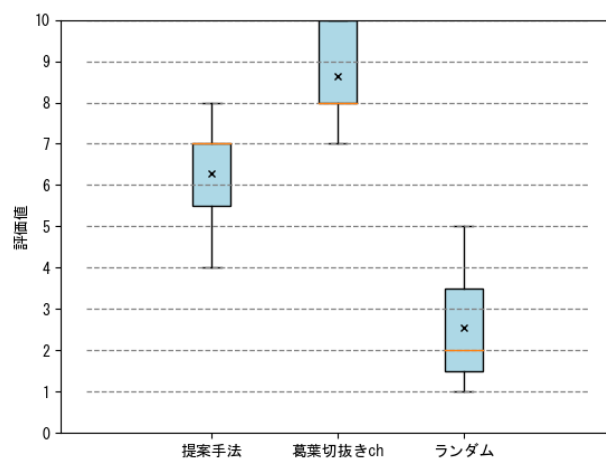


図 2: 被験者の評価を用いて作成した箱ひげ図

に対する「切り抜いた部分に関してはどれも配信の中で面白い場面ではあった」などの自由記述から、提案手法により、多くの人がハイライト部分だと感じる箇所の抽出ができていとみられる。

4 むすび

本研究では、ライブ配信中に視聴者が書き込んだテキストチャットを利用し、オリジナルのライブ配信を視聴したことがある被験者を対象に、自動生成したハイライト動画を視聴してもらった。提案手法により、多くの人が良いと感じるハイライト動画を、個別にアノテーションを行うことなく自動生成できたことを確認した。今後の展望として、他のジャンルのライブ配信のハイライト動画の作成や、提案手法の定量的な評価が挙げられる。

参考文献

- [1] 小城凱, 伊東栄典, “動画のハイライト部分自動抽出に向けた検討”, 情報科学技術フォーラム講演論文集, 22 卷 (第 2 分冊), pp.473-474, 2023.
- [2] 植野天翔, 平井辰典, “BERT を用いてライブ配信のチャットから視聴者の好みに合ったハイライト動画を生成するシステム”, 第 85 回全国大会講演論文集, 2023, pp.763-764, 2023.
- [3] “Hugging Face,” <https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>. (Accessed on 01/06/2024).
- [4] K. Kaur, P. Kaur, “BERT-CNN: Improving BERT for Requirements Classification using CNN,” Procedia Computer Science, Vol.218, pp.2604-2611, 2023.