
補完的物語生成手法に基づき拡張した あらすじ文を用いた映画の興行的成功予測

飯倉 陸

大阪府立大学大学院工学研究科

Riku Iikura

Graduate School of Engineering, Osaka Prefecture University

1 要旨

映画産業は、製作費が数 100 億円以上にも及ぶ大規模なビジネスである。そのため、企画段階でその映画が興行的に成功するか否かを予測することは、製作陣や製作費の投資家にとって重要といえる。近年の深層学習技術の発展に伴い、映画の監督や出演者、公開年などのメタ情報を特徴量として、興行収入を予測する手法が複数提案されてきた。一方で、それらメタ情報に、物語の流れを表現したあらすじ文の言語情報を合わせた特徴量を使用する研究は、ほとんど報告されていない。そこで本研究では、映画のメタ情報と、事前学習済みの BERT モデルに基づき処理したあらすじ文の言語情報を合わせたマルチモーダルな特徴量を用いて、興行的成功の予測精度の向上を図る。また、映画のあらすじ文データに関して、英語のデータセットは整備されている一方で、日本語のデータセットは未だ十分でない。そのため、日本語のあらすじ文を含む、興行的成功予測のためのデータセットを新たに構築する。しかしながら、収集可能なデータ数は限られているため、あらすじ文に対してデータ拡張手法を適用し、サイズ不足を補う。本研究では、物語的文章の特性に着目し、文章の最初と最後の文から補完的に物語を生成する、Variational Hierarchical Conversation RNN ベースの補完的物語生成手法 (VHCR-CSGM) を提案する。そして、VHCR-CSGM に基づき、映画のあらすじ文データを拡張する手法を提案する。実験の結果、映画のメタ情報のみあるいはあらすじ文のみを特徴量とするユニモーダルな手法より、マルチモーダルな特徴量を使用する手法の方が、興行的成功の予測精度は高くなった。さらに、VHCR-CSGM により拡張したあらすじ文データを含む訓練データに基づくモデルは、従来のテキスト拡張手法を用いた場合よりも高い予測性能を有することを確認した。