**DỰ ĐOÁN QUAN HỆ HỢP TÁC TRONG TƯƠNG LAI GIỮA CÁC DIỄN VIÊN**

## Lê Thái Dương, Trần Văn Vinh, Lâm Tấn Duy

Trường Đại Học Công nghệ thành phố Hồ Chí Minh, VietNam

**Abstract:** Dự đoán khả năng các diễn viên hợp tác trong các dự án phim trong tương lai là một bài toán thú vị và thử thách tại giao thoa giữa khoa học dữ liệu và ngành công nghiệp giải trí. Nghiên cứu này tập trung phát triển một mô hình dự đoán phân tích dữ liệu lịch sử về phim ảnh, mạng lưới các diễn viên, và các mô hình hợp tác để dự báo khả năng ghép các cặp diễn viên trong các bộ phim sắp tới.

Ứng dụng các kỹ thuật học máy, bao gồm phân tích mạng xã hội và xử lý ngôn ngữ tự nhiên, mô hình xét đến các yếu tố như sở thích thể loại phim, lịch sử hợp tác, và xu hướng ngành. Nghiên cứu kết hợp các bộ dữ liệu công khai như IMDb và hoạt động trên mạng xã hội để xây dựng hồ sơ diễn viên và biểu đồ hợp tác.

Kết quả nghiên cứu cho thấy những yếu tố chính ảnh hưởng đến việc ghép cặp diễn viên và minh chứng tính khả thi của việc sử dụng phân tích dự đoán trong quyết định tuyển diễn. Nghiên cứu này không chỉ cung cấp những thông tin hữu ích cho đạo diễn tuyển diễn và các nhà sản xuất mà còn đặt nền tảng cho những ứng dụng tiếp theo của mô hình dự đoán trong các ngành sáng tạo.

***Purpose:*** Mục đích của nghiên cứu này là nhằm tìm hiểu và phát triển các mô hình phân tích dữ liệu để hỗ trợ dự đoán việc ghép diễn viên, đánh giá tính khả thi của các ứng dụng học máy trong tuyển diễn, và cung cấp các góc nhìn mới về việc khai thác tiềm năng của khoa học dữ liệu trong ngành giải trí.

***Method:*** Data được thu thập bằng phương pháp crawl trên web phim nhằm lấy cách thông tin như tên diễn viên, điểm số của khán giả, tên phim, thể loại phim, … Áp dụng các mô hình đồ thị như GNN, thuật toán Louvain, thuật toán Gỉvan-newman.

***Results and Findings:***

***Originality:*** Bài nghiêm cứu này cho thấy tổng quan về những thông tin liên quan ảnh hưởng tới chất lượng bộ phim

***Limitation:*** Nghiêm cứu của chúng tôi chỉ thu hẹp ở phạm vi tìm ra những diễn viên có khả năng sẽ diễn cùng nhau trong bộ phim sắp tới.

***Implication:*** Kết quả của nghiên cứu này mang lại những ý nghĩa quan trọng đối với ngành công nghiệp giải trí. Thứ nhất, việc sử dụng mô hình dự đoán có thể giúp các nhà sản xuất và đạo diễn đưa ra quyết định sáng suốt hơn trong việc tuyển chọn và ghép cặp diễn viên, từ đó tăng cơ hội thành công của các dự án phim. Thứ hai, nghiên cứu mở ra một hướng đi mới trong việc ứng dụng khoa học dữ liệu vào việc tối ưu hóa quy trình sản xuất phim, giúp tiết kiệm chi phí và tối ưu hóa nguồn lực. Cuối cùng, những kết quả này có thể được áp dụng không chỉ trong lĩnh vực điện ảnh mà còn trong các ngành công nghiệp sáng tạo khác, nơi mà yếu tố hợp tác giữa các cá nhân đóng vai trò quan trọng.

**Keywords:** Actor collaboration, predictive modeling, film industry, data science, machine learning, social network analysis, movie success prediction.

1. **INTRODUCTION**

Trong ngành công nghiệp giải trí, việc lựa chọn diễn viên không chỉ dựa trên kỹ năng diễn xuất mà còn phụ thuộc vào khả năng tạo ra sự gắn kết và tương tác tốt trên màn ảnh. Tuy nhiên, quyết định ghép cặp diễn viên thường mang tính chủ quan, dựa trên kinh nghiệm của các đạo diễn và nhà sản xuất. Sự phát triển của khoa học dữ liệu và học máy mang lại cơ hội mới để tối ưu hóa quá trình này, bằng cách đưa ra các dự đoán dựa trên phân tích dữ liệu lịch sử và các yếu tố liên quan.

Thị trường giải trí hiện đại ngày càng cạnh tranh, đòi hỏi các nhà sản xuất phim phải tối ưu hóa chi phí và tăng cơ hội thành công của các dự án. Việc sử dụng dữ liệu để hỗ trợ các quyết định chiến lược như ghép cặp diễn viên không chỉ giúp giảm thiểu rủi ro mà còn mang lại lợi ích trong việc xây dựng thương hiệu và thu hút khán giả.

Hơn nữa, sự gia tăng của các nền tảng phát trực tuyến và các xu hướng sáng tạo mới trong ngành điện ảnh đòi hỏi việc đánh giá và phân tích các yếu tố ảnh hưởng đến thành công của bộ phim phải chính xác và nhanh chóng hơn. Đây là thời điểm lý tưởng để áp dụng công nghệ phân tích dữ liệu vào các bài toán cụ thể như dự đoán khả năng hợp tác giữa các diễn viên.

Nghiên cứu này nhằm xây dựng một mô hình dự đoán khả năng hợp tác giữa các diễn viên trong các dự án phim tương lai. Thông qua việc sử dụng dữ liệu từ các nền tảng công khai như IMDb, mạng xã hội, và các thông tin ngành, nghiên cứu khám phá cách các yếu tố như lịch sử hợp tác, thể loại phim ưa thích, và mạng lưới xã hội ảnh hưởng đến khả năng ghép cặp. Đồng thời, nghiên cứu còn đưa ra những phân tích chi tiết nhằm minh họa tầm quan trọng của việc khai thác tiềm năng của dữ liệu lớn và học máy trong ngành giải trí.

1. **METHOD**
   1. ***OBJECTIVE***

Mục tiêu của nghiên cứu này là phát triển một mô hình dự đoán dựa trên bộ dữ liệu tự thu thập được, có khả năng đánh giá và xác định mức độ phù hợp của các diễn viên cho các dự án phim trong tương lai. Nghiên cứu hướng tới việc tối ưu hóa quá trình tuyển chọn diễn viên thông qua việc khai thác dữ liệu lịch sử và các yếu tố ảnh hưởng khác. Bằng cách này, nghiên cứu không chỉ giúp tăng hiệu quả và độ chính xác của các quyết định sản xuất phim mà còn mở ra tiềm năng ứng dụng rộng hơn trong các ngành công nghiệp sáng tạo khác.

* 1. ***DATA COLLECTION***

Dữ liệu gồm có 1098 bộ phim với nhiều thông tin liên quan bên trong bộ phim, thu thập bằng cách crawl data trên web phim.

Bộ dữ liệu gồm những có cột như ‘Genre’, ‘Cast’

* 1. ***DATA PREPROCESSING***

Tiền xử lý dữ liệu chúng tôi đã tách những dữ liệu trong cột ‘Cast’, ‘Genre’ ra thành mỗi list vì trong 1 bộ phim có thể có nhiều thể loại và nhiều diễn viên và một vài bộ phim đạo diễn cùng có thể là diễn viên nên cũng loại bỏ tên đạo diễn ra.

* 1. ***MODEL IMPLEMENT***

Để thực hiện nghiên cứu, chúng tôi sử dụng hai cột chính trong tập dữ liệu: "Cast" (danh sách diễn viên tham gia) và "Genre" (thể loại phim). Phương pháp triển khai bao gồm các bước chính sau đây:

* Xây dựng đồ thị quan hệ:
* Dựa trên cột "Cast", chúng tôi xây dựng một đồ thị quan hệ giữa các diễn viên, trong đó mỗi cạnh biểu thị sự hợp tác giữa hai diễn viên trong cùng một bộ phim.
* Trọng số các cạnh được tính dựa trên số lần hợp tác hoặc mức độ tương đồng giữa các thể loại phim "Genre" mà họ tham gia.
* Phân cụm bằng thuật toán Louvain: Chúng tôi áp dụng thuật toán Louvain để phát hiện các cộng đồng trong mạng lưới diễn viên. Các cụm này thể hiện các nhóm diễn viên có khả năng hợp tác cao dựa trên lịch sử làm việc chung.
* Trích xuất đặc trưng: Các đặc trưng như số lượng cộng tác, thể loại phim phổ biến trong cụm, và mức độ trung tâm của mỗi diễn viên trong đồ thị được sử dụng làm đầu vào cho các mô hình học máy.
* Huấn luyện mô hình: Sử dụng các thuật toán sau để dự đoán khả năng hợp tác:
  + Random Forest: Một mô hình học máy mạnh mẽ và dễ triển khai, dùng để dự đoán dựa trên các đặc trưng trích xuất.
  + Graph Convolutional Networks (GCN): Một mô hình mạng nơ-ron được thiết kế đặc biệt để làm việc trên dữ liệu đồ thị, giúp tận dụng tối đa cấu trúc của mạng lưới diễn viên.
  + Jaccard Similarity: Được sử dụng để tính toán mức độ tương đồng giữa các cặp diễn viên dựa trên tập hợp thể loại phim họ đã tham gia.
  + Adamic-Adar Index: Một chỉ số trong lý thuyết mạng dùng để đánh giá khả năng hai nút trong đồ thị (diễn viên) có thể liên kết với nhau.
* Đánh giá mô hình: Chúng tôi sử dụng các chỉ số như độ chính xác (accuracy), F1-score và AUC-ROC để đánh giá hiệu suất của từng mô hình trên tập kiểm tra.

1. **RESULTS**
2. **DISCUSSION**

REFERENCES

|  |  |
| --- | --- |
|  | T. V. D. V. Ljiljana Despalatović, "Community structure in networks: Girvan-Newman algorithm improvement," *IEEE Xplore logo - Link to home,* 2014. |
|  | S. C. B. G. Sarthak Giri, "Analyzing Social Networks of Actors in Movies and TV Shows," p. https://arxiv.org, 2024. |
|  | S. N. D. S. K. F. P. G. J Krauss, "Predicting Movie Success and Academy Awards through Sentiment and Social Network Analysis," 2008. |
|  | D. Shen, "Movie Box office Prediction via Joint Actor Representations and Social Media Sentiment," p. https://arxiv.org, 2020. |
|  | K. Z. Michael T. Lash, "Early Predictions of Movie Success: The Who, What, and When of Profitability," p. https://www.tandfonline.com, 2016. |
|  | C. J. Costa, "From Screen to Success: InvestigatingMachine Learning Algorithms to Predict Movies'Success, Based on Actor Features," *Lisbon School of Economics & Management,* 2024. |
|  | K. Z. H. W. X. S. &. B. Z. Yujia Liu, Content Matters: A GNN-Based Model Combined with Text Semantics for Social Network Cascade Prediction, 2021. |
|  | A. K. Awasthi, A. K. Garov, M. Sharma and M. Sinha, "GNN Model Based On Node Classification Forecasting in Social Network," p. https://ieeexplore.ieee.org, 2023. |
|  | X. Que, F. Checconi, F. Petrini and J. A. Gunnels, "Scalable Community Detection with the Louvain Algorithm," *IEEE Xplore logo - Link to home,* p. https://ieeexplore.ieee.org, 2015. |
|  | S. Ghosh, M. Halappanavar, A. Tumeo, A. Kalyanaraman, H. Lu and D. Chavarrià-Miranda, "Distributed Louvain Algorithm for Graph Community Detection," *IEEE Xplore logo - Link to home,* p. IEEE.org, 2018. |
|  | S. H. A. H. M. S. F. S. N. B. A. Nur Nasuha Daud, "Applications of link prediction in social networks: A review," *ScienceDirect,* 2020. |