



Institut de Recherche
en Informatique de Toulouse
CNRS - INP - UT3 - UT1 - UT2J



UNIVERSITÉ
TOULOUSE III
PAUL SABATIER



Université
de Toulouse

PhD thesis in data science and machine learning for film analysis

General information

Contract type: Fixed-term contract

Level of qualifications required: Master degree or equivalent

Fonction: PhD scholarship of 3 years

Level of experience: Recently graduated

Domain: Multimedia Analysis, Machine Learning, Computer Vision, Data Science

Location:

Université Toulouse 3 Paul Sabatier, Laboratoire IRIT
18 Route de Narbonne
31062 Toulouse, France

Starting date: October 2022

Duration of contract: 3 years

Contact/Recruiter:

Prof. Florence Sèdes, Full Professor in Computer Science

Florence.Sedes@irit.fr

<https://www.irit.fr/~Florence.Sedes/>

Co-supervised with Université Côte d'Azur, CNRS, Laboratoire I3S:

Dr. Lucile Sassatelli, Associate Professor with Habilitation, Junior Fellow of IUF

lucile.sassatelli@univ-cotedazur.fr

<https://www.i3s.univ-cotedazur.fr/~sassatelli/>

Instruction to apply:

Send to above email address: CV, motivation letter detailing why you are interested in this specific position, master transcripts, 3 reference names. Any additional code repositories and/or writing samples would also be appreciated.

Any incomplete application will be automatically discarded.

About Université Toulouse 3 and IRIT

<https://www.univ-tlse3.fr/>

<https://www.irit.fr/> , <https://www.irit.fr/en/home/>

Title

Data modeling and graph neural networks for multimodal film analysis

Context and objective

Films, series, and the characters they depict shape our collective imagination and perception of sociological constructs, such as gender, race, and class. In film gender studies, the concept of “male gaze” [Mul89] refers to the way the characters – especially women – are portrayed on-screen as

objects of desire rather than subjects. In contrast, the “female gaze”, recently re-defined by Brey [Bre20], consists in conveying the corporeal experience of a woman character. These analyses show that the disparities in how women and men are represented in visual media lie in the multimodal film signs (cinematographic, iconographic, textual, sound), through which gender is produced.

ANR TRACTIVE is a 4-year research project funded by the French National Research Agency (ANR) and involving 6 academic laboratories. The goal of TRACTIVE is to analyze the representation of gender in films. TRACTIVE regroups researchers from computer science, media studies, linguistics, and gender studies. We integrate Machine Learning (ML), linguistics, and qualitative media analysis in an iterative approach that aims to qualitatively identify the multimodal (visual and textual) discourse patterns of gender in films, and quantitatively reveal their prevalence. In this context, we seek a PhD student to integrate into our multi-disciplinary team, and work towards the specific following scientific objective:

Objective of the PhD thesis: Creation of a multimodal data model for film data and design of graph neural network approaches to analyze, from visual and textual data, how film characters are represented in relation to the other characters, in connection with their genders.

Graph Neural Networks (GNN) is a type of deep neural networks designed to process relational data, which in our case will be both visual and textual. The PhD thesis offers both research and hand-on experience on a wide variety of problems and multimedia processing tools, in developing code for open frameworks, and the opportunity to work on a project of large scale and strong societal impact. The cover letter must mention the project context and discuss its scope.

Methodology

To address this objective, the major scientific sub-domains in ML considered will be Deep Learning, Graph Neural Networks (GNN), and explainability. Deep Neural Networks (DNNs) indeed act as strongly flexible parametric function approximators made of compositions of tunable functions. This flexibility however raises the difficulty to explain the causes leading to the DNN's output. That is why so-called explainability is currently a main research challenge in deep learning, with existing works for image and text classification [Rib16] or drug discovery [Jim20].

The first year of the thesis will be dedicated to the creation of the dataset, through the definition of a data model to represent the film data and features that will be later analyzed, and the definition of the pre-processing workflow to extract the multimodal properties of films. In order to give feedback on the data sets and to validate the data model and the elicited features, a SQL-like query environment will be proposed to the end-user. From low-level processing leveraging the most recent computer vision tools on the visual data in conjunction with textual data from screenplays, we will extract higher-level features representing various properties including: (i) iconographic: with object/character appearances [Par20, Cao21] and gender, (ii) filmic: with camera movements and editing, and (iii) textual. These high-level features compose the meta-data that will be analyzed by ML models. The pre-processing workflow will also develop tools that enable film and media studies experts to add manual annotations and correct errors in the feature extraction. Finally, inspired by [Suc16], the high-level features will be embedded in a symbolic narrative model to design a declarative model integrating knowledge-based reasoning. Consequently, this will support inference

and querying for qualitative analysis of HSS, but also enable the generation of refined meta-data for the ML models to be designed.

Years 2 and 3 will focus on the relational analysis of the characters from image and text in relation to their genders. We will carry this out with GraphSAGE [Ham17], a class of Graph Neural Networks (GNN) recently gaining prominence. The core of GNNs is to generate node embeddings aggregating information over a variable-sized neighborhood. We can therefore design the proper GNN to represent characters and the camera holding the viewer's point of view with nodes, these nodes holding iconic and filmic properties. We will analyze the underlying interplays between women and men by looking at visual and textual representations that are characteristic of their gender. Explainability approaches for GNNs will identify the neighborhood of each node explaining the decision, and the important features of every node in this neighborhood. We will first resort to existing works like GNNExplainer in [Yin19] or [Bal19].

References

- [Bal19] Pope, P. E. et al.. Explainability Methods for Graph Convolutional Neural Networks. In IEEE International Conf. On Computer Vision and Pattern Recognition (CVPR), 2019. Available: https://openaccess.thecvf.com/content_CVPR_2019/papers/Pope_Explainability_Methods_for_Graph_Convolutional_Neural_Networks_CVPR_2019_paper.pdf
- [Bre20] I. Brey. Le regard féminin - Une révolution à l'écran. Editions de l'Olivier, France, 2020.
- [Cao21] Z. Cao, G. Hidalgo et al.. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. IEEE Trans. on Pattern Analysis and Machine Intell., vol. 43, no. 1, pp. 172-186, 2021. Available: <https://arxiv.org/abs/1812.08008>
- [Ham17] W. L. Hamilton et al.. Inductive Representation Learning on Large Graphs. NeurIPS, 2017. Available: <https://proceedings.neurips.cc/paper/2017/file/5dd9db5e033da9c6fb5ba83c7a7e9-Paper.pdf>
- [Jim20] Jiménez-Luna, J., Grisoni, F. & Schneider, G. Drug discovery with explainable artificial intelligence. Nature Machine Intelligence, 2, 573–584 (2020). Available: <https://www.nature.com/articles/s42256-020-00236-4>
- [Mul89] L. Mulvey. Visual pleasure and narrative cinema. In Visual and other pleasures, pages 14–26. Springer, 1989.
- [Par20] O. M. Parkhi et al.. Automated Video Face Labelling for Films and TV Material. IEEE TPAMI vol. 42, no. 4, 2020.
- [Rib16] Ribeiro, M. T., Singh, S. & Guestrin, C. “Why should I trust you?” Explaining the predictions of any classifier. In Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 1135–1144 (ACM, 2016).
- [Suc16] J. Suchan and M. Bhatt. The geometry of a scene: On deep semantics for visual perception driven cognitive film, studies. IEEE Winter Conf. on Applications of Computer Vision (WACV), 2016.
- [Yin19] R. Ying et al.. GNNExplainer: Generating Explanations for Graph Neural Networks. NeurIPS, 2019. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7138248/pdf/nihms-1062398.pdf>

Skills

Required:

- Essentials skills in data (base) modelling (UML),
- Excellent background in Machine Learning and Computer Vision,
- Excellent programming skills in Python,
- Ability to read, analyze, and compare to the state of the art in relevant fields,
- Willingness to tackle ongoing research challenges, propose innovative solutions, and validate results with scientific rigor,
- Good engineering skills to debug and compile repositories with source code in C++ and Matlab,
- Strong competence and prior experience with computer vision, image processing, text processing tools and libraries (e.g., opencv, ffmpeg, cimg, dlib, nltk, Lemur),
- Prior experience with machine learning libraries (scikit learn, tensorflow, torch) and algorithms,
- Good familiarity with data processing workflows for machine learning,
- Expertise in code repository management (git),
- Seeking, understanding, using, and writing technical documentation in French and English,
- Ability to understand needs of and communicate technical details to non-technical staff and project members.

Appreciated:

- Skills in web programming (HTML5, CSS, javascript, node),
- Knowledge of virtualization and software packaging (e.g., Docker, Singularity),
- A good level of written and spoken French and English

Remuneration

2100 euros gross monthly, 1700 euros net