

Master 2 Internship — Multimodal Video Recommendation

Improving Content Representation and Similarity for TV Series Recommendation

Context

Recommender systems for audiovisual content traditionally rely on user behavior such as ratings, viewing histories, or clicks. This makes it difficult to recommend new content for which little or no user data is available. The RECO+ project explores an alternative approach based solely on the content of audiovisual works, with the goal of producing relevant and explainable recommendations while addressing the cold-start problem.

A first study developed within RECO+ introduced a multimodal representation of TV series based on visual, audio, and textual features extracted from the first episode. The current system combines 77 interpretable descriptors capturing aspects such as visual aesthetics, editing and motion, sound characteristics, and narrative themes, and computes similarities between series using cosine distance.

This internship is the first of two complementary Master internships aiming to address the main limitations identified in this initial work. In particular, the current descriptors remain relatively coarse, while the similarity computation relies on a simple and non-contextualized distance.

Internship

The objective of this internship is to improve the multimodal representation of TV series and investigate more appropriate ways of measuring similarity between audiovisual works. The student will first study the state of the art in multimodal video representation, audiovisual content analysis, recommender systems, and metric learning. Building on the existing RECO+ pipeline, the work will then focus on refining visual, audio, and textual descriptors in order to better capture the aesthetic, narrative, semantic, and rhythmic dimensions of TV series.

A particular emphasis will be placed on the normalization and fusion of the different modalities, which remains an open question in the current approach. The student will also investigate similarity learning methods that can go beyond a uniform cosine distance, for example through learned metrics, contrastive learning, or other representation-learning approaches. The proposed methods will be implemented and evaluated experimentally within the existing recommendation framework, with particular attention to understanding which modalities and features contribute to meaningful similarities between series.

The internship combines software development, machine learning experimentation, and scientific research. It offers the opportunity to work on a concrete research problem at the intersection of multimedia analysis, multimodal learning, and recommender systems.

Candidate profile

We are looking for a Master 2 student in Computer Science with a strong interest in machine learning and multimedia and :

- Good programming skills in Python
- Knowledge of deep learning libraries
- Familiarity with Computer Vision, audio processing or Natural Language Processing
- An interest in machine learning applied to multimedia and audiovisual content
- Good analytical skills and the ability to conduct and interpret experimental results
- Experience with metric learning, contrastive learning, or representation learning would be a plus.

Application

Applicants should send a CV and their academic transcripts for the last two academic years to guinaudeau@lisn.fr

Level: Master 2 / final-year Master's student in Computer Science

Duration: 5 months

Location: Laboratoire Interdisciplinaire des Sciences du Numérique - Orsay

Starting date: January - February 2026