

TV genre classification system

Overview

Automatic genre classification is an effective solution to describe semantic properties of multimedia data. The TOSCA-MP developed a framework to classify the genre of TV programmes. For this purpose, four multimodal vectors, including both low-level perceptual descriptors, and higher level human-centered features are employed. These vectors serve as the input of a parallel neural network system that performs classification of input TV content into seven genre classes.

In depth description

The TOSCA-MP genre classification framework that is able to discern between seven TV genres, named commercials, newscasts, weather forecasts, talk shows, music video clips, animated cartoons and football match videos. These genres are fairly representative of the programme formats that are currently produced and distributed either through the traditional distribution channels, such as broadcast, cable and satellite, or through new platforms like the Internet or mobile phones.

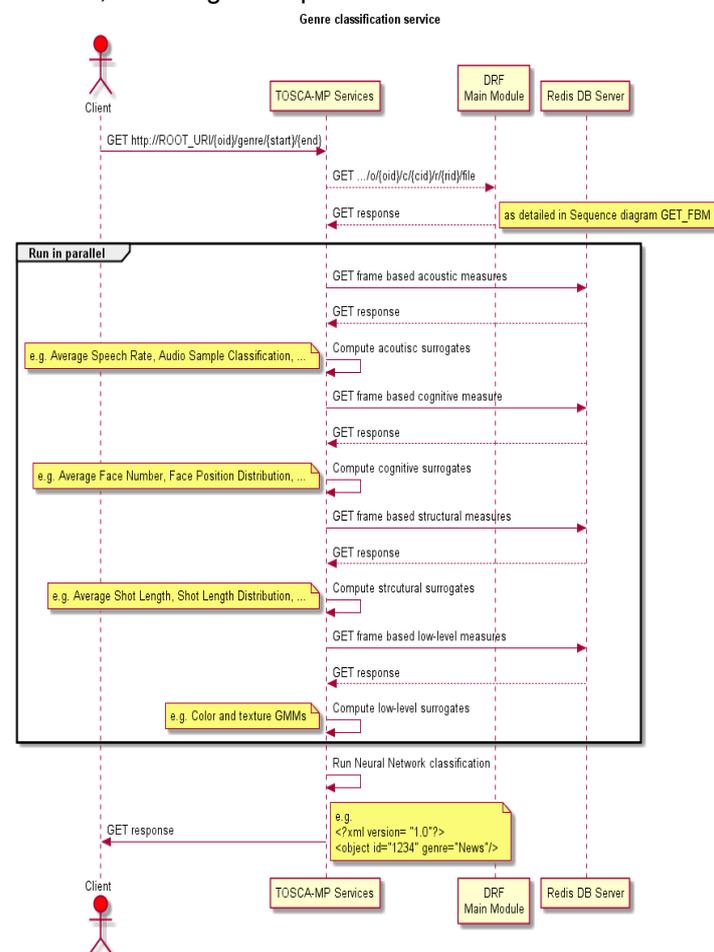


Figure 5. Sequence diagram for genre classification.

The developed system is based on two foundations: (i) *Multimodal content analysis* to derive a compact numerical representation of the multimedia content; and (ii) *Neural Network training process* to produce a classification model from those pattern vectors. The system is multimodal in that it uses features automatically extracted from different media sources. These include visual-perceptual information (colour, texture and motion), structural information (shot length, shot distribution, shot rhythm, shot clusters duration and saturation), cognitive information (face properties, such as number, positions and dimensions) and acoustic information (transcribed text, sounds). Such features are extracted on a frame-by-frame basis (e.g. there is one hue histogram for each video frame) and stored in a Redis database, as explained in Section 0. Once all the metadata are extracted and registered in Redis, genre

classification can be performed, as illustrated in Figure 5. Let *oid* the identifier of the video object to be classified, and *start*, *end* the frame interval on which classification has to be performed. Firstly, the metadata for each of the information sources (i.e. cognitive, structural, acoustic and visual) are requested to Redis through the DRF. Starting from these metadata, a set of feature surrogates (i.e. feature vectors) is computed and used as the input for a parallel Neural Network system that performs the genre classification.

Potential fields of Application

The result can be applied as a software framework specifically targeted at the implementation of semantic information extraction processes for media (pre) processing, annotation and retrieval.

Possibilities for exploitation

Exploitation of the result is on a bilateral licensing agreement basis. Licensing information is still to be defined by the owner. Some software integration (e.g. Neural Network retraining) may be needed to ensure the applicability of this result to different domains of interest.

Further Information

Further technical information is available in TOSCA-MP confidential Deliverable D2.2 “Automatic Metadata Extraction and Enrichment”.

Contact Person

Dr. Alberto MESSINA

RAI – Radiotelevisione Italiana - Centre for Research and Technological Innovation

Corso Giambone, 68, I-10135

Torino – ITALY

alberto.messina@rai.it