



UBIQUITOUS
KNOWLEDGE
PROCESSING

Ubiquitous Knowledge Processing Lab

Prof. Dr. Iryna Gurevych

Technische Universität Darmstadt

Hochschulstraße 10, 64289 Darmstadt

<https://www.ukp.tu-darmstadt.de>

Automatically Analyzing Audiovisual Content for Applications in Psychology and Pedagogy

Responsible: Prof. Dr. Iryna Gurevych, Pedro Santos

Usage: Automatic bootstrapping of manual analysis of audiovisual content from the Digital Humanities and automatic preprocessing of audiovisual data for analysis of humanistic researchers

Research Disciplines: Psychology, Communication Sciences, Pedagogy.

Example Application: Automatic identification of personality of primary school students

Code/Documentation: Currently not available publicly. Please contact the responsible(s) of this workflow to get access to code and/or documentation

Used in the following CEDIFOR Pilot Project: [Automatically Extracting Audiovisual Features to Analyze Videos in Educational Research](#)

1. Hypotheses Formulation

Many psychological constructs are externalized by the so-called external cues, where each different cue can have a different importance depending on the construct under analysis, according to the Brunswik's Lens Model. Some of those cues can only be captured using audio- and/or video data. Since manually analyzing the relationship between certain cues and psychological constructs on a larger sample is not feasible, automatic methods need to be applied. In the first step, a hypotheses about this relationship need to be formed, and in later steps empirically verified.

Example: The personality of students is externalized by their non-verbal behavior in daily classroom situations.

Best Practice: Since the identification of important cues depends on the psychological construct under analysis, it is necessary to have a well-defined method from the humanities discipline being employed to obtain a good intuition and evidence for the task under analysis.

2. Data Generation

This step requires to identify suitable data for the analysis of the phenomena that should be analyzed (defined in step 1).

Example: Video recording of students working spontaneously in daily activities, e.g. class room or lab situations during physics experiments.

Best Practice: It is important that the quality of the recordings is good enough to allow further processing of the audiovisual content.

3. Data Annotation

The phenomena under analysis must be coded according to a precisely defined model (e.g. using finite categories). The result is the so-called "gold standard".

Example: Questionnaires for assessing the personality of students, such as the BFI-10 (self-evaluation).

Best Practice: The questionnaires should be concise and not tedious, such as the BFI-10, otherwise the subjects under analysis might not be fully engaged when filling up the questionnaires.

4. Experimental Processing/Exploration

Here, computer vision techniques are employed together with machine learning methods to automatically identify the psychological constructs under analysis.

Example: Since a video is a sequence of images, tools for extracting all the frames from a video file are used, such as [FFmpeg](#). The next step after segmenting the video into images is to extract features from each image present in the video sequence. These features can be raw features, such as the pixels values from the RGB (Red-Green-Blue) channels, or high-level features such as facial landmarks, which can be extracted using tools like the [Cambridge Face Tracker](#). The last step is the application of machine learning algorithms, such as Support Vector Machines using [Weka](#) or artificial neural networks using [Theano](#) or [Keras](#).

Best Practice: This step requires programming knowledge and basic experience with machine learning techniques (application, parametrization). It should be carried out by a computer science expert.

5. Data Analysis and Interpretation

This step needs to be linked to the hypotheses formed in step 1, by discussing whether the defined cues are able to predict the psychological constructs under analysis. Error analysis is employed to understand which cases are not correctly identified by the automatic procedure in step 4.

Example: Some but not all personality traits (as externalized by the questionnaires in step 3) can reliably be predicted by the automatic procedure in step 4. Misclassified instances need to be identified and analyzed, so that error patterns in the automatic procedure can be found. During test phase, instances should have assigned an ID, so that the indexes of the misclassified instances can be linked to the instances' IDs.

Best Practice: It is important to store the outcome of the classifier in a format that makes the further analysis as convenient as possible. Most machine learning frameworks such as [Weka](#), [Scikit-learn](#), [Theano](#) or [Keras](#) offer this feature.

6. Evaluation

The best approach is to use a training, development and testing set approach, however if the number of instances is small, which is something quite common with data from the humanities, then a cross-validation setup should be employed.

Example: Leave-one-student-out cross-validation, where all the students are used for training and one student is left for testing (repeatedly).

Best Practice: Avoid having bias in the sets, so instances which are present in the training set do not have similar instances present in the testing set. The balance of the labels should be also similar in the creation of the sets.

7. Data Storage

All generated data should be securely stored to allow for reproducibility of the results (step 6).

Example: The student recordings and questionnaires are stored on a secure institutional data management server.

Best Practice: Depending on privacy issues, license and usage of the data, the video recordings (step 2), the annotated data (step 3) as well as a means to map between them need to be stored internally or publicly accessible. The formats used should be open source rather than proprietary.