EDITORIAL

Special issue on data-driven personalisation of television content

Lyndon J. B. Nixon[1] · Jeremy Foss[2] · Vasileios Mezaris[3]

Television content is no longer consumed only via traditional, linear TV broadcasting. The aim of this Special Issue on "Data-driven Personalisation of Television Content" is to address the increasing importance and relevance of richly granular and semantically expressive data about TV and immersive audiovisual content in the media value chain. Such data need software, specifications, standards, and best practices for extraction, modelling, and management before it can be meaningfully reused in new, innovative services for TV or other immersive audiovisual settings (e.g., 360° video in AR or MR). For this, the topics of interest to this Special Issue include but are not limited to:

- Content understanding and summarization (e.g., to provide highlights of a program according to a specific user, theme, or channel).
- Recommendation and scheduling across publication channels (broadcast, streaming, and social networks).
- In-stream personalisation of content (both spatial and temporal modification of text, audio, and video).
- Personalised and adaptive presentation for various media experiences, including user-user or network-user delivery using interworking media presentation formats.

Following an open call for papers and a rigorous peer review of the received contributions, four papers were finally accepted for inclusion in this Special Issue.

The paper entitled "Understanding Videos with Face Recognition: A Complete Pipeline and Applications", co-authored by P. Lisena, J. Laaksonen, and R. Troncy, deals with the topic of content understanding; in particular, understanding who are the (famous) people appearing in a video's scenes. The paper presents a complete pipeline for face recognition of celebrities, using a combination of state-of-the-art techniques for detecting faces, extracting face representations (embeddings) and training celebrity face classifiers. The face recognition results obtained at the individual frame level are the combined over time using object tracking. For training the complete face recognition system, celebrity images are automatically crawled from the web. The paper shows that such a person-centric approach to video understanding is very useful for navigating in and making sense of large TV content collections, including contemporary TV news broadcasts as well as older archive footage that is used for historical research.

The paper entitled "Combining Semantic and Linguistic Representations for Media Recommendation", co-authored by I. Harrando and R. Troncy, studies the use of various textual representations, as well as semantic representations, in a content-based recommender system. Two recommendation tasks are considered: (i) user-specific, which is performed by suggesting new items to the user given a history of interactions, and (ii) item-based, i.e., based solely on content relatedness. The paper reports on a thought experimental investigation of different possible choices in designing such a recommender system, e.g., examining how using automatically extracted content via ASR compares to using human-written content summaries. It then proposes deriving a semantic content representation by combining manually created metadata and automatically extracted annotations, showing that Knowledge Graphs, through their embeddings, constitute a suitable means for seamlessly integrating extracted knowledge with pre-existing metadata. The paper finally studies how combining both semantic and textual representations can lead to superior performance on both of the considered recommendation tasks.

The paper entitled "Towards Automatic Placement of Media Objects in a Personalised TV Experience", co-authored by B. Allan, I. Kegel, S. H. Kalidass, A. Kharechko, M. Milliken, S. McClean, B. Scotney, and S. Zhang, presents technologies enabling the in-stream personalization of video content. Specifically, the paper presents a Proof-of-Concept automated system that allows a new media object to be automatically positioned on a multi-angle broadcast video sequence without occluding any key action, thus enabling

✉ Vasileios Mezaris
bmezaris@iti.gr

1   MODUL Technology, Vienna, Austria

2   Birmingham City University, Birmingham, England

3   CERTH-ITI, Thessaloniki, Greece

additional graphic or video content to be used to personalise the broadcast for individual viewers. This system is showcased on team sports content, in particular video of football matches. To achieve its goal, the proposed system initially uses a DNN-based object detection algorithm for detecting the players and the ball. Following the identification of the region where the action takes place within each frame, the remaining space within the frame sequence is analysed and, considering also some broadcaster-defined templates for the placement of inserted objects, suitable locations for object insertion are proposed. The paper shows that the proposed system is capable of handling video sequences with multiple camera angles and proposing the start time and duration of inserted media objects without these insertions to occlude any key action.

Finally, the paper entitled "Data-driven personalisation of Television Content: A Survey", co-authored by L. Nixon, J. Foss, K. Apostolidis, and V. Mezaris, is a comprehensive survey of technologies, tools, and other resources that contribute to the vision of TV broadcasting being personalised and personalisation being data-driven. The paper starts by surveying various classes of technologies from the scientific literature that can contribute to this vision, ranging from technologies for TV content decomposition (temporal video fragmentation, object detection) and annotation (classification of media assets, annotation with classes or instances) to TV content re-purposing and personalization. The latter classes of technologies cater to the needs of finding the right content for different users/usages, e.g., by means of cross-modal content representation and retrieval; transforming this content using techniques such as video summarization, highlight-detection, super-resolution, or aspect-ratio adaptation; personalising the TV content by, e.g., inserting new content inside pre-existing video streams. The paper then continues with surveying deployable applications that contribute to the vision of data-driven personalised TV experience, i.e., relevant tools and Web services, and also presents the standards that have been defined so far to support activities in the media trading and value chain and can be utilised in enabling data-driven personalised TV. The paper concludes with a discussion on the current open problems for all relevant technologies and standards, proposing future research and innovation directions.

We hope that this Special Issue contributes to a better understanding of the current state of research and technology in personalised television.