

Automatic Content Curation of Visual Heritage

Authors : Delphine Ribes Lemay, Valentine Bernasconi, André Andrade, Lara DéFayes, Mathieu Salzmann, Frédéric Kaplan, Nicolas Henchoz

Abstract : Digitization and preservation of large heritage induce high maintenance costs to keep up with the technical standards and ensure sustainable access. Creating impactful usage is instrumental to justify the resources for long-term preservation. The Museum für Gestaltung of Zurich holds one of the biggest poster collections of the world from which 52'000 were digitised. In the process of building a digital installation to valorize the collection, one objective was to develop an algorithm capable of predicting the next poster to show according to the ones already displayed. The work presented here describes the steps to build an algorithm able to automatically create sequences of posters reflecting associations performed by curator and professional designers. The exposed challenge finds similarities with the domain of song playlist algorithms. Recently, artificial intelligence techniques and more specifically, deep-learning algorithms have been used to facilitate their generations. Promising results were found thanks to Recurrent Neural Networks (RNN) trained on manually generated playlist and paired with clusters of extracted features from songs. We used the same principles to create the proposed algorithm but applied to a challenging medium, posters. First, a convolutional autoencoder was trained to extract features of the posters. The 52'000 digital posters were used as a training set. Poster features were then clustered. Next, an RNN learned to predict the next cluster according to the previous ones. RNN training set was composed of poster sequences extracted from a collection of books from the Gestaltung Museum of Zurich dedicated to displaying posters. Finally, within the predicted cluster, the poster with the best proximity compared to the previous poster is selected. The mean square distance between features of posters was used to compute the proximity. To validate the predictive model, we compared sequences of 15 posters produced by our model to randomly and manually generated sequences. Manual sequences were created by a professional graphic designer. We asked 21 participants working as professional graphic designers to sort the sequences from the one with the strongest graphic line to the one with the weakest and to motivate their answer with a short description. The sequences produced by the designer were ranked first 60%, second 25% and third 15% of the time. The sequences produced by our predictive model were ranked first 25%, second 45% and third 30% of the time. The sequences produced randomly were ranked first 15%, second 29%, and third 55% of the time. Compared to designer sequences, and as reported by participants, model and random sequences lacked thematic continuity. According to the results, the proposed model is able to generate better poster sequencing compared to random sampling. Eventually, our algorithm is sometimes able to outperform a professional designer. As a next step, the proposed algorithm should include a possibility to create sequences according to a selected theme. To conclude, this work shows the potentiality of artificial intelligence techniques to learn from existing content and provide a tool to curate large sets of data, with a permanent renewal of the presented content.

Keywords : Artificial Intelligence, Digital Humanities, serendipity, design research

Conference Title : ICDH 2021 : International Conference on Digital Heritage

Conference Location : London, United Kingdom

Conference Dates : November 18-19, 2021