Detecting the Structure of Videos by Extracting Features from Time Series User Interaction Data

Yakura, Hiromu

Mainstream content on the Internet is becoming less text-based, with more emphasis given to videos. As a result, the demands on automatic summarization are rapidly increasing. In the summarizing of music videos, the audio signal is often utilized. In this research, a novel method is proposed that estimates the climax portion of a music video by analyzing time-series comments from users, a technique that will be applicable to a wide range of videos. Frequent and specific words contained in comments were extracted and normalized by morphological analysis. To extract specific slang words correctly, a new dictionary which complied slang words from the wiki including Japanese Internet slang words was prepared. Feature vectors were calculated based on the frequency distribution of extracted words for each section determined by audio signal analysis. The system scored the distance between the vectors and feature values, and estimated the climax portion based on the results. For the evaluation, 100 videos and their comments from nicovideo.jp, the largest Japanese video-sharing platform, were used. The proposed method correctly estimated the climax portions for 89% of the dataset, whereas the accuracy was respectively 76% and 79% when only the audio signal and unweighted comment quantity with section data detected by audio signal analysis were used. The method drastically improves the precision of the climax portion estimation. Since the method is based on user-interactions, it can be applied not just to music videos but to various genres of videos on the Internet by introducing the proposed feature vectors. Furthermore, novel means for determining suitable positions for inserting advertisements could utilize time-series user-interaction data such as SNS using this method.

Awards Won: