Comparative Analysis of different Text Segmentation Algorithm on Arabic News stories

The undertaking of message division addresses a significant stage in numerous applications and keeping in mind that much work has been completed to address this errand for the English language, work on message division for different dialects is as yet slacking. In this paper a similar examination of three distinct message division calculations on Arabic reports is introduced. To survey how well every calculation deals with Arabic reports, each was applied on an Arabic Reuters report dataset and the outcomes were looked at. The work in this paper likewise portrays a blend of two of these calculations that was found to create improved results than any of the introduced individual calculations. It likewise presents a bunch of mistake decrease channels that were found to diminish division blunders in the recognition of lines in Arabic based reports fundamentally.

Fusion of News Reports using Surface-Based Methods

Occasions happened in the genuine word covering by news reports from various sources. Each report for the most part contains data that is tracked down in others yet may likewise contain remarkable data. To realize all the data about a specific occasion, a client should peruse every one of the various reports. This is a duplication of exertion since most data will be rehashed in the various reports. In our examination, we endeavor to meld news reports about a similar occasion into a solitary lucid record disposing of reiteration however safeguarding all the data contained in the source reports utilizing just surface-based techniques. Data in every news report is addressed by a bunch of element relationship diagrams. The diagrams addressing each report are then converged into a solitary chart while monitoring the source sentences. The melded report is created utilizing the maximally expressive arrangement of sentences - the sentences that convey most data about the substances and their connections in the news report and guaranteeing that all elements and connections are communicated in the combined archive. Our Record combination framework was assessed utilizing a bunch of information reports downloaded from MSNBC News that refer to their sources and utilizing human assessment. We demonstrate the way that our framework can catch the greater part of the data found across various source reports while keeping up with meaningfulness. Catchphrases Archive combination; news; theoretical diagrams; element connection charts.

News Hotspots Detection and Tracking Based on LDA Topic model

With the fast spread of Web and the portable web, the quantity of information pages is expanding rapidly as well as the substance of information turns out to be exceptionally unique. It's challenging for typical clients to get explicit data contained in a mass of information streams. Thus, it's of extraordinary examination importance to concentrate on the most proficient method to break down enormous news, recognize and follow news areas of interest consequently. This exploration proposes to apply LDA (Inert Dirichlet Designation) model to the utilization of subject identification and following. The news stories gathered by crawlers are displayed by the LDA model in a type of record point word dissemination. We propose a technique to process the intensity of points in view of the conveyance and to recognize the news areas of interest. Also, we track the advancement of the subject patterns in various time-cuts. Jenson-Shannon distance is utilized to quantify the closeness between points to recognize subject legacy and point transformation. We directed probes a dataset comprising of 3462 news texts from news entryways. The outcome uncovered that the proposed model has a decent impact both in identifying areas of interest and finding significant effective development patterns. Catchphrases — LDA theme model; news reports; areas of interest; location; following;

Extracting semantic information from news and sport video

Improvement of frameworks supporting successful recovery by satisfied of recordings expects to play out a wide range of procedure on video transfers, including fleeting division, examination of the sound and video tracks, ID, and acknowledgment of text. Low level elements are then handled to give some more elevated level portrayal of video content, as the majority of the client inquiries are normally connected with more elevated level punctuation and semantics, rather on the lower-level lexical level. Also, the particularity of various application areas involves those various arrangements be taken on in various settings. This might influence both the decision of low-level elements to be separated, as well as the displaying of explicit space information required address the issue of more significant level of semantics. In this paper we will write about our involvement with the application settings of information and sports recordings. We will show arrangements embraced to adapt to explicit necessities of various application spaces.

Twitter vs News: Concern Analysis of the 2018 California Wildfire Event

During disasters, discover people's concerns dynamically is crucial to disaster rescue and relief. In this paper, we propose a social media-based framework to analyse people's concerns, to access the importance and to track the dynamic changes of these concerns. To better understand people's concerns across platforms and to monitor the dynamics, we make comparisons between Tweets and news on the mentioned aspects and disclosed some interesting findings. Specifically, we take 2018 Camp Fire, the most destructive wildfire on record in history of California as a case study. We find that despite their keen attentions towards the disaster, social media, and news media focus on different aspects of the disaster, so are the contents and dynamic changes of their concerns

Overlay Text Extraction From TV News Broadcast

The text data present in overlaid bands convey brief descriptions of news events in broadcast videos. The process of text extraction becomes challenging as overlay text is presented in widely varying formats and often with animation effects. We note that existing edge density-based methods are well suited for our application on account of their simplicity and speed of operation. However, these methods are sensitive to thresholds and have high false positive rates. In this paper, we present a contrast enhancement based pre-processing stage for overlay text detection and a parameter free edge density-based scheme for efficient text band detection. The second contribution of this paper is a novel approach for multiple text region tracking with a formal identification of all possible detection failure cases. The tracking stage enables us to establish the temporal presence of text bands and their linking over time. The third contribution is the adoption of Tesseract OCR for the specific task of overlay text recognition using web news articles. The proposed approach is tested and found superior on news videos acquired from three Indian English television news channels along with benchmark datasets.

A DETECTION-BASED APPROACH TO BROADCAST NEWS VIDEO STORY SEGMENTATION

A detection-based paradigm decomposes a complex system into small pieces, solves each subproblem one by one, and combines the collected evidence to obtain a final solution. In this study of video story segmentation, a set of key events are first detected from heterogeneous multimedia signal sources, including a large-scale concept ontology for images, text generated from automatic speech recognition systems, features extracted from audio track, and high-level video transcriptions. Then a discriminative evidence fusion scheme is investigated. We use the maximum figure-of-merit learning approach to directly optimize the performance metrics used in system evaluation, such as precision, recall, and F1 measure. Some experimental evaluations conducted on the TRECVID 2003 dataset demonstrate the effectiveness of the proposed detection-based paradigm. The proposed framework facilitates flexible combination and extensions of event detector design and evidence fusion to enable other related video applications.

Beaming into the News: A System for and Case Study of Tele-Immersive Journalism

This work was funded under the European FP7 Integrated Project BEAMING (contract number 248620). We thank Carlota Crusafon (UB) and Aitor Rovira (UCL) for helping to set up the transporter system at UCL and Evans Hauser for helping at the UB destination. We also thank Marie Schutgens and Bregtje van der Haak of VPRO Tegenlicht, and UBTV for providing film footage. We also thank Engineered Arts Ltd for the loan of a Robothespian at UCL, and NVIS for the loan of an NVIS SX111 at USC. We would like to thank Dr Javier Martínez Picado, Dr Perla Kaliman, and Mr Miguel Aguilar for taking part in the beaming experiences.

NEWS VIDEO ANALYSIS BASED ON IDENTICAL SHOT DETECTION

The paper presents a method to detect identical video segments from video footages in broadcasted video archives, and its application to news video analysis. We define identical video segments as distinctively similar video segments. After giving the definition of identical images, efficient algorithm to detect identical shots in videos is shown. Its effectiveness is shown by applying the algorithm to news video analysis. As the first experiment, the algorithm successfully locates identical shots derived from the same video materials shared by semantically related topics. This obviously is useful for topic tracking by a visual cue. Then the second experiment shows that aggregating detected identical shots can tell news video structure, i.e., can locat~filler shots such as opening shots, anchor shots, weather forecast. etc. The result is quite helpful for news video parsing. The experiments reveal that, although the method does not use any news-specific a priori knowledge, it could be used as a powerful tool to explore useful knowledge from large-scale news video archives.

Breaking News Detection and Tracking in Twitter

Twitter has been used as one of the communication channels for spreading breaking news. We propose a method to collect, group, rank and track breaking news in Twitter. Since short length messages make similarity comparison difficult, we boost scores on proper nouns to improve the grouping results. Each group is ranked based on popularity and reliability factors. Current detection method is limited to facts part of messages. We developed an application called “Hot stream” based on the proposed method. Users can discover breaking news from the Twitter timeline. Each story is provided with the information of message originator, story development and activity chart. This provides a convenient way for people to follow breaking news and stay informed with real-time updates. Keywords-Twitter, Topic Detection and Tracking, Real-time text-mining, Information Retrieval.

Hash tagger+: Efficient High-Coverage Social Tagging of Streaming News

News and social media now play a synergistic role and neither domain can be grasped in isolation. On one hand, platforms such as Twitter have taken a central role in the dissemination and consumption of news. On the other hand, news editors rely on social media for following their audience's attention and for crowd-sourcing news stories. Twitter hashtags function as a key connection between Twitter crowds and the news media, by naturally naming and contextualizing stories, grouping the discussion of news, and marking topic trends. In this work, we propose Hash tagger+, an efficient learning-to-rank framework for merging news and social streams in real-time, by recommending Twitter hashtags to news articles. We provide an extensive study of different approaches for streaming hashtag recommendation and show that pointwise learning-to-rank is more effective than multi-class classification as well as more complex learning-to-rank approaches. We improve the efficiency and coverage of a state-of-the-art hashtag recommendation model by proposing new techniques for data collection and feature computation. In our comprehensive evaluation on real data, we show that we drastically outperform the accuracy and efficiency of prior methods. Our prototype system delivers recommendations in under 1 minute, with a Precision@1 of 94 percent and article coverage of 80 percent. This is an order of magnitude faster than prior approaches and brings improvements of 5 percent in precision and 20 percent in coverage. By effectively linking the news stream to the social stream via the recommended hashtags, we open the door to solving many challenging problems related to story detection and tracking. To showcase this potential, we present an application of our recommendations to automated news story tracking via social tags.

Our recommendation framework is implemented in a real-time Web system available from insight4news.ucd.ie.

Learning Approaches for Detecting and Tracking News Events

THE RAPIDLY GROWING AMOUNT of electronically available information threatens to overwhelm human attention, raising new challenges for information retrieval (IR) technology. Traditional query-driven retrieval is useful for content-focused queries but deficient for generic queries such as "What happened?" or "What's new?"

Consider, for example, a person who has just returned from an extended vacation and needs to find out quickly what happened in the world during her absence. Reading the entire news collection is a daunting task, while generating specific queries without any knowledge of recent events is rather unrealistic. Or consider a foreign-policy specialist who wants to study the Asian economic cri sis, including precursor and consequent events. A keyword-based search on the query "Asian economy crisis" would most likely miss many relevant stories about the stock market crashes in Indonesia or Korea, banking-sector insolvency in Japan, or Jusuf Habibi's rise to power in Indonesia.

In other words. query-based retrieval is useful when you know more precisely the nature of the events or facts you're seeking. It is less useful when you want specific information but can only formulate a larger cate gory sharing few if any terms with the potentially most useful texts. In short, retrieval based on immediate-content-focused queries is often insufficient for obtaining a variety of relevant stories and tracking the gradual evolution of events through time.

A SHOT BOUNDARY DETECTION METHOD FOR NEWS VIDEO BASED ON OBJECT SEGMENTATION AND TRACKING

As a critical step in many multimedia applications, shot boundary detection has attracted many research interests in recent years. The most of existing methods measure the similarity among video frames based on its low-level feathers. However, they are sensitive to the change in not only brightness, colour, motion of object, but also camera motions and the quality of video. This paper proposes an innovative shot boundary detection method for news video based on video object segmentation and racking. It combines three main techniques: the partitioned histogram comparison method, the video object segmentation and tracking based on wavelet analysis. The partitioned histogram comparison is used as the first filter to effectively reduce the number of video frames which need object segmentation and tracking. The unsupervised video object segmentation and tracking based on wavelet analysis is robust to those problems mentioned above. The efficacy of the proposed method is extensively tested with more than 3 hours of CCTV and CNN news programs, and that 96.4% recall with 97.2% precision have been achieve.