SITUATION AWARE TOPIC ANALYSIS FOR KEYNOTE/SPEAKER EVENTS

Rajagopal Srinivasan
Tuli Kundu
CD Venkatesh
Suyash Saluja

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation
Srinivasan, Rajagopal; Kundu, Tuli; Venkatesh, CD; and Saluja, Suyash, "SITUATION AWARE TOPIC ANALYSIS FOR KEYNOTE/SPEAKER EVENTS", Technical Disclosure Commons, (October 29, 2018)
https://www.tdcommons.org/dpubs_series/1615

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.
SITUATION AWARE TOPIC ANALYSIS FOR KEYNOTE/SPEAKER EVENTS

AUTHORS:
Rajagopal Srinivasan
Tuli Kundu
CD Venkatesh
Suyash Saluja

ABSTRACT

Techniques are described herein to provide speaker insights for a talk by correlating audience mood and engagement data (derived from video feed) with speech data. With this framework a speaker obtains a summary of topics that were relevant to the crowd. This technique helps a speaker to design future content based on the metrics and topics obtained using this framework. This solution can bring a lot of value to people who intend to design content for keynotes that will be relevant to the crowd.

DETAILED DESCRIPTION

The speaker is provided with a summary of topics that captivate the mood of the audience in an automated fashion. Our method involves identifying keywords (from audio) that resonate with the audience (e.g., elevated moods, engagement score, etc.) and directly mapping the keywords to the summary timeline so that a speaker knows which topics/phrases were appreciated by the audience.

At the end of the event, a speaker can look at the metrics generated and identify the topics relevant to the audience. This output helps the speaker understand the interests of the audience, while winnowing topics that are irrelevant.

This solution can be extended from company meetings to professional talks and events, ensuring that the speakers get a clear indicative way of designing topics for future presentations/talks.

Speech data is correlated with the audience mood and engagement statistics available from the video feed. This ensures that there are fewer words to process in real time, enabling identification of the topics and relevant moods close to real-time. This also creates a higher reliability as fewer text characters are processed without overloading the topic modelling techniques with entirely audio content (e.g., speech audio for an entire conference).
Figure 1 below illustrates an example for mapping the mood and engagement level of the audience to the relevant topic on the video timeline in real-time.

![Figure 1](image1.png)

Figure 2 below illustrates a summary of an event hosted by a speaker. The size of the word in this word cloud is directly proportional to the mood and engagement level of the audience, which indicates to the speaker the relevant topics of speech.

![Figure 2](image2.png)

Figures 1 and 2 above provide a reasonable summary about the talk and the topics relevant to the situation.
Figure 3 below illustrates engagement analysis for a speaker. In this example, there is an overall engagement score of 85%.
Figure 4 below illustrates an example Audio/Video (AV) frame and how the region of interest (ROI) in the frame is extracted. Audio extracted from this ROI is then submitted to the topic modelling engine.
Figure 5 below illustrates an example architecture
Data may be reported to a speaker as follows:

1) A computing language framework is used to read from a JavaScript Object Notation (JSON) payload generated by topic analysis. This JSON payload may contain information such as time, engagement score, topic list, etc.

2) A Hyper Text Markup Language (HTML) script may be rendered through a framework to an external Internet Protocol (IP) address so the speaker can view the dashboard.

Figure 6 below illustrates an example wireless architecture.
Figure 7 below illustrates an example wired architecture.

Figure 8 below illustrates an example flowchart.
Figure 9 below illustrates an example algorithm for topic modeling.

The algorithm proceeds as follows -

1) The system obtains input from any previous modules.
2) The system performs Regular expression (Regex) cleaning to remove unwanted words and patterns (e.g., fix multiple spaces or duplicate words from text to speech errors).
3) The system adds tags to words in order to mark prepositions, articles, and other unimportant classes of words for removal.
4) The system uses lemmatization to use context to reach the word’s lemma to avoid similar words existing in different forms (e.g., “run” and “running”).
5) After the preprocessing steps, the system calculates the Term Frequency (TF) scores for each term in the speech extract.
6) The system uses the TF scores to determine which words occur most frequently in the speech.
7) The system passes the text to a neural network based Named-Entity Recognition (NER) model to extract the classes or named entities in the text. The sixth and seventh steps may occur simultaneously.
8) The system assigns high weightage to words with higher TF scores and also having a named entity (e.g., person, organization, technology, etc.) attached thereto.
9) The system builds the topic list and sends the modified JSON string to the web interface.

The speech data will be recorded and corroborated with the mood/engagement score of the audience. When the mood/engagement score passes a certain threshold, the speech data is buffered for the past 45 seconds. This enables obtaining a high performance level to retrieve insights from the speech Application Programming Interface (API).
The system also performs topic modeling for a speaker event by extracting “audio worthy” data from the voice using correlated video data (the efficiency may be increased by reducing the number of frames processed by reducing sampling frequency to a few frames averaged over the session). The amount of speech processed is reduced to 45 seconds only (as discussed in the above paragraph) rather than processing all the voice content (enough to generate topics from nouns/technology words used in the phrase). By correlating the audience engagement video and speech data, only the relevant content is extracted and the unwanted data, which usually gets processed by speech-based topic modeling engine, is eliminated. This creates higher accuracies in topic modeling.

The techniques described herein leverage one or more features pertaining to emotion/sentiment recognition to extract the interesting part of audio from video. Accuracy can be achieved using proven frameworks. Cloud based frameworks may be useful as they can reduce the computation power required on the local machines thereby allowing for portable deployment of the solution.

In summary, this publication describes a technique for providing insights into a talk by correlating mood data based on video with speech data. With this framework a speaker obtains a summary of topics that are relevant to the crowd being addressed. Situational awareness through video content helps a speaker to design future content based on the metrics and topics obtained. This solution can bring a lot of value to people by helping them design their keynotes, speeches, etc. to stay relevant to the crowd.