

Technical Disclosure Commons

Defensive Publications Series

November 30, 2015

Method For Providing Metrics To Determine Transcription/Translation Quality

Dimitri Kanevsky

Mark Mao

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

Recommended Citation

Kanevsky, Dimitri and Mao, Mark, "Method For Providing Metrics To Determine Transcription/Translation Quality", Technical Disclosure Commons, (November 30, 2015)

http://www.tdcommons.org/dpubs_series/85



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

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.

METHOD FOR PROVIDING METRICS TO DETERMINE TRANSCRIPTION/TRANSLATION QUALITY

ABSTRACT

A system and method are disclosed for determining the transcription/translation quality of web video content based on metrics derived from indirect feedback of users on existing captioning. The method may take into account how often the closed caption (CC) option was activated by a user on videos and the number of times users stayed through the whole video content using the closed captions. The system can also be used in assessing quality of manual transcription for languages that do not have automated speech recognition and to validate acoustic and language models in machine translation/transcription.

BACKGROUND

A large number of videos are created and posted on the internet every day. These videos are provided with subtitles in original and foreign languages to be accessible by a user and compatible to run applications, such as indexing, searching, summarization, context analyses etc. Transcription and translation of videos are done via automated or manual means. Manual assessment of translation and captioning quality is done using human experts who would listen to audio and compare it with a transcription or translation. This method could be expensive and might require a lot of time. It may also be impractical because of the sheer volume of new audio/video data loaded onto the internet every day. Automated transcription and/or translation could therefore be useful. In many cases, transcription and/or translation subtitles that are created are of insufficient quality for one type of application while they may be acceptable for others. Thus, there is a need to identify transcription and/or translation quality to determine compatibility with different types of applications.

For languages for which speech recognition or machine translation are available, there are automated means to provide confidence scores, (for example, lattice posteriors) to judge

the transcription/translation quality. However, for languages that do not have automated translation means, judging transcription/translation quality becomes difficult. Currently available methods based on automated recognition or machine technologies are generated without knowledge of the ground truth and are hence not accurate enough for all types of applications. Thus, there is a need for a better method to determine transcription/translation accuracy.

DESCRIPTION

This disclosure presents a method and system for determining the transcription/translation quality of web video content based on metrics derived from indirect feedback of users on existing captioning. The method may take into account how often the ‘closed caption’(CC) option was activated by a user on videos and the number of times users stayed through the whole video content using the ‘closed caption’ option. The user’s behavioral characteristics are thus used to arrive at ratings as to the accuracy of transcription/translation.

In one scenario, where a significant percentage of users activate CC and watch the video completely, the system registers the quality of the transcription to be good. In an alternative scenario, where a user watches a popular video with the CC option activated for only a short time, or discontinues watching the video before its completion, the system registers that the user is not satisfied with corresponding transcription/translation quality.

The system separately determines if a user stopped watching a video due to bad quality of the video and not due to its transcription/translation quality, by baselining with % of users discontinuing viewing the video without the use of captions. For a same video, if $x\%$ of users watch the complete video without captions, and $y\%$ of users watch the complete video with captions, then if $y > x$, the transcription and/or translation is deemed to be good and if $y < x$, then it is deemed to be bad.

In some implementations, the system may require both y to be greater than x and the difference or ratio to be greater than some threshold value to classify the captioning to be good. There can be instances where the transcriptions are of good quality while translations are bad, or vice versa. Then the system could compare how often people dropped a video with translation vs. people who dropped watching a video with transcription, to determine the quality of transcription vs. translation. In some cases, the transcription in one language may have a better quality over another, which the system is configured to identify.

The above system, in addition to being used in machine-generated transcription, can also be used in assessing manual transcription for languages that do not have automated speech recognition systems. In other applications of the system and method, videos with high transcription scores can be used to validate acoustic and language models in machine translation/transcription.