October 11, 2016

Language Importance Score

Jonasz Pamula

Christoph Oehler

Aleksandar Stupar

Daniel Henell

LANGUAGE IMPORTANCE SCORE

ABSTRACT

A system and method are disclosed for classifying videos based on a Language Importance Score (LIS) indicating how important it is to understand a certain language to be able to understand a given video. The method extracts video features directly from the video after it has been uploaded rather than by analysing viewing patterns from logs. The level of LIS is classified as LOW, MID, or HIGH based on the degree of ease for a user to understand. Global/multilingual video platforms can use this system to optimize search to improve viewer experience effectively.

BACKGROUND

Watch time of a video resource is the total length of time that users spent watching the content. About 80% watch time of online media is from users outside the US. In this global environment, the concept of understandability of media plays a key role in its dissemination. Suggesting videos to viewers that they're not able to understand may lead to bad user experience. For example, videos without verbal text can be understandable by everyone, whereas a physics lecture demands interpretation. Thus, there is a need for a better method to indicate the importance of a certain language to understand a given video effectively.

DESCRIPTION

This disclosure presents a system and method for classifying videos based on a Language Importance Score (LIS) indicating how important it is to understand a certain language to be able to understand a given video. Classification on video features can be extracted directly from the video after it has been uploaded rather than by analysing viewing patterns from logs. The system can be used in a global/multilingual video platform.

LIS for a video is a value [0,1] that indicates the importance of understanding a certain language to be able to understand or benefit from the video. For a video v, let population A be the viewers who understand the language in v and population B the viewers who do not. LIS can be defined as the fraction of the watch time for population B: LIS=1-min(1, E[WT(v | viewer drawn from B)] / E[WT(v | viewer drawn from A)]). This definition assumes that these two populations are large and they are equally exposed to the video. The LIS may have a bias if the users use actual views of the video to compute the score because the population size could vary and the viewers who understand the video are more likely to be exposed to the video. Examples of LIS scores are as follows:

LIS = 1.0: No one watches the video if they do not understand the language.

LIS = 0.5: Half the watch time for a group of people who do not understand the language.

LIS = 0.0: The language is not of importance or the video does not contain any language.

The level of LIS can be classified as LOW, MID, or HIGH based on the degree of ease for a user to understand. Examples of low, medium, and high LIS genres are given below:

Low LIS genres: cat videos, fail compilations, Mr Bean sport, music

Medium LIS genres: Unboxing gadgets, Japanese game shows, instructional programs

High LIS genres: Lectures, vlogs, news, documentaries, instructional programs.

The method of classifying video features can be implemented by the following use cases:

User Language Profiles: Assign weight factors for languages of videos when aggregating the videos that a user has been watching.

Recommendations: Assign a weight factor for the language overlap between a video and a user. Recommend creators that their low LIS videos would benefit from translated metadata. Potentially enable auto translations for low LIS videos’ metadata.

Training Model: Edutainment products are used as training model for video classification, and logistic regression is implemented to do multiclass classification. To preserve the natural order of the classes (e.g. LOW, MID, HIGH – in that order), the method is implemented as a tree of binary classification problems as depicted in FIG.1. If classified as LOW or MID, then the video is auto translated. If MID_HIGH, then it is further classified between MID and HIGH. If it is MID, then it is classified for auto translation, and if HIGH, then auto translation is not used. More steps can be included in the method if a user wants finer granularity of classes.

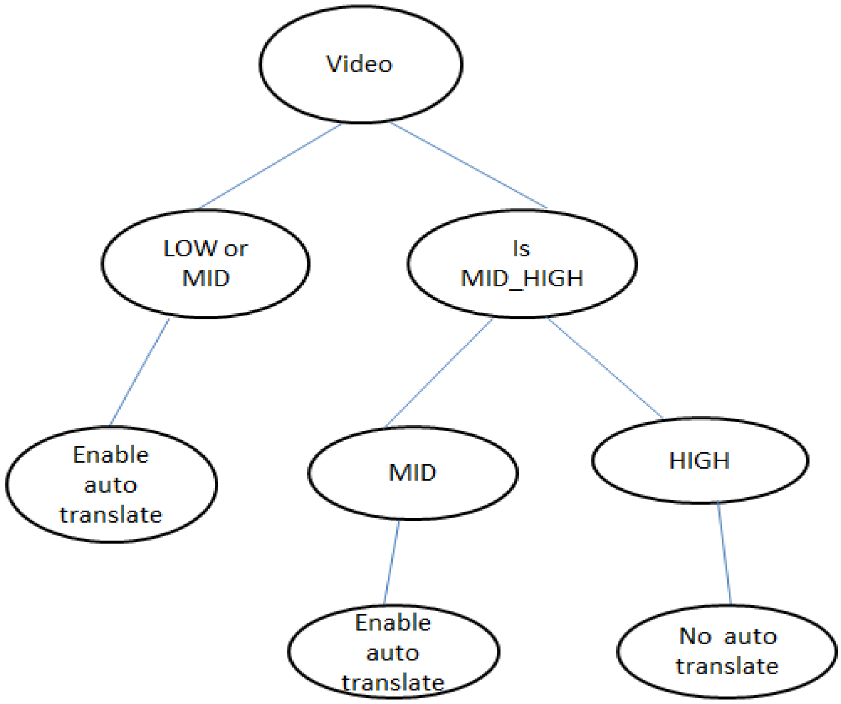

FIG. 1: Binary tree representation of LIS

Feature Set: Word clusters from metadata provided by an existing edutainment product can be used. Additionally, the amount of speech in a video is an important signal to identify videos with high LIS. A video may contain low level audio features, but may require numerous training iterations data due to high dimensionality. A low dimension audio track containing less speech can be created, which would require fewer training data than competing schemes.

Dataset Population: The population contains non-music videos, exemplarily with WT> X last 30 days. Music videos are excluded entirely from classification. The naive assumption is that music is universal and always has low LIS.

Eval Set Sampling: In the first iteration, a user can reuse the eval set for an edutainment product. The product eval set contains a certain number of videos that are sampled every quarter. The sample videos weighted by the last 30 days of watch time are used as training data.

Alternative implementation of sampling method might use uniform sampling from the dataset population and stratified sampling with automatically assigned product categories. A user can take more samples from categories with higher LIS variance. The dataset size for an eval set could use a certain number of videos. This disclosed system and method can optimize search to improve viewer experience effectively.