Data Mining Exam 2 - 4283724 Nadine Kanbier

Question 1

Overview classification scheme

To investigate the influence of video thumbnails on YouTube's recommendation algorithm, a good classification scheme is a prerequisite. Previous research has shown that the selection of a good thumbnail is correlated with objective quality metrics, such as frame texture and sharpness [1]. YouTube also recommends selecting a high resolution of a thumbnail as possible [2]. Because the effect of visual aesthetics has already been researched, it is interesting to look at the content of the thumbnail rather than only the quality metrics.

A good classification scheme consists of the following pillars: inter-subjectivity (so it can be learned by a computer), avoiding specific labels and a concise number of labels. Following these pillars, the classification scheme for this research purpose distinguishes three categories for video thumbnails: 1) People, 2) Graphics, 3) Text. Definitions and examples of each label can be found in Table 1.

Label	Definition	Examples
People	The thumbnail has a focus on (real) people	● https://i.ytimg.com/vi/TdIsTLg7RQM/hqdefault.jpg
Graphics	The thumbnail consists of graphics. Comes in different forms: ● Animated people ● Graphics	● https://i.ytimg.com/vi/XCrOde-JYso/hqdefault.jpg ● https://i.ytimg.com/vi/ng22Ucr33aw/hqdefault.jpg

	● Background	● https://i.ytimg.com/vi/r4C5p8m-f14/hqdefault.jpg
Text	The thumbnail consists of mostly text OR the text in the thumbnail is the dominant feature, drawing the attention.	● https://i.ytimg.com/vi/0ztF-419fac/hqdefault.jpg ● https://i.ytimg.com/vi/RNRZchHaKgw/hqdefault.jpg

Table 1.

What type of questions can be answered?

There are multiple types of questions that can be answered after classifying and labeling thumbnails using this classification scheme. For instance, you can ask a correlational question such as: *"How does the content of a thumbnail drive user engagement in the YouTube social network?"* (provided that you have information on the user engagement). Furthermore, predictive analytics questions can be answered. Questions such as: *"Can we predict the popularity of a video based on its thumbnail?"* can be answered (provided that you have information on popularity of the videos).

Can it computationally be reproduced?

Because the classification scheme is based on the pillar of inter-subjectivity, it is made sure that the labels can be trained by a computer. A cascading classifier can be used to detect faces (People label) and Tesseract can recognize text (Text label). However, in a lot of the thumbnails there is also text on a People-labeled thumbnail (example). Therefore, it would be necessary to figure out where certain objects are located and how large they are, using a model trained on the MS Coco for example.

That being said, it would be more convenient to build a classification model after labeling a set of images manually (creating a 'ground truth'). It is important to calculate the Fleiss Kappa when labeling with multiple people: it is a way to measure the inter-annotator agreement. After this, you can use a ResNet50 classification model to label the rest of the thumbnails.

References

1. Song, Y., Redi, M., Vallmitjana, J., & Jaimes, A. (2016, October). To click or not to click: Automatic selection of beautiful thumbnails from videos. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (pp. 659-668).

2. https://creatoracademy.youtube.com/page/lesson/thumbnails#strategies-zippy-link-2