

Literature Review

Detection of spam comments in youtube videos

Primitives:

1. **Text classification** is the process of assigning tags or categories to **text** according to its content. It's one of the fundamental tasks in Natural Language Processing (NLP) with broad applications such as sentiment analysis, topic labeling, spam detection, and intent detection.

2. **Spam Comments :**

Spam based on the text of a **comment**, or by the behavior from a particular commenter. For instance, repeatedly posted **comments** are detected as **spam**.

Introduction:

Media resources in social Web spaces trigger social interactions, as they consist of motivating means to create and exchange user-generated content. The massive social content could provide rich resources towards deriving social profiles to augment user models and improve adaptation in simulated learning environments. However, potentially valuable social contributions can be buried within highly noisy content that is irrelevant or spam.

Methodology for the detection of intrusive comments or spam on Youtube -

We describe spam comments as those which have a promotional intent or those who deem to be contextually irrelevant for a given video. The prospects of monetisation through advertising on popular social media channels over the years has attracted an increasingly larger number of users. This has in turn led to the growth of malicious users who have begun to develop automated bots, capable of large-scale orchestrated deployment of spam messages across multiple channels simultaneously. The presence of these comments significantly hurts the reputation of a channel and also the experience of normal users.

Youtube themselves have tackled this issue with very limited methods which revolve around blocking comments that contain links. Such methods have proven to be extremely ineffective as Spammers have found ways to bypass such heuristics.

Standard machine learning classification algorithms have proven to be somewhat effective but there is still room for better accuracy with new approaches. In this work, we attempt to detect such comments by applying conventional machine learning algorithms.

Types Of Spam On Youtube:-

We can classify the majority of comment spam on Youtube as one of the following types.

Link Based Spam

This is a very common form of spam often seen on Youtube. Comments contain Hypertext (HTTP) links to other websites, usually other videos on Youtube itself. Many links redirect the the user to potentially malicious webpages often without the knowledge of the user.

Channel Promotional Spam

This is the most prevalent form of spam on Youtube. These types of comments usually consist of users who attempt to promote their own channel by requesting for subscribers, posting links to their videos etc.

Main Body:

1. Youtube spam detector program to be achieved upon analyzing the results of SVM with the cosine similarity kernel .It is important to note that an SVM is ideal in this problem since there are a disproportionate amount of spam instances when compared to the ham instances. In an SVM only the "edge-cases are regarded" and the model will be built according to these instances, so it is not a problem when one class has more instances than the other.

2. When there is a large dataset and it is important to avoid expensive algorithms, a Naive Bayes model is ideal since it is not computationally expensive. It just uses statistical models to analyze and then predict.

Conclusion

- A method for automated machine assisted detection of spam comments on the Youtube platform
- Combat link-based spam by blocking all comments containing Hypertext links

References

- Alberto, Túlio C, Johannes V Lochter, and Tiago A Almeida. "Tubespam: comment spam filtering on YouTube." In Machine Learning and Applications (Icmla), IEEE 14th International Conference on, 138–43. IEEE. (2015).
- N-Gram Assisted Youtube Spam Comment Detection
Shreyas Aiyara, Nisha P Shettyb , El Sevier (2018) International Conference on Computational Intelligence and Data Science (ICCIDS 2018)
- Ammari A., Dimitrova V., Despotakis D. (2012) Identifying Relevant YouTube Comments to Derive Socially Augmented User Models Ardissono L., Kuflik T. UMAP 2011. Lecture Notes in Computer Science, vol 7138. Springer, Berlin, Heidelberg