



Finding User Sentiments Using YouTube Comments

Shannon Baldwin, Dr. Aravind Mohan

Department of Computer Science, McMurry University, Abilene TX

Abstract

In the big data era, social media platforms like YouTube have emerged as massive repositories of information and opinions, reflecting the diverse voices and perspectives of users worldwide. Amidst this wealth of data, understanding user sentiment becomes paramount for organizations and individuals alike, serving as a compass for navigating the digital landscape and informing strategic decision-making processes. By discerning the prevailing sentiments expressed within the vast array of social media content, stakeholders can refine content strategies, tailor marketing approaches, and anticipate trends with greater accuracy. Addressing the challenge of extracting valuable insights from public opinions shared through social media platforms presents a multifaceted research endeavor. At the forefront of this inquiry lies the critical question of how to effectively harness the power of social media data to gain actionable insights. One promising avenue for exploration involves connecting topics to YouTube videos and conducting sentiment analysis on the accompanying comments. By categorizing sentiments as positive and negative, researchers can capture the prevailing public opinion on distinct topics and themes discussed in YouTube videos, providing a better understanding of audience perceptions and preferences. This entails extracting specific comments on topics of interest, ensuring a targeted and focused analysis. Using advanced natural language processing (NLP) techniques implemented in the Java programming language. By advancing our understanding of user sentiment on social media platforms like YouTube, this research aims to empower individuals and organizations with actionable insights for informed decision-making and strategic planning. We can unlock the full potential of social media data as a valuable resource for understanding human behavior, societal trends, and consumer preferences in the digital age.

Introduction

In the age of big data, social media platforms like YouTube serve as rich sources of information and opinions. Understanding the sentiments expressed by users within this vast sea of data is crucial for making informed decisions, refining content strategies, and adapting marketing approaches. However, deciphering user sentiment from YouTube comments poses a significant challenge due to the sheer volume of data and the diverse range of expressions. This research addresses the pressing need to extract valuable insights from public opinions shared on YouTube. The central question revolves around how to effectively analyze sentiments expressed in YouTube comments, categorizing them as positive, and negative and thereby capturing prevailing public opinion on various topics. To tackle this challenge, I proposed a systematic approach that begins with the collection of datasets from social media platforms, particularly YouTube. Our methodology involves employing natural language processing (NLP) techniques implemented in the Java programming language, with a focus on an object-oriented approach for enhanced efficiency and accuracy. While others may have approached sentiment analysis using similar NLP techniques, our research distinguishes itself by integrating a publicly available dataset containing sentiment-labeled words to improve the accuracy of sentiment classification. This correlates with approaches solely reliant on engagement metrics like likes, dislikes, and views. In this research, I outline the importance of understanding user sentiment on YouTube and I highlight the challenges associated with sentiment analysis. We also compared existing solutions and their limitations, paving the way for the presentation of my innovative methodology and its verification through rigorous experimentation. Through this research, I aim to provide a robust framework for effectively analyzing user sentiments expressed through YouTube comments, facilitating informed decision-making and content strategy development in the digital era.

Methodology

The methodology outlined in the provided section involves the analysis of comments on YouTube videos related to a specific topic. Our YouTube Data Analyzer (Y-DAZ), method begins by collecting videos related to a particular topic of interest. For each video, comments related to that video are gathered. The Xscore, representing the comparison of positive and negative words in each comment, is calculated. This score is derived by evaluating the ratio of positive words to total words minus the ratio of negative words to total words. Subsequently, an average X*score is computed across all comments for each video. Additionally, metrics such as the number of likes, dislikes, and view count for each video are recorded. Using these metrics, the Yscore, which reflects the sentiment of the audience towards the video, is calculated. The Yscore is determined by the ratio of likes to views minus the ratio of dislikes to views. Finally, a similarity score (S) is calculated for each video, representing the degree of similarity between the Yscore and the average Xscore. This score is obtained by adjusting the Yscore based on the difference between the Yscore and the average Xscore, normalized to a percentage scale. The resulting similarity scores provide insights into the alignment between audience sentiment and the characteristics of comments for each video.

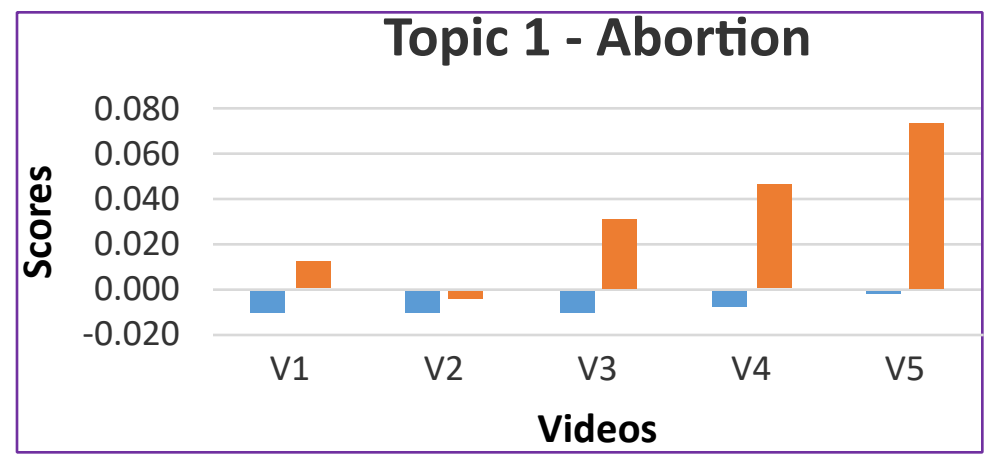
Y-DAZ ALGORITHM

- 1) $U \leftarrow$ Collect videos related to topic T
- 2) for each video u in U
- 3) $v \leftarrow$ Collect comments related to video U
- 4) $X*score = 0$
- 5) for each comment in v (individual) in V
- 6) $C1 \leftarrow$ find # of words in v
- 7) $C2 \leftarrow$ find # of positive in v
- 8) $C3 \leftarrow$ find # of negative in v
- 9) $Xscore = C2/C1 - C3/C1$ for each comment
- 10) $Xscore += Xscore$
- 11) end for
- 12) $X*score = Xscore / |Comments|$
- 13) $M1 \leftarrow$ # of Likes for video u
- 14) $M2 \leftarrow$ # of Dislikes for video u
- 15) $M3 \leftarrow$ View Count for video u
- 16) $Yscore = M1/M3 - M2/M3$
- 17) $S = X*score/Yscore$ for video u, if $X*score < Yscore$
- 18) $S = Yscore/X*score$ for video u, otherwise
- 19) end for
- 20) return S

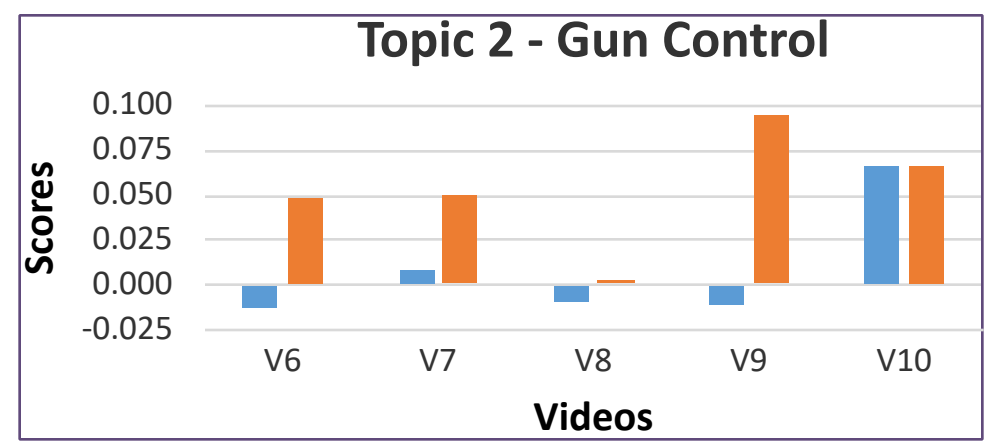
Sample Data of the Similarity Score of V1

List of Topics & Videos					
Video Topics	Video 1	Video 2	Video 3	Video 4	Video 5
Abortion	V1	V2	V3	V4	V5
Gun Control	V6	V7	V8	V9	V10
Artificial Intelligence	V11	V12	V13	V14	V15
Border Control	V16	V17	V18	V19	V20
Climate Change	V21	V22	V23	V24	V25
List of Comments for Topic 1					
Videos	Comments	Likes	Dislikes	Views	Yscore
V1	C1, C2, C3, C4, C5	726	18	56,667	0.012
V2	C6, C7, C8, C9	14,467	18,077	1,025,740	-0.004
V3	C10, C11, C12	14,176	935	416,140	0.032
V4	C13, C14	7,051	694	135,015	0.047
V5	C15	1,569	18	21,029	0.074
List of Comments for Topic 1					
Comments	Total Words	Pos. Words	Neg. Words	Xscore	
C1	13	0	0	0.000	
C2	41	2	0	0.049	
C3	51	1	1	0.000	
C4	23	1	0	0.043	
C5	45	0	1	-0.022	
			X* for V1 =	0.014	
			S for V1 =	89.20	
			Similarity Score (S) =	89%	

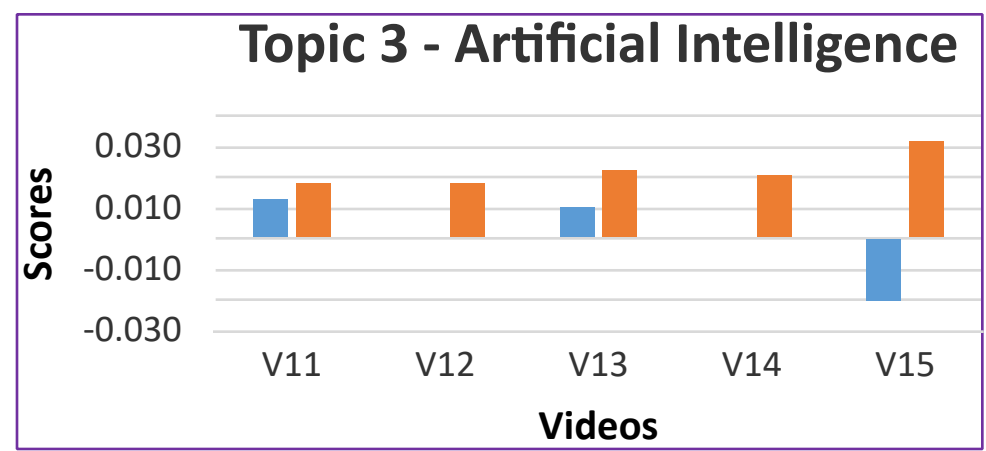
Experiments



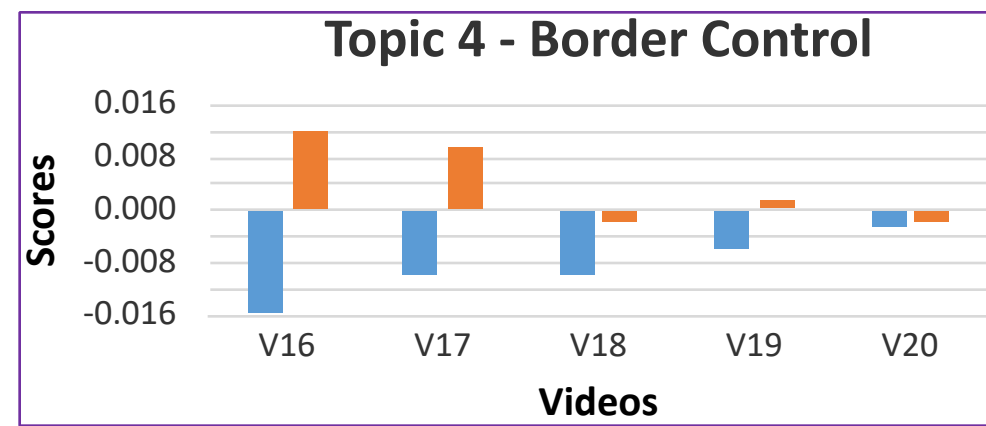
When analyzing the bar graph, we interpret values below zero as indicating a negative response to the YouTube video, while values above zero signify a positive response. For instance, let's consider the video, V3, which boasts 416,140 views along with 14,176 likes and 935 dislikes. At first glance, the video appears to have garnered an overall positive reception, as indicated by the Yscore. However, upon closer examination, we observe a disparity between the engagement metrics and the sentiment expressed in the comments (Xscore). Despite the video's popularity, the prevalence of negative words suggests a discrepancy between viewer reactions and expressed opinions.



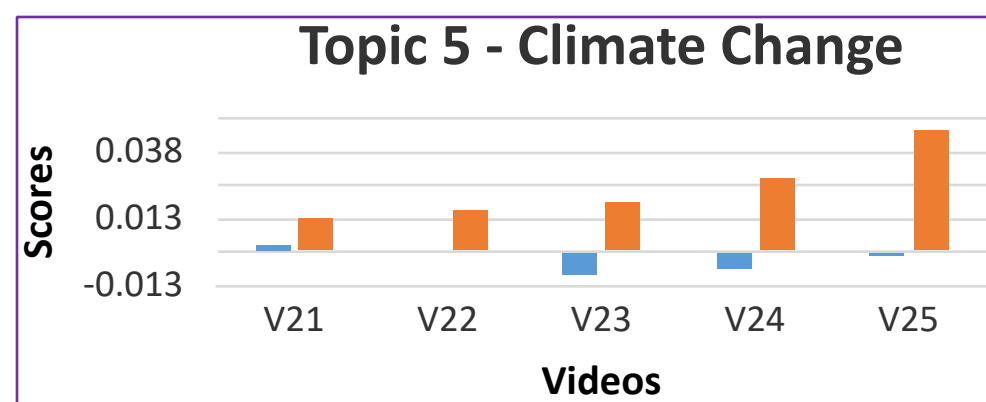
The video, V10, has metrics such as 726,727 views, 49,567 likes, and just 1,380 dislikes. Additionally, it's received a substantial amount of feedback, totaling 3,478 words across 100 comments. Remarkably, among these comments, 103 express positive sentiments while only 85 are negative. The synchronization between the engagement metrics and sentiment analysis implies a harmonious alignment between viewer reactions and the content of the comments. In essence, the positive response observed in both metrics indicates that the YouTube reactions effectively mirror the sentiments expressed in the comments.



The video, V12, has 69,688 views, accompanied by 1,293 likes and merely 64 dislikes. The video has garnered feedback through 2,273 words across 100 comments, with both positive and negative sentiments totaling 57 each. This balance of positive and negative sentiment words translates to an Xscore of 0. Notably, while the engagement metrics are substantial, the evenly distributed positive and negative sentiments imply a neutral sentiment expressed in the comments. Indicating a mixed reception or possibly a topic of debate among viewers.



The video, V16, which has 347,351 views, along with 4,357 likes and 187 dislikes. The video has generated 2,324 words across 100 comments, comprising 33 positive words and 69 negative words. Upon analyzing the bar graphs, it becomes apparent that the Xscore is negative, indicating a predominance of negative sentiment in the comments. Furthermore, the negative score surpasses the Yscore, signifying that the volume of negative words in the comments outweighs the positive reactions to the video. This discrepancy highlights a disparity between viewer sentiment and engagement metrics, suggesting a divergence in viewer perception and the actual content of the video.



The video, V25, has 458,369 views, accompanied by 23,319 likes and only 2,348 dislikes. Additionally, the video has a total of 5,547 words across 100 comments, with 143 expressing positive sentiments and 151 conveying negative sentiments. Notably, upon analyzing the bar graphs, we observe that the Yscore is overwhelmingly positive, reflecting the substantial viewer appreciation for the video. Conversely, the Xscore appears minuscule in comparison, indicating a negligible presence of negative sentiment in the comments relative to the overwhelmingly positive engagement metrics.

Conclusion

In summary, this project delves into analyzing user sentiment on YouTube to understand audience reactions and preferences better. By collecting and processing YouTube comments related to specific topics, we've developed a systematic approach to categorize sentiments as positive, negative, or neutral. Our methodology involves calculating sentiment scores like the Xscore and Yscore, which provide quantitative measures of sentiment and audience engagement. By comparing these scores with engagement metrics such as likes, dislikes, and views, we aim to uncover insights into the alignment between viewer reactions and expressed opinions in comments. Through experimentation and validation, we've demonstrated the reliability of our sentiment analysis methodology. This project contributes to understanding audience perception and content reception on YouTube, empowering content creators, marketers, and decision-makers with actionable insights for strategic decision-making and content strategy development in the digital age.