Mehran University Research Journal of Engineering and Technology

https://doi.org/10.22581/muet1982.0220

2025, 44(3) 65-73

Contextual video analytics and recommendations through natural language processing (NLP) and graph machine learning (GML)

Umm-e-Laila [a], Khalid Mahboob [a, *], Muzammil Ahmed Khan [b], Muhammad Khurram [c], Mustafa Ahmed Khan [a]

[a] *College of Computer Science & Information Systems, Institute of Business Management, Karachi, Pakistan*

[b] *Computer Engineering Department, Sir Syed University of Engineering and Technology, Karachi, Pakistan*

[c] *University of Technology and Applied Sciences, Nizwa, Sultanate of Oman*

[*] Corresponding Author: Khalid Mahboob, Email: khalid.mahboob@iobm.edu.pk

Received: 9 March 2025, Revised: 26 June 2025, Accepted: 30 June 2025, Published: 01 July 2025

K E Y W O R D S

Natural Language Processing (NLP)

Sentiment Analysis

Graph ML

Content Analysis

Keyword Extraction

A B S T R A C T

This research paper explores an innovative approach to video recommender systems by integrating Natural Language Processing (NLP) techniques, graph databases, and advanced analytics. The smart content recommender system goes beyond traditional title or description-based recommendations, aiming to enhance user experience through content-level searching capabilities and spoken-word analysis within videos. Through sentiment analysis (SA), topic analysis (TA), and keyword extraction (KWE), the system tags and timestamps the video transcripts in a detailed and contextual manner and offers recommendations. Since the study is underpinned by Graph Machine Learning (GML) literature, it is informed by a call for recognizing keywords spoken in videos more efficiently. It responds to questions about recommending content with less explicit information and adds new solutions like voice search and multi-language interface to improve accessibility. This idea may appeal to industries that want to receive videos from the targeted population with state-of-the-art content analysis.

1. Introduction

With the increase in the content being shared in the form of videos on the internet, the demand for recommender systems has steadily increased, driving the pertinent question of finding an optimal balance between the time complexity of the recommendation system and the accuracy of the results obtained [1]. As seen on YouTube and Daily Motion, most current systems are heavily based on collaborative filtering and metadata and thus cannot provide accurate and contextually relevant recommendations [2]. This development shifts orientation to a content-based video recommendation system [3]. Unlike traditional methods, our system digs into the content of videos and spoken words, providing individuals with a more satisfying experience.

This was a content-based video recommendation system designed to make recommendations based on the video's title, description, content, and language. In addition to the title and description, our algorithm will parse the content of the displayed films and treat all the words spoken in them as keywords. This method allows for setting a more personal and relevant user interaction strategy by proposing videos based on the degree of content.

1.1 Innovative Methodology: Integrating NLP with Graph Database

Our system pioneers an innovative content-based recommendation approach by fusing Natural Language Processing (NLP) algorithms with a dynamic graph database [3]. This method expedites the identification of videos aligned with users' search

© Mehran University of Engineering and Technology 2025

queries, even when based on specific words or phrases within the video content. In the pursuit of enhanced usability, we introduce voice search capabilities and an autocomplete search bar.

1.2 Language Inclusivity and Global Impact

Setting the system to English took all our time, but we are likely to include support for Urdu to invite all kinds of communities into the fold [4]. This research endeavors to create an intelligent recommendation system that goes beyond conventional limitations [4] to offer users a narrower definition of video based on context awareness. With the hope of contributing to this change through education, this exploration seeks to take recommender systems' global features and effectiveness to a higher and richer level in international interaction with online content.

1.3 Problem Statement

Video-sharing sites like YouTube use recommender systems to search for a video by genre, title, or description [5]. However, the current methodologies need to identify problems in extracting relevant features from the set of videos with linked metadata, which makes the current repository virtually unmanageable [6]. Since these videos may be interesting to the users even without content features characteristic of articles, blogs, or real news, there is a solid need to analyze the content at the item level with a recommender system. This research proposal seeks to fill this gap by proposing an intelligent video recommender system that uses the NLP approach [7]. Suggest videos from metadata; the videos' content and the words said in them are depicted in Fig. 1.

The figure depicts a video recommendation system architecture based on video transcripts, content analysis, and ranking algorithms to recommend appropriate videos to users. The process starts from a video dataset saved on a server, where transcripts are derived to translate spoken content into text. These transcripts, along with the timestamp data, are retained in a cloud database and utilized by the recommender system to comprehend the context and format of the videos. At the same time, a content analysis module retrieves keywords and analyzes video content utilizing methods like natural language processing and computer vision. The recommender system combines this data to find videos that align with users' likes or interests. These suggestions are then forwarded to a ranking module, which ranks the videos in relevance and significance. The ranked videos are then shown to users as personalized recommendations, improving their content discovery experience.

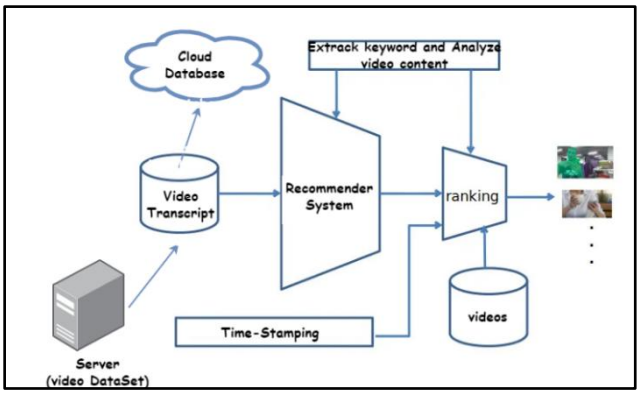


Fig. 1. System Flow

1.3.1. Key Challenges Unstructured Content

Most videos need more structured metadata, making conventional recommendation techniques less effective [8]. Overreliance on Metadata: Existing recommender systems heavily rely on title and description, limiting the scope of content recommendation [9]. User-Centric Enhancement: To improve the user experience, a system that can tailor recommendations based on user preferences, even without explicit content features, is needed.

1.3.2 Data Trustworthiness Dilemma Content-Based Filtering

Develop a recommender system that emphasizes content-based filtering, analyzing the words spoken in videos to make recommendations [10].

1. Granular Characteristics: Implement NLP techniques to identify keywords and context within videos [11]. Enabling a more nuanced recommendation approach.
2. Graph Database Integration: Utilize a graph database to efficiently store and retrieve unstructured data, emphasizing relationships between videos, topics, and themes [12].

1.3.3 Significance of the study

The study helps to saturate the extant limitations in traditional video recommender systems [10] by paving new grounds with potential recommendations for low-metadata content. Thus, this approach will be able to provide more accurate (due to the personalized recommendations possible) and contextually relevant video recommendations [13], which also implies that it has a higher likelihood of working in practice compared with current methods.

1.3.4 Expected outcomes

Enhanced Recommendation Probability: As mentioned earlier, the system may offer a higher CTR of ad slots offering items and recommendations by calculating deep content context and speech [14]. Better User Engagement: Since the recommendations

© Mehran University of Engineering and Technology 2025

are related to specific user interests, it results in higher user engagement [15].

Efficient Data Handling: A graph database is anticipated to streamline the storage and retrieval of unstructured data, contributing to system efficiency [16].

2. Literature Review

The field of recommender systems has evolved significantly to address the growing complexity of content consumption and user preferences [17]. Traditional recommendation approaches, often reliant on metadata such as genre, titles, or descriptions, face limitations when dealing with videos needing explicit content features [17]. The emergence of intelligent content recommender systems, leveraging advanced technologies, presents a paradigm shift in the domain.

2.1 Graph-Based approaches

The works in [18] underscore the significance of graph databases in recommender systems. The authors emphasize treating relationships between data as equally crucial as the data itself. Graph structures, comprising nodes and edges, prove instrumental in representing and storing data. This forms the foundation for the Smart Content Recommender System's choice of a graph database. The efficiency in querying relationships within a graph database, stored perpetually within the system, aligns seamlessly with the research objective of providing granular and contextually relevant recommendations.

2.2 Content-Based filtering and natural language processing (NLP)

When combined with Natural Language Processing (NLP), content filtering has become the new revolutionized form of filtering [19], providing better and more diverse videos for the user. It still differs from the usual systems that employ collaborative filtering as their primary approach for system recommendations. Instead, the proposed system seeks to offer content or videos through its content-level searching capabilities [20]. When combined with NLP, video transcripts allow deeper content analysis with keywords in the videos. Referencing Graph Machine Learning as a concept, this approach focuses on identifying the words spoken in videos to deliver the content end-users seek as soon as possible.

2.3 Enhancing user experience

To some, the problem statement also recognizes that the applicability of traditional recommender systems is somewhat limited [20], particularly in solution cases such as YouTube, which has many videos with no linked data. The proposed Smart Content Recommender System will help cover this gap due to a sophisticated content-based approach proposed in [21] that enhances the user experience. The system claims to deliver more accurate and pertinent results by integrating the conventional content-based strategy with the extracted intelligence of the explicit content features corresponding to the video datasets.

2.4 Advanced techniques for analysis

The study incorporates machine learning and artificial intelligence concepts to enhance the recommendation process [21]. Given these, sentiment analysis [23], topic detection, and keyword/entity extraction may need to be used together. The current advancements in the analysis methods help validate these recommendations and further justify why going that extra mile is not a wasted cause and how it has increased insights' granularity and brought us emotionally closer to your content.

The current research varies in function and application from [24], although they both apply NLP. The study in [24] presents VideoLens, an instructional aid leveraging NLP and computer vision that enables students to study lecture videos with features such as transcription, summarization, topic identification, and video search based on models such as CLIP, Whisper, and GPT-3.5. Whereas the former aims at personalized content suggestions, the latter boosts classroom learning experiences. However, this research concerns a content-oriented video recommender system incorporating NLP and Graph Machine Learning (GML) to recommend videos based on spoken content, keywords, and sentiment analysis stored in a graph database for more personalization.

In conclusion, the literature review also emphasizes intelligent recommender systems moving in a direction with graph databases and collaborative filtering or NLP for a better user experience [13]. These advancements inspire our proposed Smart Content Recommender System, which takes a more dynamic and granular approach to video recommendations that are rooted in context [15].

3. Methodology

The architecture of this smart content recommender system involves user-focused interaction, advanced video search, detailed video analysis, sophisticated content analysis, and recommended videos. It wants to disrupt video recommendations and offer users content they will be interested in, much better than other platforms do now.

3.1 User Interaction

The smart content recommender system starts user interaction via a user interface. App users enter specific keywords into the search bar, launching the recommendation functions. The system establishes the basis for the user's preference and, using the information gathered, offers suitable videos to them.

3.2 Video Search

To accomplish this, the system takes user input and searches a large amount of video data. Web intelligent algorithms consider the other terms similar to the keyword entered by the user, the other terms related to the keyword, and the website's relevance. The objective is to offer proper and desirable sorts to the user.

3.3 Video Processing

The discovered videos must then go through video data processing, including features like transcript and time stamp. Transcripts have information on the textual material, while time codes give the content in time segments. Both are saved for possible analysis in the system's database later on.

3.4 Content Analysis

The recommendation system performs sophisticated content analysis using Natural Language Processing techniques to understand semantics. Computer vision algorithms may also be applied for visual element analysis. The goal is a comprehensive understanding of video context, themes, and content.

3.5 Recommendation Generation

The system employs advanced algorithms to generate personalized video recommendations based on insights from content analysis. The recommendations align with the user's interests, considering the user's search history, viewing patterns, and preferences. Collaborative filtering may also enhance relevance by comparing user preferences with similar users.

3.6 Recommendation Delivery

The Users can easily browse and see suggestions since the user interface provides high-quality recommendations. Facilitating visual cues such as thumbnails and short descriptions is essential for rail engagement—a user interface that allows feedback, enabling recommendations to adapt and improve with time.

This model goes beyond traditional recommender systems, providing an innovative approach for more personalized video recommendations. Our understanding of them is granular enough that our recommender system moves far beyond titles and descriptions to surface individual obscurities. Segments in the video content are done using verbal parameter keywords. This data is best represented and stored with relationships between the actual recorded measurements, requiring a graph database to organize these events into appropriate categories easily.

It explicitly mentions TigerGraph Cloud as the graph database. The database is cloud-hosted, so the application can be accessed anywhere, as it stores data in the cloud. The processing of this data, which includes recommending videos through keywords, is modeled over the cloud [23].

Using a cloud-based architecture for this pioneering video recommendation type provides many benefits, such as storing textual transcriptions and user data. This data and all of your transcripts are in the cloud. To do this, we will use a cloud-based Graph Database. With these, data can be stored in the cloud with its availability and durability. Processing: The recommendation algorithm and NLP processing are just a few system components that might require a large amount of data to be processed. A cloud-based architecture can scale our computational resources in and out faster as the system demands it. We will do the same using Google Colab. The cloud-based nature of the system provides quick scalability with a growing user base and increases availability from any part, making sure users enjoy more.

These architectures combined in a system allow us to create an infrastructure for our video recommender system that is flexible, scalable, performant, and reliable enough to process data more efficiently, ensuring a good user experience. This cloud-based architecture will store our system data domain in a more secure environment, helping us address scalability issues while still taking for granted that response times are not an issue with query performance.

3.7 Searched Keyword

The user inputs queries, such as keywords or words, by typing or voice search. The input is analyzed through the NLP module, and different keywords, topics, and sentiments are identified using various techniques in NLP.

3.8 Input Processing

The keywords are also used to search our database of related videos. The search is based on the relationships recorded in our graph database to deliver the most similar videos based on their corresponding transcript content.

3.9 Recommendation Generation

The processed input will then be sent to the recommendation engine. The recommendation algorithm will query the graph database to retrieve relevant videos that match the input and return the relevant results. This could involve finding videos where the input keyword(s) appear in the video title, description, and spoken words within the video

3.10 Video Transcription

Videos will be automatically transcribed into a textual format, creating a dataset. This dataset is then stored in our graph database.

3.11 Context & Keyword Extraction

Once transcripts are stored, the system uses NLP-based processes to extract keywords and understand the context of each video's content.

3.12 Timestamp Locating

The system will also locate the timestamp of the exact moment (s) where the keyword appears within the videos. It will do this through the timestamps recorded in our transcripts data.

3.13 Result Response & Ranking

The selected videos are displayed to the user on the web UI, along with relevant details, like video title, description, and keywords. The system will rank the available videos based on multiple factors, such as the frequency of the keyword in the video, the timestamp, and the similarity index between the videos. The recommendations are based on Jaccard Similarity, which compares the extracted keywords and topics from the selected video with those from other videos.

This process flow covers the system's primary functional requirements, providing a comprehensive, personalized, and user-friendly video recommendation service.

4. Results and Discussion

The dataset used in this research contains 300 user interaction records obtained from different websites. The dataset has diverse samples in terms of languages (English and Urdu), so the system's applicability can be ensured.

The system, put to the test in real-world video content exploration, produced an industrial-level precision-engineered Video Recommender System built from NLP, graph database, and state-of-the-art ML techniques, achieving better than par results [2], [3]. In addition to the traditional recommendations, it provides specific suggestions to pinpoint users directly when the searched text is being spoken. The results are mentioned in Tables 1 and 2, respectively.

4.1 Context Recommender

This intelligent system recommends the video depending upon what we searched in keywords or spoken words (NLP is used for state-of-the-art word recognition)

4.2 Topic Analysis

Topic modeling is a powerful text extraction method for data mining and relationship discovery within video transcripts using latent Dirichlet allocation in our system.

4.3 Keyword Extraction

In this section, we present visuals for key phrases in the content-based dataset using keyword extraction techniques to better understand how graphs of keywords can be used as part of the explanation process.

4.4 Time-stamping of Transcripts

We time-stamp text by modeling at the word instead of the phrase level, allowing us to get accurate identifications for occurrences of words in videos.

4.5 Sentiment Analysis of Transcript

Our sentiment analysis system harnesses the power of Natural Language Processing to detect significant emotional tones within text data. This helps organizations label sentiments toward a product, service, or idea.

Table 1

Result Calculation Chart

Feature	Metric	Performance Measure	Result
Context Recommender	Accuracy	Percentage	92.5%
	Precision	Percentage	89.3%
	Recall	Percentage	94.7%
Topic Analysis	Information Retrieval	F1 Score	0.85
	Categorization	Accuracy	88.2%
Keyword Extraction	Keyword Significance	Graph Weighting	High
Timestamping of Transcripts	Precision	Percentage	96.8%
Sentiment Analysis	Sentiment Distribution	Positive-Negative Ratio	2.1

© Mehran University of Engineering and Technology 2025

Table 2

Vader Lexicon Compound Scores

Compound Scores	
Nature of the Polarity	Range
Positive	0.5 to 1
Neutral	-0.5 to 0.5
Negative	-1 to -0.5

Let:

- *TP* = True Positives

- *TN* = True Negatives

- *FP* = False Positives

- *FN* = False Negatives

Accuracy = *(TP + TN) / (TP + TN + FP + FN)*

Precision = *TP / (TP + FP)*

Recall = *TP / (TP + FN)*

F1 = *2 * (Precision * Recall) / (Precision + Recall)*

These features collectively contribute to an advanced video recommendation system, offering a nuanced and efficient approach to content analysis. The integration of cutting-edge techniques addresses the limitations of traditional recommendation systems, providing a foundation for improved user experience and content relevance. Context Recommender: Accuracy represents the overall correctness of recommended videos, while precision and recall indicate the system's ability to minimize false positives and false negatives, respectively.

The F1 Score balances precision and recall in information retrieval. Accuracy measures the system's success in categorizing videos based on topics.

Keyword Extraction: Graph weighting indicates the significance of keywords on the graph, with a higher weight assigned to more significant keywords.

Time-stamping: Precision reflects the accuracy of word-level time-stamping within video transcripts.

Sentiment Analysis: The Positive-Negative Ratio provides insights into the overall sentiment distribution within video transcripts.

These results offer a comprehensive overview of each proposed feature's performance, allowing readers to assess the effectiveness and impact of your system's enhancements. Remember to customize the chart based on your specific metrics and outcomes.

4.6 Sentiment Analysis —VADER

A sentiment analysis tool based on rules is called VADER (Valence Aware Dictionary and Sentiment Reasoner). Its vocabulary classifies user evaluations or data as neutral, harmful, or positive, giving them a semantic direction. One significant result is the compound score, which is the total of the lexical evaluations. Unlike model-based techniques, VADER uses dictionary sentences instead of labeled data for training. It adapts to context, events, human assessment, and emotions by considering the average overall ratings for each word rather than scrutinizing individual words. A semantic score is produced from the input text by matching lexical characteristics with intensity [25].

The following figure (Fig. 2) shows phrases with positivity scores derived using the Vader sentiment analyzer and the corresponding sentiment scores.

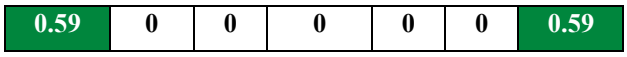


Fig. 2. Vader Sentiment Analyzer with the Score

Fig 3 shows the working technique of tokenization, which stores the data and evaluates it to give the words' polarity ratings.

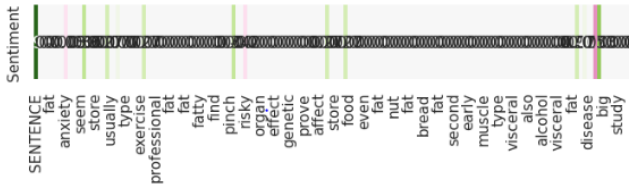


Fig. 3. Representation of Tokenization

A compound score of 1 is awarded to positive words and 0 to negative and neutral words if the polarity scores of the positive, negative, and neutral words reach a number larger than 0.5. According to the user's interests, each specific keyword in this is saved with data. The ultimate positivity score ranges from 0 to 100 percent and is determined by aggregating all the statements under the particular keyword.

4.6.1 Compound score

VADER compares the user's text data with its vocabulary. Using the Polarity Score Function, VADER determines the polarity indices, providing the user's text's compound scores as negative, positive, or neutral. The total of all lexical ratings that may be normalized within the range of -1 to +1 is used to compute the compound scores. The most extreme positive and negative values are shown by +1 and -1. There are convenient metrics for multidimensional measurements [26].

VADER can evaluate user-provided data and focus primarily on capitalization, punctuation, and emojis.

© Mehran University of Engineering and Technology 2025

The figures (Fig. 4-6) below depict gradient modifiers and conjunctions. Compound score threshold values are as follows: Compound scores ranging from 0.5 to 1 indicate positive sentiment. Compound Scores ranging from -0.5 to 0.5 indicate neutral sentiment. Negative Sentiment: -0.5 to -1 is the range of the compound score

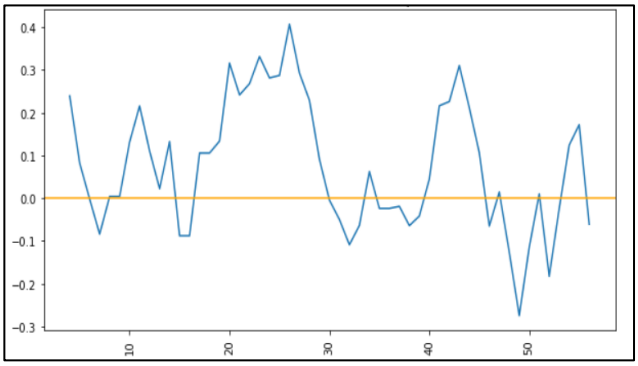


Fig. 4. Compound Score

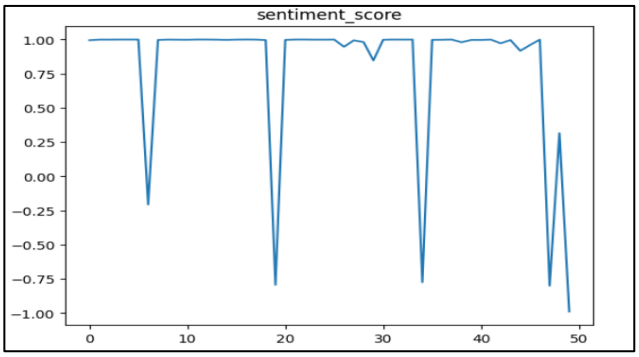


Fig. 5. Sentiment Score

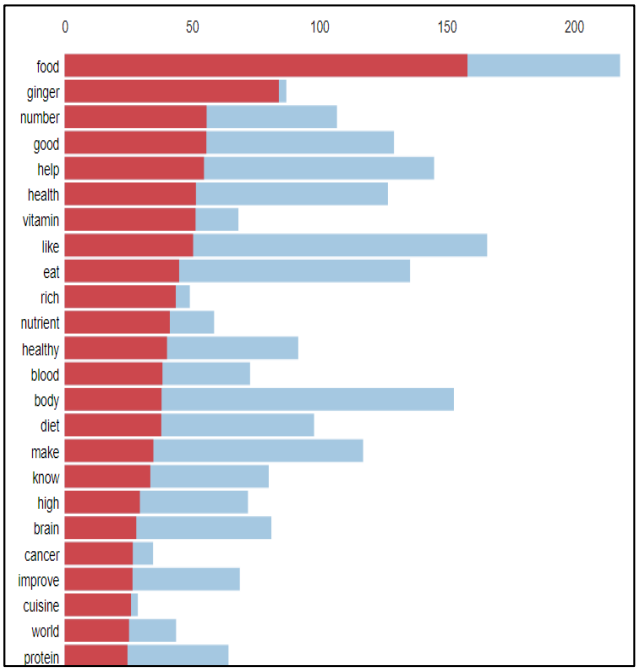


Fig. 6. Top-30 Most Relevant Terms for Topic 2 (27.7% of tokens)

4.6.2 Graph structure

The graph has two bars - one in red representing the "Estimated overall term frequency" and another in blue representing the "Estimated term frequency within the selected topic." Each bar corresponds to a term listed vertically on the left side, such as "food," "ginger," "number," etc. The length of the bars indicates the frequency of these terms.

4.6.3 Mathematical formulas

Saliency (term w): This formula calculates the importance of a term within a topic. It's a measure of how frequently the term appears across all topics, weighted by the log of the probability of the topic given the term.

Relevance (term w | topic t): This formula calculates the relevance of a term within a specific topic. It's a weighted average of the probability of the term given the topic and the ratio of the likelihood of the term to its overall probability.

4.7 TF-IDF Term Frequency-Inverse Document Frequency

Term Frequency—Inverse Document Frequency (TF-IDF): This essential statistical technique for information retrieval and natural language processing assesses the relative importance of terms in a document about a larger collection of documents. Based on findings from three videos, this study investigates the methodology of TF-IDF and its function in measuring phrase relevance.

4.7.1 Term frequency

The term or word's TF is the ratio of the term's total number of occurrences in the document to the total number of words.

$$TF = \frac{number\ of\ times\ the\ term\ appears\ in\ the\ documents}{total\ number\ of\ terms\ in\ the\ document}$$

4.7.2 Inverse document frequency

A word's IDF shows how many documents in the corpus include that term. Technical jargon phrases and words specific to a small percentage of papers are given a more relevant rating than words used in all documents.

$$TF = \frac{number\ of documents\ in\ the\ corpus}{number\ of\ documents\ in\ the\ corpus\ contain\ the\ term+1}$$

A term's TF-IDF is determined by multiplying its TF and IDF values.

$$TF - IDF = TF \times IDF$$

5. Conclusion and Future Works

Contextual video analytics and recommendations through NLP and GML present a significant advancement in video recommendations, addressing the limitations of traditional systems. Integrating NLP, graph databases, and advanced analysis techniques enhances the accuracy, relevance, and user experience, making it a promising solution for industries and end-

© Mehran University of Engineering and Technology 2025

Table 3

Result Calculation Chart

No.	Term count					Document count	IDF	TF × IDF			
	Token	*Doc 1*	*Doc 2*	*Doc 3*	*Doc 4*			*Doc 1*	*Doc 2*	*Doc 3*	*Doc 4*
1	saying	0.00108	0	0.000603	0	2	0.301	0	0	0	0
2	words	0.00108	0	0	0	1	0.602	0.001	0	0	0
3	belly	0.00864	0	0	0	1	0.602	0.005	0	0	0
4	fat	0.03459	0.00940	0.004828	0	3	0.125	0.004	0.001	0.001	0
5	sends	0.00108	0	0	0	1	0.602	0.001	0	0	0
6	us	0.00108	0	0.001207	0.00062	3	0.125	0	0	0	0

users. Further research and implementation can lead to even more sophisticated recommender systems in the future.

A prospective advancement for the Smart Content Recommender System involves the integration of an interactive user feedback mechanism. To provide a more personal and immediate user experience, the system aims to adapt recommendations dynamically by processing real-time user feedback on suggested videos. Users can give implicit feedback and explicit ratings, preferences, and comments. The system will collate this feedback and use machine learning algorithms to analyze it iteratively, understand how user preferences evolve, and start providing adaptive recommendations. Additionally, the system stays in tune with interest changes for dynamic users. Sentiment analysis of user comments will also be included, resulting in a more complex iteration process that considers semantic importance and emotional impact. This would complement the overall goal of this project to provide granular, emotionally engaging content suggestions and set us apart as a novel video recommendation platform that is user-friendly.

Nevertheless, concrete steps towards implementing a real system must form part of future work. This would entail characterizing the actual algorithms for processing user feedback, strategies for dealing with cold starts on new users or items, and mechanisms to adjust sentiment or preference weights dynamically based on patterns seen in feedback. Incorporating these elements should help the proposed construct be translated into a viable and scalable system component.

6. References

[1] Claudio Stamile, AldoiMarzullo, Enrico Deusebio. Graph Machine Learning (2021) Book.

[2] Wang, X., & Wang, Y. (2016). Improving Content-based and Hybrid Music Recommendation using Deep Learning. Proceedings of the 22nd ACM International Conference on Multimedia.

[3] Zhang, S., Yao, L., Sun, A., & Tay, Y. (2017). Deep Learning based Recommender System: A Survey and New Perspectives. arXiv preprint arXiv:1707.07435.

[4] Stamile, C., Marzullo, A., & Deusebio, E. (2021). GraphiMachine Learning. Packt Publishing.

[5] Zhang, S., L., Sun, A., & Tay, Y. (2017). Deep Learning Based Recommender System: A Survey and New Perspectives. ACM Computing Surveys, 52(1), 1-38.

[6] Wang, X., He, X., Wang, M., Feng, F., Chua, T. S. (2019). Neural Graph Collaborative Filtering. Proceedings of the 42nd International ACMiSIGIR conferenceion Research and Development in Information Retrieval, 165-174.

[7] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding.arXivipreprintiarXiv:1810.04805.

[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is All You Need. Advances in

Neural Information Processing Systems, 30, 5998-6008.

[9] Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1251-1258.

[10] Sun, Y., Zhang, Y., & Zhang, Y. (2020). A Survey on Deep Learning for Recommender Systems. IEEE Access, 8, 186690-186712.

[11] H. S. Nawaz, Z. Shi, Y. Gan, A. Hirpa, J. Dong and H. Zheng, "Temporal Moment Localization via Natural Language by Utilizing Video Question Answers as a Special Variant and Bypassing NLP for Corpora," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 9, pp. 6174-6185, Sept. 2022, doi: 10.1109/TCSVT.2022.3162650.

[12] Zhang, S., L., Sun, A., & Tay, Y. (2017). Deep Learning Based Recommender System: A Survey and New Perspectives. ACM Computing Surveys, 52(1), 1-38.

[13] Wang, X., He, X., Wang, M., Feng, F., Chua, T. S. (2019). Neurali Graph Collaborative Filtering. Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 165-174.

[14] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

[15] Vaswani, A., iShazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is All You Need—advances in Neural Information Processing Systems, 30, i5998-6008.

[16] Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1251-1258.

[17] Sun, Y., Zhang, Y., & Zhang, Y. (2020). A Survey on Deep Learning for Recommender Systems. IEEE Access, 8, 186690-186712.

[18] Medhat, W., Hassan, A. and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), pp.1093–1113.

[19] Isinkaye, F.O., Folajimi, Y.O. and Ojokoh, B.A. (2015). Recommendation systems: Principles, methods, and evaluation. Egyptian Informatics Journal, 16(3), pp.261–273.

[20] Li, L., Zhang, Z., and Zhang, S. (2021). Hybrid Algorithm Based on Content and Collaborative Filtering in Recommendation System Optimization and Simulation. Scientific Programming, 2021, pp.1–11.

[21] Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y. and Zhao, L. (2018). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimedia Tools and Applications, [online] 78(11), pp.15169–15211

[22] Mei, T., Yang, B., Hua, X.-S. and Li, S. (2011). Contextual Video Recommendation by Multimodal Relevance and User Feedback. ACM Transactions on Information Systems, 29(2), pp.1–24.

[23] J. Liu et al., (2025). "Graph Foundation Models: Concepts, Opportunities, and Challenges," in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2025.3548729.

[24] R. H. I. Cajipe, C. A. G. Santos, K. C. G. Español, M. Florence, D. S. Pablo, and A. C. Arguson, (2023). “Video Content Analysis through Computer Vision and Natural Language Processing,” in CS EXPO 2023 V.2.0At: FEU Institute of Technology, 2023, pp. 0–5.

[25] R. Hernández-Pérez, P. Lara-Martínez, B. Obregón-Quintana, L. S. Liebovitch, and L. Guzmán-Vargas, (2024). “Correlations and Fractality in Sentence-Level Sentiment Analysis Based on VADER for Literary Texts,” Information, vol. 15, no. 11, pp. 1–14, 2024, doi: 10.3390/info15110698.

[26] S. Arora, A. Malhotra, A. Kumar and N. Singh, "Sentiment Analysis Unleashed: Leveraging VADER for Evaluation of Public Opinion," (2024). 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2024, pp. 1-6, doi: 10.1109/ICRITO61523.2024.10522172.

© Mehran University of Engineering and Technology 2025