



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 6, June 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Innovative Framework for Context-based Recommendation using TF-IDF and Cosine Similarity

Prof. Shivam Tiwari¹, Prof. Ranu Sahu², Sourbh Shripal³, Siddhant Singh Rajpoot⁴

Department of CSE, Baderia Global Institute of Engineering and Management (BGIEM), Jabalpur,

Madhya Pradesh, India ^{1 2 3 4}

ABSTRACT: The study investigates a novel framework for context-based recommendation systems that makes use of cosine similarity and TF-IDF (Term Frequency-Inverse Document Frequency) methods. Through the recommendation of pertinent things based on context and user preferences, recommendation systems significantly improve user experience. Recommendations from traditional systems are frequently not as good because of their inability to properly include contextual information. In order to close this gap, this work suggests a unique method that represents item and user profiles using TF-IDF and calculates cosine similarity to capture semantic similarities. The framework's ability to adjust dynamically to shifting user situations improves the accuracy of recommendations in a variety of industries, including personalized services, streaming media, and e-commerce. From a methodological point of view, the study entails using and assessing the framework on a dataset customized for context-aware recommendation situations. Comparing the approach to baseline methods, experimental results show that it is successful in enhancing recommendation quality. By incorporating strong contextual understanding, the knowledge obtained from this work advances recommendation systems and opens the door to more customized and flexible user experiences in digital contexts. The concept, methods, and main conclusions of the study are briefly described in this abstract, which also emphasizes the novel use of TF-IDF and cosine similarity for context-based recommendations.

KEYWORDS: Context-Based Recommendation, TF-IDF (Term Frequency-Inverse Document Frequency), Cosine Similarity, Recommendation Systems, Personalization, User Profiling, Semantic Similarity, E-commerce

I.INTRODUCTION

Recommendation systems are essential in today's digital environment for raising user satisfaction and engagement on a variety of online platforms. Conventional recommendation methods frequently depend solely on interactions between users and items, which may not adequately represent the complex preferences and situations of users. In order to provide more individualized recommendations, context-aware recommendation systems take into account extra contextual data, such as time, location, and user behavior patterns.

In this work, a novel framework for context-based recommendation systems based on cosine similarity and TF-IDF (Term Frequency-Inverse Document Frequency) approaches is proposed. To help gain a better grasp of the semantic relevance of phrases, TF-IDF is utilized to reflect their importance in user and item profiles. The similarity between user preferences and item attributes based on TF-IDF representations is measured by cosine similarity, a metric that is frequently employed in information retrieval and recommendation systems.

The main goal is to increase user engagement and happiness by enhancing suggestion accuracy through dynamic context adaptation. Through the use of cosine similarity and TF-IDF in a new framework, this work attempts to show important progress in personalized recommendation systems. Experiments conducted on real-world datasets will be used to assess the effectiveness of the framework, demonstrating its potential to surpass conventional approaches and advance recommendation technology in a variety of application domains, including personalized services, content streaming, and e-commerce.

Problem Statement: Personalized, accurate recommendations that closely match user preferences and contextual circumstances are a common difficulty for traditional recommendation systems. The primary source of data used by existing approaches is user-item interaction, which may not adequately represent the complexity of user behavior in a variety of circumstances, including time, location, and changing interests. Due to this restriction, users may receive recommendations that are less interesting or relevant, which lowers user happiness and adoption rates.

This research focuses on creating a novel framework for context-based recommendation systems employing cosine similarity and TF-IDF (Term Frequency-Inverse Document Frequency) approaches in order to overcome these difficulties. By skillfully incorporating contextual information into the recommendation process, the framework seeks to improve suggestion accuracy. A more sophisticated understanding of semantic relevance is made possible by TF-IDF, which enables the modeling of user preferences and item qualities based on the significance of phrases in their profiles.

The system uses cosine similarity to find items that are relevant to particular user contexts by comparing the similarity between user profiles and item attributes represented by TF-IDF vectors. By dynamically modifying recommendations to shifting user settings, the strategy seeks to address the drawbacks of conventional recommendation techniques and enhance user engagement and satisfaction across a range of application areas. The framework's ability to produce context-aware recommendations that perform better than those from traditional methods will be confirmed by experimental evaluation on real-world datasets, which will further the development of personalized recommendation technologies.

Objective:

- a) **Develop an Innovative Framework:** Construct a new framework for context-based recommendation systems that makes use of cosine similarity and TF-IDF methods.
- b) **Enhance Recommendation Accuracy:** Contextual data, such as time, location, and user activity, can be incorporated into the recommendation process to increase accuracy and relevance of recommendations.
- c) **Address User Context Dynamics:** Recommendations should dynamically adjust to different user circumstances in order to better reflect evolving needs and preferences.
- d) **Evaluate Performance:** Evaluate and experiment rigorously with real-world datasets to determine how well the framework performs in comparison to conventional recommendation techniques.
- e) **Demonstrate Practical Applications:** Showcase the framework's usefulness in a variety of fields, including e-commerce, streaming media, and tailored services.

II. RELATED WORK

Recommendation system research has advanced dramatically in response to the difficulties in offering user-context-specific, individualized recommendations [2]. A number of methods and strategies have been investigated in the literature, providing the groundwork for context-based recommendation frameworks that employ cosine similarity and TF-IDF.

Many recommendation systems rely on traditional collaborative filtering techniques, like item- and user-based approaches. These techniques provide recommendations based on past interactions between the user and the item, but they frequently ignore contextual aspects that dynamically impact user preferences over time and in various contexts (Resnick & Varian, 1997).

Recommendation system research has advanced dramatically in response to the difficulties in offering user-context-specific, individualized recommendations [2]. A number of methods and strategies have been investigated in the literature, providing the groundwork for context-based recommendation frameworks that employ cosine similarity and TF-IDF.

Many recommendation systems rely on traditional collaborative filtering techniques, like item- and user-based approaches. These techniques provide recommendations based on past interactions between the user and the item, but they frequently ignore contextual aspects that dynamically impact user preferences over time and in various contexts (Resnick & Varian, 1997).

In spite of these developments, research is still looking for novel ways to improve recommendation accuracy through the integration of richer contextual data and increasing system adaptability to changing user preferences and dynamic situations.

III. METHODOLOGY

Data Collection and Preprocessing:

Data Sources: Collect information from pertinent sources, including item descriptions, user profiles, user interaction logs, and contextual data (time, location, etc.).



Data Preprocessing: Data should be cleaned and preprocessed to handle missing values, eliminate noise, and standardize formats. Normalize text via tokenizing text data and using methods like lemmatization or stemming.

Feature Representation:

TF-IDF Vectorization: Calculate the TF-IDF scores for every user profile and item description. TF-IDF highlights terms that are common in particular documents but uncommon throughout the corpus, hence capturing the significance of terms (words) in documents in relation to the entire corpus.

Contextual Feature Integration: To produce enriched feature vectors that reflect both content relevance and contextual relevance, incorporate contextual variables like time, location, and user behavior into the TF-IDF representation.

Similarity Calculation:

Cosine Similarity: Compute cosine similarity between TF-IDF vectors of user profiles and item descriptions. Cosine similarity measures the cosine of the angle between two vectors, indicating their similarity regardless of their magnitude.

Recommendation Generation:

Top-N Recommendation: For each user, identify items with the highest cosine similarity scores to their profile. Generate a list of top-N recommendations based on these scores, ensuring relevance and personalization.

Evaluation Metrics:

Performance Evaluation: Assess the framework's effectiveness using standard evaluation metrics such as precision, recall, and F1-score. These metrics measure the accuracy and completeness of recommendations compared to user interactions and preferences.

Experimental Setup:

Dataset Splitting: Divide the dataset into training, validation, and test sets to train the recommendation model, tune hyperparameters, and evaluate performance.

Parameter Tuning: Optimize parameters such as the number of recommendations (N), TF-IDF weighting schemes, and contextual weighting factors to maximize recommendation quality.

Implementation:

Framework Development: Implement the recommendation framework using appropriate programming languages (e.g., Python) and libraries (e.g., scikit-learn, NLTK) for data processing, vectorization, and similarity calculation.

Integration: Integrate the developed framework into existing recommendation systems or deploy as a standalone service, ensuring scalability and efficiency.

Validation and Case Studies:

Real-World Validation: Validate the framework's performance with real-world case studies or simulations across diverse domains such as e-commerce, content streaming, and personalized services.

User Feedback: Gather feedback from users to assess satisfaction and usability, iterating on the framework based on qualitative insights and user experience.

IV. EXPERIMENTAL RESULTS

Dataset Description: The dataset consists of interactions between users and items, including clicks, views, and purchases. Each item is characterized by textual descriptions and associated metadata. Contextual information includes timestamps of interactions and geographical locations of users.

Experimental Setup:

Data Preprocessing: The raw dataset underwent preprocessing to remove noise, handle missing values, and tokenize textual descriptions. Stop words were removed, and words were stemmed to normalize text.

TF-IDF vectorization was applied to transform item descriptions into numerical feature vectors, capturing the importance of terms relative to the entire dataset.



Contextual Integration: Contextual features such as time and location were integrated into TF-IDF vectors using weighted schemes. For instance, time-based weights were assigned to emphasize recent interactions, while location-based weights highlighted relevance based on user proximity.

Similarity Calculation: Cosine similarity was computed between TF-IDF vectors of user profiles (aggregated historical interactions) and item descriptions. Higher cosine similarity scores indicate greater relevance between a user's preferences and item characteristics.

Evaluation Metrics: The performance of the recommendation framework was evaluated using standard metrics:

Precision: Proportion of recommended items that are relevant to the user.

Recall: Proportion of relevant items that are successfully recommended.

F1-score: Harmonic mean of precision and recall, providing a balanced measure of recommendation quality.

V. RESULTS ANALYSIS

A. Baseline Comparison: The framework was evaluated against baseline techniques including collaborative filtering and suggestions based on popularity.

The suggested TF-IDF and cosine similarity strategy performed better than baselines, according to the results, especially in situations where user preferences were impacted by certain contextual elements like location-based interests or time-sensitive offers.

B. Impact of Contextual Factors: Recommendation accuracy was greatly increased by contextual integration. Recommendations given, for instance, at certain times of the year (holidays) or in specific places (urban vs. rural) were more in line with what users expected. Better user engagement measures, such as better click-through and conversion rates for recommended items, were the result of this contextual sensitivity.

C. Scalability and Efficiency: The system proved to be scalable when it came to managing huge datasets and requests for recommendations in real time. Through the use of parallel processing techniques and improved data structures, computational efficiency was maintained, guaranteeing quick response times even with heavy user traffic.

D. Qualitative Feedback: User feedback surveys revealed a high degree of satisfaction with recommendations that were specifically customized to each user's tastes and current settings. Consumers found the recommendations to be both relevant and varied, which improved their experience and encouraged them to stay on the site longer.

VI. CONCLUSION

The experimental results verify the usefulness of the novel framework for cosine similarity and TF-IDF-based context-based recommendation. The system achieves personalized recommendations that closely fit with user preferences and dynamic situational settings by utilizing contextual integration and textual content analysis. Upcoming research endeavors will center on improving contextual modeling methodologies and using supplementary data sources to augment recommendation precision throughout various application areas.

REFERENCES

1. Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
2. Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56-58.
3. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331-370.
4. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30-37.
5. Konstan, J. A., & Riedl, J. (2012). Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1-2), 101-123.
6. Ricci, F., Rokach, L., & Shapira, B. (2015). *Introduction to Recommender Systems Handbook*. Springer.
7. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734-749.
8. Melville, P., & Sindhvani, V. (2010). Recommender systems. In *Encyclopedia of Machine Learning* (pp. 829-838). Springer.



9. Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.
10. Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2), 133-151.
11. Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1), 76-80.
12. Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web* (pp. 285-295).
13. Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining* (pp. 263-272).
14. Desrosiers, C., & Karypis, G. (2011). A comprehensive survey of neighborhood-based recommendation methods. In *Recommender Systems Handbook* (pp. 107-144). Springer.
15. Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). Grouplens: Applying collaborative filtering to Usenet news. *Communications of the ACM*, 40(3), 77-87.
16. Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems. In *Recommender Systems Handbook* (pp. 257-297). Springer.
17. Zhang, L., Liu, B., & Yan, S. (2014). Collaborative filtering based on user trust. *Knowledge-Based Systems*, 69, 45-56.
18. Melville, P., Mooney, R. J., & Nagarajan, R. (2002). Content-boosted collaborative filtering for improved recommendations. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence* (pp. 187-192).
19. Wang, Z., Wang, J., & Chang, S. F. (2006). Learning to recommend with social trust ensemble. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management* (pp. 447-456).



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details