CHALMERS
UNIVERSITY OF TECHNOLOGY

A Data-driven Approach for Mining Software Features based on Similar App Descriptions and User Reviews Analysis

RESEARCH-ARTICLE

A Data-driven Approach for Mining Software Features based on Similar App Descriptions and User Reviews Analysis

KHUBAIB AMJAD ALAM, Al Ain University, Al Ain, Abu Dhabi, United Arab Emirates

RAMSHA ALI, National University of Computer and Emerging Sciences Islamabad, Islamabad, Islamabad, Pakistan

ZYENA KAMRAN, National University of Computer and Emerging Sciences Islamabad, Islamabad, Islamabad, Pakistan

SABEEN FATIMA, National University of Computer and Emerging Sciences Islamabad, Islamabad, Islamabad, Pakistan

IRUM INAYAT, University of Gothenburg, Gothenburg, Vastra Gotaland, Sweden

Open Access Support provided by:

Al Ain University

National University of Computer and Emerging Sciences Islamabad

University of Gothenburg

Published: 27 October 2024

Citation in BibTeX format

ASE '24: 39th IEEE/ACM International Conference on Automated Software Engineering
October 27 - November 1, 2024
CA, Sacramento, USA

Conference Sponsors:
SIGAI
SIGSOFT

A Data-driven Approach for Mining Software Features based on Similar App Descriptions and User Reviews Analysis

Khubaib Amjad Alam
College of Engineering, Al Ain University
United Arab Emirates
khubaib.alam@aau.ac.ae

Ramsha Ali
National University of Computer emerging Sciences, Islamabad, Pakistan
i200839@nu.edu.pk

Zyena Kamran
National University of Computer emerging Sciences, Islamabad, Pakistan
i200802@nu.edu.pk

Sabeen Fatima
National University of Computer emerging Sciences, Islamabad, Pakistan
i200505@nu.edu.pk

Irum Inayat
University of Gothenburg, Chalmers University of Technology, Sweden
irum@chalmers.se

ABSTRACT

Mobile app development necessitates extracting domain-specific, essential, and innovative features that align with user needs and market trends. Determining which features provide a competitive advantage is a complex task, often managed manually by product managers. This study addresses the challenge of automating feature mining and recommendation by identifying similar apps based on user-provided descriptions. The proposed approach integrates Named Entity Recognition (NER) for feature extraction from mined Google Play app data with BERT (Bidirectional Encoder Representations from Transformers) and Topic Modeling to find comparable apps. Our top-performing model, which uses Non-negative Matrix Factorization (NMF) for Topic Modeling with Sentence-BERT (SBERT) embeddings, achieves an F1 score of 87.38%.

ACM Reference Format:
Khubaib Amjad Alam, Ramsha Ali, Zyena Kamran, Sabeen Fatima, and Irum Inayat. 2024. A Data-driven Approach for Mining Software Features based on Similar App Descriptions and User Reviews Analysis. In *39th IEEE/ACM International Conference on Automated Software Engineering (ASE '24), October 27-November 1, 2024, Sacramento, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3691620.3695342

1 RESEARCH PROBLEM AND MOTIVATION

Mobile app features typically stem from user demands, market dynamics, and business goals, necessitating considerable investment of time and resources. Although app stores have been identified as valuable sources for automating feature extraction, prior research has primarily focused on user reviews. However, other underutilized data sources, such as app descriptions also offer substantial untapped potential. While existing studies often target bug reports and problematic features from user reviews [1], there remains a gap in leveraging these data for both inspiring new app development and enriching requirements elicitation.

Previous research has explored feature identification and recommendation through data mining and natural language processing (NLP) techniques. Works such as by Alves et al. [2] extracted features from requirement specifications and forums, yet these methods were less effective for new apps. Jiang et al. [3] introduced SAFER, a model for recommending new features by identifying similar apps using Latent Dirichlet Allocation (LDA). Similarly, Scalabrino et al. [6] focused on extracting features from user reviews, while Liu et al. [4] analyzed app descriptions and UI texts to facilitate feature updates. In contrast to existing studies, we propose using NER and BERT models for automated feature extraction.

2 APPROACH AND NOVELTY

Our approach enables users to input App URLs or text descriptions of new app ideas to receive insights on competitor apps, recommended features, and user review analytics. The methodology, illustrated in Fig. 1, is structured around three core modules:

a) Feature Extraction from Google Play Data: We extract features from Google Play App data using a specialized NER model trained on datasets with annotated entities from app descriptions and user reviews. For app descriptions, we directly labeled and trained the NER model. For user reviews, we applied sentiment analysis to filter positive feedback and utilized manually curated custom linguistic patterns with RAKE (Rapid Automatic Keyword Extraction) to pinpoint key aspects. A semi-supervised learning approach was then used to refine feature classification. This process results in two app feature datasets: one derived from user reviews and another from app descriptions.

b) Identifying the Optimal Mobile App Recommendation Technique: This module explores BERT-based methods combined with Topic modeling to determine the optimal technique for recommending similar apps. We create contextual representations of app descriptions using lightweight BERT models such as *'all-MiniLM-L6-v2'*, *'stsb-roberta-base'*, and *'msmarco-distilbert-base-v2'*. We assess three approaches:

- Using BERT embeddings alone.
- Applying LDA on BERT embeddings.
- Applying NMF on BERT embeddings.

We evaluate these methods by processing sample input user descriptions (queries) with true labels to align with the format of app descriptions. The embeddings and topic probability vectors are then compared to identify the top five similar apps for each query. The results demonstrate the most effective approach (Table 1).

c) Top-Ranked Features and User Review Insights: We compute the similarity between user queries and app descriptions using

ASE '24, October 27-November 1, 2024, Sacramento, CA, USA

ACM ISBN 979-8-4007-1248-7/24/10
https://doi.org/10.1145/3691620.3695342

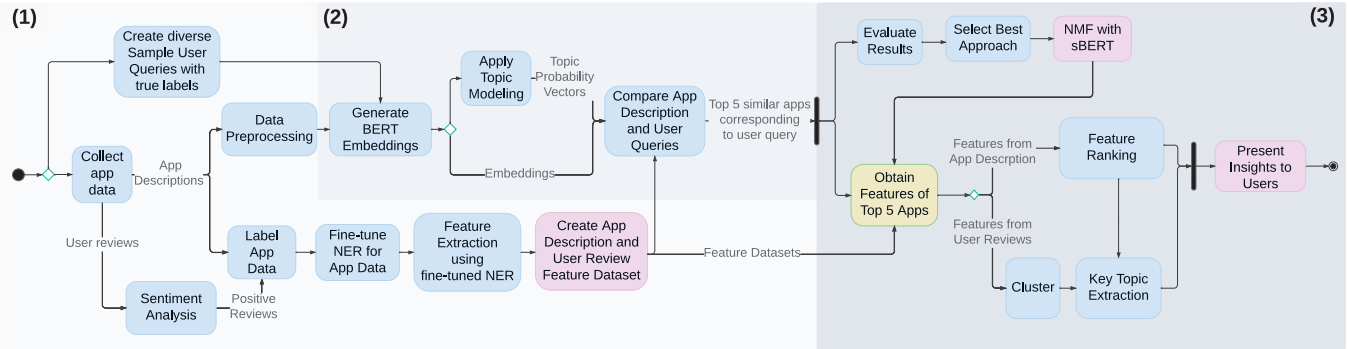

Figure 1: Step-wise Flow of Proposed Approach

cosine similarity followed by ranking. To recommend features, we present the top features from the five most similar apps. User reviews, clustered with K-means (k=6) and annotated with KeyBERT to highlight central themes.

The novelty of this study lies in the utilization of a custom-tailored NER model for feature extraction from app store data and applying Non-negative Matrix Factorization (NMF) with BERT embeddings to find similar apps, outperforming conventional LDA methods [3]. This method offers improved flexibility, adaptability, and resilience to noise for crafting personalized recommendations. Unlike popular models such as GPT and Falcon, our lightweight BERT models achieve an optimal balance between performance and real-time processing.

3 RESULTS AND ANALYSIS

Table 1 summarizes the average performance metrics for three BERT-based models—SBERT, RoBERTa, and DistilBERT—evaluated across three different pipelines: Simple BERT, LDA with BERT, and NMF with BERT. The NMF with BERT embeddings pipeline notably excels compared to the others, particularly in recall and F1-score. For example, SBERT combined with NMF achieved an F1-score of 87.38%, demonstrating NMF's effectiveness in capturing detailed features within complex datasets.

We evaluated our NMF with SBERT approach against the Data-based Raw Domain Model (DRDM) introduced by Liu et al. [5], which uses app descriptions to build a domain model for identifying similar apps. As shown in Table 2, our method significantly outperforms the DRDM across all key metrics, including precision, recall, and F-score. To further evaluate, we assessed the diversity of features recommended by our NER system using the Intra-List Diversity (ILD) metric (1). With an ILD score of *0.7564*, our system demonstrates a balanced and diverse set of suggestions, enhancing user satisfaction by minimizing redundancy.

$$\text{ILS}(P) = \frac{\sum_{p_i \in P} \sum_{p_j \in P,\, p_i \neq p_j} \text{sim}(p_i, p_j)}{2} \quad (1)$$

Table 1: Relative Performance of Three solution Pipelines

Pipeline	Model	Precision	Recall	F-Score
BERT	SBERT	89.00%	69.00%	68.00%
	RoBERTa	83.00%	57.00%	53.00%
	DistilBERT	84.00%	57.00%	52.00%
LDA with BERT	SBERT	81.00%	18.00%	17.00%
	RoBERTa	88.00%	15.00%	14.00%
	DistilBERT	96.00%	13.00%	9.00%
NMF with BERT	SBERT	98.63%	87.50%	87.38%
	RoBERTa	94.00%	56.00%	56.00%
	DistilBERT	95.00%	75.00%	69.00%

Table 2: Performance Results of APPFIRE vs DRDM

Method	Category	Precision	Recall	F-Score
APPFIRE	Business	100.00%	90.00%	91.00%
	Education	98.50%	82.50%	82.50%
	Tools	100.00%	100.00%	100.00%
	Entertainment	96.00%	77.50%	76.00%
DRDM	Travel	87.50%	80.77%	84.00%
	Education	85.71%	66.67%	75.00%
	Sports	88.24%	78.94%	83.33%
	Photograph	90.91%	71.43%	80.00%
Average (APPFIRE)		**98.63%**	**87.50%**	**87.38%**
Average (DRDM)		**88.09%**	**74.45%**	**80.58%**

REFERENCES

[1] A. Ali, Y. Xia, Q. Umer, and M. Osman. 2024. BERT based severity prediction of bug reports for the maintenance of mobile applications. *Journal of Systems and Software* 208 (Feb 2024), 111898.
[2] V. Alves et al. 2008. An Exploratory Study of Information Retrieval Techniques in Domain Analysis. In *12th International Software Product Line Conference*.
[3] H. Jiang, J. Zhang, X. Li, Z. Ren, D. Lo, X. Wu, and Z. Luo. 2019. Recommending new features from mobile app descriptions. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 28, 4 (2019), 1–29.
[4] H. Liu, Y. Wang, Y. Liu, and S. Gao. 2021. Supporting features updating of apps by analyzing similar products in App stores. *Information Sciences* 580 (2021), 129–151.
[5] Yuzhou Liu, Li Zhang, and Yong Zhang. 2019. Information recommendation based on domain knowledge in app descriptions for improving the quality of requirements. *IEEE Access* 7 (2019), 9501–9514.
[6] S. Scalabrino, G. Bavota, B. Russo, M. D. Penta, and R. Oliveto. 2019. Listening to the Crowd for the Release Planning of Mobile Apps. *IEEE Transactions on Software Engineering* 45, 1 (January 2019), 68–86.