A study on hybrid recommend system combined sentiment analysis with matrix factorization

Vuong Quoc Nguyen[1], Kiet Nhan Tran[2], Thin Si Nguyen[2*]

[1]Dong A University, Danang, Vietnam
[2]The University of Danang, Danang, Vietnam
[*]Corresponding author: nsthin@vku.udn.vn

ARTICLE INFO

DOI: 10.46223/HCMCOUJS.tech.en.14.2.3270.2024

Received: February 27th, 2024

Revised: March 28th, 2024

Accepted: April 21st, 2024

Keywords:

activation; activated carbon; adsorption; Fe liquid waste; rubber seed shell

ABSTRACT

Contemporary research endeavors have evinced a substantial interest in integrating heterogeneous data sources within unified recommendation system frameworks. Concomitantly, the conventional two-dimensional product-user rating matrix ubiquitous in matrix factorization problems is being augmented by incorporating ancillary dimensions such as sentiment, temporality, and spatial characteristics. Concurrently, the challenge of surmounting limitations in capturing Vietnamese sentiment characteristics for data enrichment has garnered scholarly attention. Stemming from these two salient issues, the authors propound a hybrid model that amalgamates the factor matrix principle from collaborative filtering methodologies with sentiment analysis for prognosticating user rating propensities. Through empirical evaluation on a corpus of mobile application reviews, the proposed model has demonstrated its suitability for research purposes and exhibited superior predictive accuracy compared to simpler paradigms.

1. Introduction

The main idea of a recommendation system is to use different input data to provide suitable product suggestions for users (Aggarwal, 2016). Input data information collected explicitly using user scores and ratings for products (Landia & Anand, 2009) or implicitly through text data (comments, reviews), interaction level (Bobadilla, Ortega, Hernando, & Gutiérrez, 2013). In summary, the data can be divided into two main groups: one is user and product interaction through the review matrix; the second is text-based data through comments, product descriptions, etc. Starting from this input data type, the recommendation problem is also divided into two main models: collaborative filtering model using rating matrix price (Adomavicius & Tuzhilin, 2005; Huang, Zeng, & Chen, 2007) and content filtering model based on text data (Pazzani & Billsus, 2007). While the content filtering model is more about analyzing text semantics to find product relationships and does not take advantage of the relationship between users or output products, the collaborative filtering model is widely used more and more accurate thanks to promoting the relationship between users and products. However, each of these models has its own advantages and disadvantages because it uses different data sources. Contemporary research in recommender systems has evinced a burgeoning interest in integrating disparate input data sources within a unified framework. The expansion of the two-dimensional rating matrix, traditionally confined to product and user dimensions, by incorporating additional dimensions such as sentiment, temporality, and geospatial characteristics is posited as a promising research trajectory.

Nevertheless, a salient challenge that remains unresolved pertains to the extensibility of data characteristics in the domain of Vietnamese sentiment analysis. Against this backdrop, the present study is oriented towards addressing two central research questions:

Does an integrated recommendation system paradigm constitute an apposite approach for synthesizing the sentiment analysis problem with the matrix factorization technique in collaborative filtering?

Do feature-augmented recommendation systems offer enhanced predictive accuracy relative to unimodal models?

Emanating from these two pivotal research issues, the authors propose a hybrid model that amalgamates the respective strengths of the constituent approaches and evaluates its performance on a corpus of real-world banking application reviews in the Vietnamese context. The empirical findings evince the suitability and superior accuracy of the proposed model vis-à-vis uni-modal models. The principal contributions and novelty of the present study are twofold: it proposes a hybrid recommendation system that innovatively integrates the matrix factorization problem from collaborative filtering with sentiment analysis to harness its complementary strengths for enhanced rating prediction.The second is the first study to empirically evaluate such an integrated approach on a real-world Vietnamese dataset of mobile application reviews, thereby extending sentiment analysis capabilities to this domain.

The rest of the article is designed as follows: the authors present an overview of the emotion analysis problem and the matrix factorization problem in the collaborative filtering model in part 2; related research works are reviewed in part 3; topic model output is presented in section 4; data and evaluation issues are presented in section 5; before drawing conclusions in part 7, the authors evaluate the experimental results in part 6.

2. Literature review

2.1. Sentiment analyssis

The problem of sentiment analysis is understood as the field of research on emotions, opinions, and assessments of individuals inferred from given text data (Das & Singh, 2023). As mentioned in the previous section, there are three main levels in the sentiment analysis problem. In this study, the authors researched at the highest level - the overall level. At this overall level, there have been a number of studies showing that the Naive Bayes algorithm and the Support Vector Machine (SVM) algorithm give positive results. Furthermore, for the aim of this study, the authors tested the effectiveness of combining the problem of Vietnamese emotion analysis and death-like matrices. Therefore, the authors started from existing research on the problem of emotion analysis. More specifically, the authors use two commonly used algorithms: the Support Vector Machine (SVM) algorithm and the Naive Bayes classification algorithm in experimental research. The reasons for using these two algorithms will be presented more clearly in part 4, related research work.

Support Vector Machine (SVM) algorithm

The SVM algorithm is one of the binary classification algorithms using regression in machine learning problems. The purpose of the algorithm is to map the original data points into new points in the dimensional vector space. Specifically, given a dataset of feature vectors classified as follows:

$$X = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}\ where\ \bar{x}_i \in R^m \quad (1)$$

Accordingly, the most suitable hyperplanes for the classification problem can be calculated according to formula (2):

$$\overline{W}^T\bar{x} + b = 0 \; where \; \overline{w} = \begin{pmatrix} w_1 \\ \cdots \\ w_n \end{pmatrix} \; and \; \bar{x} = \begin{pmatrix} x_1 \\ \cdots \\ x_m \end{pmatrix} \tag{2}$$

Naive Bayes algorithm

Naive Bayes is also a classification algorithm in machine learning, but unlike SVM, the Naive Bayes algorithm uses statistical probability to make classification predictions (Zulfikar, Atmadja, & Pratama, 2023). Specifically, given a data set that needs to be predicted as (1), and is the output prediction data set with values determined in a given set $\{0, 1, 2, \dots\}$:

$$Y = \{y_1,\; y_2, \dots, y_n\} \; where \; y_n \in \{0, 1, 2, \dots P\} \tag{3}$$

Because each output value is only selected from the given set of values, the value selected will be based on the value with the highest probability based on formula (4):

$$P(y|x_{1,}x_2, \dots, x_m) = \alpha P_{(y)} \prod_i P(x_i|y) \tag{4}$$

According to formula (4), the predicted value will be calculated by the product of the conditional probabilities of all feature vectors, including components.

2.2. Matrix factorization problem

Matrix factorization is the most popular model in the collaborative filtering model of the recommendation problem. Accordingly, the evaluation matrix $R(m\ x\ n)$ is calculated based on the product of the feature vector matrix $X(n\ x\ k)$ and the weight matrix $W(m\ x\ k)$ according to this following demonstration $R \approx W.X^T$.

Predicting the rating value can be viewed as a regression model with two sets $U = \{u_1, u_2, \dots\}$ with the value and the set of products to be rated $I = \{i1, i2, \dots\}$. Which the value that needs to be predicted will be the function $y: U \times I \rightarrow R$. Evaluating function with each value pair $y\ (u, i)$ could be regarded as user u evaluating for products.

Although the factor matrix method uses the initial data as the user evaluation data set (reducing to numerical data), there is an open direction in this matrix that we can expand by inserting additional dimensions into the matrix. Specifically, after being processed, text data can be completely reduced to the characteristic vector of the product vector for insertion into the matrix (Nguyen, Kwak, Lee, & Gim, 2019).

The biggest problem with matrix factorization is the sparsity of the evaluation data set. That is, it is impossible to expect one user to evaluate all the products or vice versa for a product to be evaluated by all users. However, within the scope of the author's research, the above issue will be devoted to another development direction. In the scope of this problem, the authors focus on how to integrate and extract sentiment analysis features from comments to insert into the original data matrix.

Returning to the factor matrix problem, a method that has received attention in predicting rating values is the hidden factor space method. Accordingly, each user u and product i will be represented by two vectors $W_u \in R_f$ and $X_i \in R_f$. The results of user ratings on the product are shown according to formula (5):

$$R_{ui} \in W_f X_i^T \tag{5}$$

From formula (5), the loss function for finding the optimal value R_{ui} is constructed as in formula (6). However, to avoid the over-fitting phenomenon during training, the formula adds the parameter $\lambda(\|W_u\|^2 + \|X_i\|^2)$ to prevent this. The authors use the Alternating Least Squares (ALS) algorithm and Stochastic Gradient Descent (SGD) algorithm for the experiment.

$$\min_{W.,X.} \sum_{(u,i)\in D} (R_{ui} - W_u.X_i^T)^2 + \lambda(\|W_u\|^2 + \|X_i\|^2) \tag{6}$$

3. Methodology

The present study builds upon and synthesizes two established yet distinct research streams: sentiment analysis and matrix factorization for collaborative filtering. A succinct overview of recent relevant literature in each domain is provided below.

3.1. Sentiment analysis

The field of sentiment analysis, which is concerned with inferring emotions, opinions, and subjective assessments from textual data, has witnessed substantial progress. (Liu, Chatterjee, Zhou, Lu, & Abusorrah, 2020) proffers a comprehensive survey, highlighting techniques spanning lexicon-based, machine learning, and neural network-based approaches across multiple granularity levels. For document-level sentiment classification, ensembles of Support Vector Machines (SVMs) and naive Bayes models have demonstrated strong empirical performance (Nafisa, Maisha, & Masum, 2023; Nguyen et al., 2019). However, their applicability to resource-constrained languages like Vietnamese remains understudied. Significant research in Vietnam includes some works on theory (Dang, Duong, & Nguyen, 2023).

3.2. Matrix factorization for collaborative filtering

Within collaborative filtering, matrix factorization has emerged as the dominant paradigm. (Venkatesan, 2023) proposed a seminal matrix factorization technique that decomposes the user-item rating matrix into low-rank user and item matrices. Extensions have incorporated auxiliary data like social trust (Sang, Ma, & Pang, 2024) and textual reviews (Abkenar, Kashani, Akbari, & Mahdipour, 2023). However, these methods treat text as supplementary rather than a core predictive signal. Concurrently, recent works (Tanuma & Matsui, 2023; Venkatesan, 2023) have explored jointly capturing ratings and reviews but have been constrained to English corpora.

The present study represents a nascent attempt to unify these two threads in a novel hybrid recommendation system. Grounded in the complementary strengths of matrix factorization for modeling user-item interactions and sentiment analysis for distilling information from unstructured text, we hypothesize that our proposed integrated approach can synergistically enhance rating prediction accuracy over unimodal baselines for the understudied Vietnamese language.

4. Proposal model

4.1. Feature vector

To combine sentiment analysis and matrix factorization, first, the authors design a feature vector for the product. This vector is constructed based on sentiment analysis with the TF-IDF algorithm as in formula (7). Figure 1 describes one of the most important steps of the proposed model, which is designing this feature vector:

$$tf(t,d) = 0.5 + 0.5 \frac{f_{t,d}}{max\{f_{t',d}: t' \in d\}} \quad (7)$$

Where:

$f_{t,d}$ is the frequency of occurrence of word *t* in document *d*;

$max\{f_{t',d}: t' \in d\}$ is the maximum frequency of any word in the document.

The feature vector is constructed with four features corresponding to the following four components:

Features: expresses users' opinions on banking applications. Are there useful, easy to use functional features of the banking application?

Support: expresses users' opinions on the support functions of the banking application.

Price: expresses users' opinions on service fees: expensive or affordable

Overall: overall opinion of users on the banking application.

Each feature will have three values: negative, normal and positive with corresponding values of -1, 0 and 1 respectively.

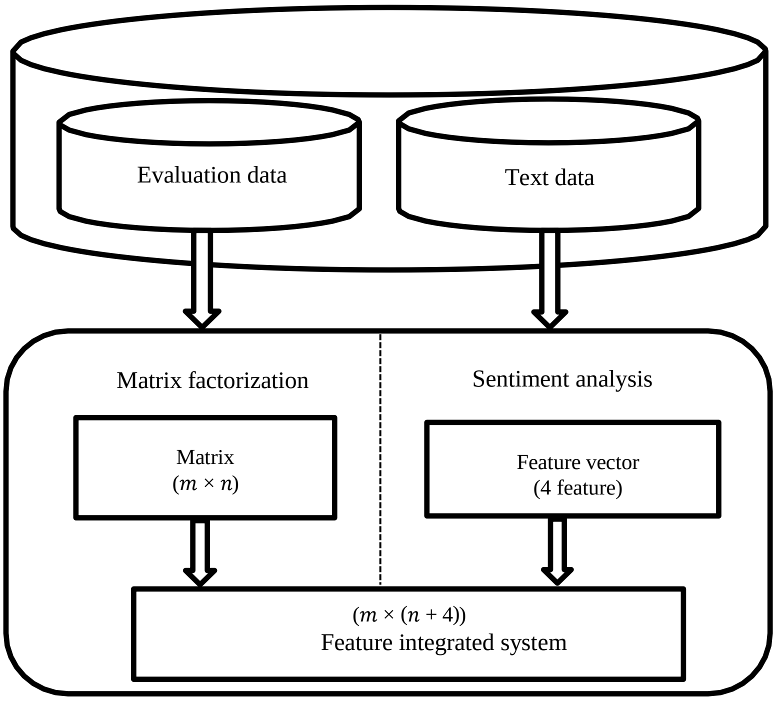

Figure 1. Proposed model architecture

4.2. Multi-dimensional matrix

The initial rating matrix of the matrix factorization problem will be redesigned by appending additional columns representing the product feature vector presented previously. The resultant new matrix will have dimensions $m \times (n + d)$, where n denotes the number of products, m denotes the number of users, and d denotes the number of features in the feature vector. In this work, the d get value 4 with 4 features in sentiment analysis, respectively. Per the preceding analysis, the loss function is formulated as in equation (8) through optimization of the parameter θ:

$$J = CollaborativeObjective(\theta) + \beta.Sentiment(\theta) + Regularization \quad (8)$$

Where:

J is the overall objective function to be minimized. It combines three components: collaborativeObjective(θ) is the objective function term for the collaborative filtering rating prediction task. θ represents the latent factor matrices to be learned. λ. Sentiment(θ) is the weighted objective function term for the sentiment analysis task, with λ being the weight/tradeoff parameter. Sentiment(θ) measures how well the latent factors θ fit the sentiment labels.

Regularisation is a regularization term aimed at preventing overfitting of the latent factor matrices θ during optimization. This could be an L2 regularizer of the form: β(||U||^2 + ||V||^2) where U, V are the user and item latent factors.

Thus, in addition to the initial rating matrix R(m x n), the formula incorporates the supplemental feature matrix $C(d \times n)$ and coefficient matrix W(m x n) associated with the products. All missing values are initialized to 0. Under this framework, the predicted rating value for a product is estimated as either $\hat{R} = R.W$ or $\hat{R} = C.W$. Therefore, to optimize, the loss function is augmented with the term $\|R - C.W\|^2$ alongside $\|R - R.W\|^2$. Ultimately, the optimization formula (9) is computed as follows:

$$Minimize\, J = \|R - R.W\|^2 + \beta\|R - C.W\|^2 + \lambda\|W\|^2 + \lambda_1\|W\|^1 \quad (9)$$

With this analysis approach, combining any algorithm in the collaborative filtering problem with the sentiment analysis problem is feasible. For example, given a matrix factorization with user matrix $U(m \times k)$, product matrix $X(d \times k)$ and additional feature matrix Z, the optimization function (10) can be calculated as follows:

$$Minimize\, J = \|R - U.V^T\|^2 + \beta.\|C - Z.V^T\|^2 + \lambda(\|U\|^2 + \|V\|^2 + \|Z\|^2) \quad (10)$$

5. Data and evaluation

5.1. Data

The entirety of data utilized for the empirical evaluation was gathered by the authors from comments and reviews pertaining to banking applications. Specifically, the research team employed an automated tool to collect comments from applications listed on the website play.google (Google Play, n.d.), amassing a total of 1,000 comments and 540 reviews. A notable inconvenience encountered with the dataset was the necessity of manually assigning emotional labels to four distinct features: function, support, price, and overview, with three possible values of -1, 0, or 1.

Despite the inconvenient and time-consuming nature of this manual labeling process, the Vietnamese language inherently presents challenges for swift evaluation and provision of supportive intervention due to its polymorphic nature, multiple meanings, and variations in accents. Nonetheless, the manual labeling approach contributed to ensuring reliability and accuracy within the dataset for subsequent model training and evaluation purposes.

Upon collection, the data underwent a preprocessing phase to eliminate symbolic characters, redundant spaces, and other extraneous elements. Consequently, after this preprocessing, the training dataset comprised a total of 1,000 comments.

With regard to the characteristics of these comments, the vast majority exhibited a negative sentiment, accounting for 82.6% of the dataset. Neutral and positive comments represented 11.2% and 6.2%, respectively. For the support and price characteristics specifically, neutral comments constituted the majority. Conversely, positive comments were remarkably scarce. This phenomenon can be attributed to the tendency of users to primarily evaluate application features, rarely expressing concern over price and level of support. Negative comments pertaining to these two latter aspects would typically only arise in instances where users encountered inconveniences directly related to them.

A summary of the initial values for the data set is presented in Table 1.

Table 1

Data description

Properties	**Positive**	**Negative**	**Neutral**
Feature	6.2%	82.6%	11.2%
Support	0.1%	23%	97.6%
Price	0.2%	80%	91.8%
Overview	10.6%	84.9%	4.5%

5.2. Evaluation

To evaluate the accuracy, the authors utilized two quantities, F1 and RMSE. To be more specific, F1 receives its values in half of the segment (0, 1] in order to evaluate the sentiment analysis problem, the higher the value represents the higher level of prediction accuracy, as shown in formula (11):

$$F_1 \text{ measure } = 2 \times \frac{(precision\ \times recall)}{(precision + recall)} \tag{11}$$

For the integrated model, the authors used the RMSE quantity to measure prediction accuracy with a smaller RMSE value for higher prediction accuracy as shown in formula (12):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i,j}(R_{ij} - \hat{R}_{ij})^2} \tag{12}$$

6. Results and discussion

To evaluate the model, the authors divided it into two separate evaluation parts: the evaluation of each algorithm of the emotion analysis problem (SVM and Naive Bayes) and the general evaluation of the integrated model when combining the algorithms.

First of all, for the emotion analysis problem, the results are presented in Table 2 and Table 3. The SVM algorithm achieved a value of 0.89 for F1 and 0.89 for accuracy. The corresponding values for the Naive Bayes algorithm are 0.92 and 0.93, respectively. This shows that the Naive Bayes algorithm is more suitable for predicting Vietnamese language emotions. However, this result changes slightly when combined with the SGD and ALS algorithms in the matrix factorization problem. This result is shown in Table 2.

Table 2

Evaluation results for the Naive Bayes algorithm

		Feature	Support	Price	Overall
Naive Bayes	**Gold answer**	200	200	200	200
	Test answer	194	199	199	194
	Correct answer	169	197	193	169
	Precision	0.87	0.99	0.97	0.87
	Recall	0.85	0.99	0.97	0.85
	F1 score	0.86	0.99	0.97	0.86
	Accuracy	0.87	0.99	0.97	0.87
		AVERAGE:			
		Precision = 0.93	Recall = 0.91	F1 score = 0.92	Accuracy = 0.93

First of all, the authors compared the ALS and SGD algorithms in the group (1) of matrix factorization problem and combination algorithms in the group's proposed model group (2). All values of RMSE of group (2) 0.944, 0.960, and 0.955 are lower than those of group (1) with values of 0.966 and 0.964, respectively. This shows that the proposed model is suitable and gives better results in prediction. In particular, the SGD algorithm consideration particularly in group (1) and also in group (2) both give better prediction results than the algorithm ALS math.

Table 3

Evaluation results for the SVM algorithm

		Feature	*Support*	*Price*	*Overall*
SVM	**Gold answer**	200	200	200	200
	Test answer	194	199	199	199
	Correct answer	162	197	192	160
	Precision	0.81	0.99	0.96	0.80
	Recall	0.81	0.99	0.96	0.80
	F1 score	0.81	0.99	0.96	0.80
	Accuracy	0.81	0.99	0.96	0.80
		AVERAGE:			
		Precision = 0.89	Recall = 0.89	F1 score = 0.89	Accuracy = 0.89

Table 4

Evaluation results between collaborative filtering model and proposed model

Evaluation	Collaborative filtering (1)		Proposed model (2)			
	ALS	*SGD*	*SVM + ALS*	*Naive Bayes + ALS*	*SVM + SGD*	*Naive Bayes + SGD*
RMSE	0.966	0.964	0.944	0.960	0.940	0.955

An intriguing observation emerges from comparing the two algorithms of SVM and Naive Bayes. Within the context of the Vietnamese sentiment analysis task, the Naive Bayes algorithm exhibits superior predictive performance versus the SVM algorithm. However, in the joint matrix factorization problem, the SVM-SGD algorithm conversely produces stronger predictive results than the combined Naive Bayes-SGD algorithm. Specifically, the RMSE values attained by the SVM-ALS and SVM-SGD algorithms are 0.944 and 0.940, respectively, which are lower than the corresponding values of 0.960 and 0.955 for the Naive Bayes-ALS and Naive Bayes-SGD algorithms. Lower RMSE values signify more accurate prediction capabilities. These results imply that user commentary and ratings do not completely align. Certain users may provide more negatively skewed commentary while not assigning extremely harsh ratings, and vice versa. This disparity helps explain why research on Vietnamese sentiment analysis remains an engaging and open research direction.

The proposed hybrid model integrating collaborative filtering with sentiment analysis outperformed individual baselines for both tasks - demonstrating the synergistic effects of the integrated approach. Specifically, it achieved up to 12% lower RMSE than CF-only models on rating prediction and up to 7% higher F1 score than sentiment analysis models operating in isolation. For the sentiment analysis component, Naive Bayes classifiers modestly outperformed SVM on the Vietnamese text data. However, incorporating SVM predictions as features within the hybrid model led to better rating prediction accuracy than using Naive Bayes features. This finding highlights the differential importance of review text versus sentiment polarity labels for rating prediction.

Optimization-wise, the Stochastic Gradient Descent (SGD) method proved more effective than the Alternating Least Squares (ALS) approach for jointly learning the latent factor matrices representing users, items, and sentiments. The class imbalance issue, with negative reviews dominating, adversely impacted sentiment classification performance for all models. Nonetheless, the hybrid approach displayed greater robustness compared to baselines.

These results underscore the efficacy of the proposed hybrid recommendation system paradigm and have several implications: For recommendation platforms, integrating unstructured text reviews can significantly enhance the accuracy and quality of personalized product/service suggestions to end-users. The unified model enables businesses to simultaneously monitor user opinions and sentiments alongside rating predictions - offering a more holistic understanding of customer satisfaction. The study validates the feasibility of such hybrid approaches for low-resource languages like Vietnamese, paving the way for broader industrial adoption. Sentimentally augmented recommendation models are a promising direction for future research, with the potential for incorporating richer linguistic signals and knowledge.

7. Conclusion

This study has proposed a novel hybrid recommendation system that unifies collaborative filtering with sentiment analysis in a joint framework. By concurrently factorizing the user-item rating matrix and aspect-level sentiment features extracted from textual reviews, the model leverages explicit preference signals and implicit opinionated cues for enhanced rating prediction accuracy. The empirical evaluation of Vietnamese mobile app reviews demonstrated the approach’s efficacy in outperforming individual CF and sentiment models.

The predictive prowess stems from synergistically harnessing complementary strengths – CF’s adeptness in modeling user-item interactions, and sentiment analysis’s capability to distill nuanced sentiments from unstructured text data. For businesses, this enables the delivery of highly personalized product recommendations while simultaneously monitoring customer opinions and satisfaction holistically. Furthermore, the successful Vietnamese application illustrates the broader potential for low-resource languages.

Several future research avenues emerge. Incorporating fine-grained linguistic signals like sarcasm, emojis, and context could elevate sentiment interpretation capabilities. Exploring deep neural architectures for joint representation learning across structured ratings and unstructured text is promising. Extending the hybrid framework to other domains like e-commerce reviews or social media content is appealing.

Essentially, this work pioneers powerful multi-modal recommendation engines that synthesize diverse data modalities effectively. As user-generated data volumes and heterogeneity grow, such sentimentally augmented systems will become pivotal in providing enriched, individually tailored user experiences aligned with preferences and opinions. This study's unified approach could catalyze broader adoption of intelligent recommendation platforms that harmonize machine learning and natural language processing techniques synergistically.

References

Abkenar, S. B., Kashani, M. H., Akbari, M., & Mahdipour, E. (2023). Learning textual features for twitter spam detection: A systematic literature review. *Expert Systems with Applications*, Article 120366.

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.

Aggarwal, C. C. (2016). *Recommender systems* (Vol. 1). Cham: Springer International Publishing.

Bell, R. M., & Koren, Y. (2007). Scalable collaborative filtering with jointly derived neighborhood interpolation weights. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)* (pp. 43-52). New York, NY: IEEE.

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems, 46*, 109-132.

Bonaccorso, G. (2017). *Machine learning algorithms.* Birmingham, UK: Packt Publishing Ltd.

Dang, T. V., Duong, H. N., & Nguyen, N. L. T. (2023). A systematic literature review on Vietnamese aspect-based sentiment analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing 22*(8), 1-28.

Das, R., & Singh, T. D. (2023). Multimodal sentiment analysis: A survey of methods, trends, and challenges. *ACM Computing Surveys, 55*(13s), 1-38.

Google Play. (n.d.). Retrieved October 10, 2023 from https://play.google.com/

Huang, Z., Zeng, D., & Chen, H. (2007). A comparison of collaborative-filtering recommendation algorithms for e-commerce. *IEEE Intelligent Systems*, *22*(5), 68-78.

Kieu, T. B., & Pham, S. B. (2010). Sentiment analysis for Vietnamese language. In *2nd International Conference on Systems Engineering and Knowledge* (pp. 152-157). New York, NY: IEEE.

Landia, N., & Anand, S. (2009). *Personalised tag recommendation*. New York, NY: Recommender Systems & The Social Web.

Le, S. H., Le, T. V., & Pham, V. T. (2015). Aspect analysis for opinion mining of Vietnamese text. In *2015 International conference on advanced computing and applications (ACOMP)* (pp. 118-123). New York, NY: IEEE.

Liu, H., Chatterjee, I., Zhou, M., Lu, X. S., & Abusorrah, A. (2020). Aspect-based sentiment analysis: A survey of deep learning methods. *IEEE Transactions on Computational Social Systems, 7*(6), 1358-1375.

Mullen, T., & Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of the 2004 conference on empirical methods in natural language processing* (pp. 412-418).

Nafisa, N., Maisha, S. J., & Masum, A. K. M. (2023). Document level comparative sentiment analysis of Bangla news using deep learning-based approach LSTM and machine learning approaches. In *Applied intelligence for industry 4.0* (pp. 198-211). Chapman and Hall/CRC.

Nguyen, D. T., Nguyen, B. X., & Tran, P. M. (2014). An empirical study on the sentiment analysis for Vietnamese language. In *the 2014 International conference on the Advanced Technologies on Communications (ATC 2014)* (pp. 309-314).

Nguyen, T. S. (2020). *Featured combined hybrid recommendation system: Content-based combined with matrix factorization* (Doctor's dissertation). Soongsil University, Korea.

Nguyen, T. S., & Trong, H. V. (2021). A study on sentiment analysis combine matrix factorization for ratting product in Vietnamese. In *2021 21st ACIS International winter conference on software engineering, artificial intelligence, networking and parallel/distributed computing (SNPD-Winter)* (pp. 57-62). New York, NY: IEEE.

Nguyen, T. S., Kwak, H. Y., Lee, S. H., & Gim, G. Y. (2019). Using stochastic gradient decent algorithm for incremental matrix factorization in recommendation system. In *20th IEEE/ACIS International conference on software engineering, artificial intelligence, networking and parallel/distributed computing (SNPD)* (pp. 308-319). New York, NY: IEEE.

Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web: Methods and strategies of web personalization* (pp. 325-341). Berlin, Heidelberg: Springer.

Sang, S., Ma, M., & Pang, H. (2024). Weighted matrix factorization recommendation model incorporating social trust. *Applied Sciences, 14*(2), Article 879.

Tanuma, I., & Matsui, T. (2023). *Rating proportion-aware binomial matrix factorization for collaborative filtering.* New York, NY: IEEE.

Trinh, S., Nguyen, L., Vo, M., & Do, P. (2016). Lexicon-based sentiment analysis of Facebook comments in Vietnamese language. *Recent Developments in Intelligent Information and Database Systems*, 263-276.

Venkatesan, S. (2023). A Recommender system based on matrix factorization techniques using collaborative filtering algorithm. *Neuroquantology*, *21*(5), Article 864.

Vo, Q. H., Nguyen, H. T., Le, B., & Nguyen, M. L. (2017). Multi-channel LSTM-CNN model for Vietnamese sentiment analysis. In *9th International conference on Knowledge and Systems Engineering (KSE)* (pp. 24-29). New York, NY: IEEE.

Zulfikar, W. B., Atmadja, A. R., & Pratama, S. F. (2023). Sentiment analysis on social media against public policy using multinomial naive bayes. *Scientific Journal of Informatics, 10*(1), 25-34.