Integrating Sentiment Features in Factorization Machines: Experiments on Music Recommender Systems

JAVIER WANG, Universidad Autónoma de Madrid, Spain
ALEJANDRO BELLOGÍN, Universidad Autónoma de Madrid, Spain
IVÁN CANTADOR, Universidad Autónoma de Madrid, Spain

Music recommender systems play a pivotal role in catering to diverse user preferences and fostering personalized listening experiences. At the same time, sentiments can profoundly influence music by shaping its emotional expression and evoking specific moods onto listeners. Expressed in textual content, these sentiments may be analyzed through natural language processing techniques to gauge emotions or opinions, hopefully increasing their relevance when exploited for recommendation. This work aims to investigate how to better integrate such information and understand its potential impact on personalized music suggestions, attempting to enhance recommendation models by incorporating sentiment features into factorization machines. For this purpose, a dataset was collected from Last.fm and enhanced with sentiment information extracted from Wikipedia. Empirical results evidence that not all sentiment-related features are equally useful, showing that each tested factorization machine approach varies in sensitivity to these features. Source code and data are available at https://github.com/abellogin/SentiFMRecSys.

CCS Concepts: • **Information systems → Recommender systems**.

Additional Key Words and Phrases: Recommender systems, music recommendation, sentiment analysis, factorization machines

ACM Reference Format:
Javier Wang, Alejandro Bellogín, and Iván Cantador. 2024. Integrating Sentiment Features in Factorization Machines: Experiments on Music Recommender Systems. In *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP '24), July 1–4, 2024, Cagliari, Italy*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3627043.3659561

1 INTRODUCTION

The field of music recommender systems has witnessed significant advancements in recent years [31], with effective algorithms to provide personalized suggestions of tracks to users, on popular music social networks like Last.fm or music streaming platforms such as Spotify. Music is inherently emotional and has the ability to evoke strong feelings in listeners [19]. Therefore, understanding the sentiment of music could provide valuable insights about the emotional characteristics and appeal of different songs [30, 16]. By applying sentiment analysis techniques on music textual content, we could identify sentiment factors of tracks, such as the levels of happiness, sadness, excitement, or relaxation. This information may be used to tailor recommendations based on the user's current emotional state or mood, entailing a more personalized and engaging music experience.

While the role of emotions and mood has been widely studied by the recommender systems community [23], in general, a user perspective has been considered, that is, studying how user situations and contexts affect the user experience when dealing with music recommendations [1, 4]. Moreover, in order to capture and exploit sentiment in a recommender system, researchers have tended to rely on textual reviews, as stated in [33], which limits the application of proposed recommendation approaches to cases where reviews are not so common.

Differently to previous studies, the primary goal of this work is to explore and understand the effectiveness of exploiting sentiment features in the domain of music recommendation, by investigating its effects on different recommendation approaches. For such purpose, factorization machines [25] are particularly useful, as they are able to seamlessly embed additional features into the recommendation process, and have only been applied in music recommendation with user information, not item characteristics [31]. This is accomplished through the collection of a new dataset using Last.fm's API, and the extraction of sentiment features with a text sentiment analyzer from Wikipedia summaries, thus not relying on textual reviews. This dataset is then used to build different sentiment models, and to test the integration of sentiment information through the evaluation of recommendation accuracy.

More specifically, we aim to answer the following two research questions:

- **RQ1. How sentiment information can be automatically acquired for music items and modeled to its exploitation by personalized recommender systems?** To address this question, in Section 2, we propose a simple method based on tags to extract sentiment models in agreement with previous proposals from the literature, namely the Valence-Arousal (VA) [29] and Valence-Arousal-Dominance (VAD) [18] models.
- **RQ2. Can the performance of music recommendation methods be improved through the integration of sentiment features?** Throughout empirical experiments on the built dataset (presented in Section 5), we contrast our proposal to integrate sentiment features for recommendation, as shown in Section 3.

In summary, by embedding sentiments into music recommendation, we aim at measuring the effectiveness of such musical properties, and the role they play in the representation of items by a recommender through the generation of suitable candidates that cater to the user preferences. We thus contribute to the development of music recommenders that not only consider user-item information and interactions, but also incorporate sentiment aspects of music.

2 EXTRACTING SENTIMENT FEATURES FOR MUSIC RECOMMENDATION

Sentiment analysis, also known as opinion mining, aims to identify the sentiment or subjective information expressed in textual content. Among other tasks, it involves analyzing a piece of text to identify its overall sentiment, such as positive, negative, or neutral, and sometimes even more nuanced emotions. One popular model used in sentiment analysis is the VAD model. Its traditional version assigns scores to words based upon three dimensions: valence, arousal and dominance [18]. Valence represents the pleasantness or positivity of the sentiment, arousal represents the level of excitement or intensity, and dominance represents the degree of control or influence.

Frequently, the valence and arousal dimensions have been considered sufficiently independent to convey mood and musical emotions, since they allow representing a reliable sentiment spectrum, striking a balance between complexity and predictive potential [29]; with dominance being necessary for a wider range of emotions, at the expense of simplicity and scoring accuracy. Their scores provide a comprehensive understanding of the emotions conveyed in a text [28].

A common approach in sentiment analysis is computing a sentiment score from the frequency of positive and negative words in the text. This score can be computed by subtracting the number of negative words from the number of positive words and dividing it by the total number of words in the text. This technique provides a sentiment ratio, a quantitative measure of sentiment polarity that is akin to valence, but ignores neutral words [37].

$$Sentiment\ Ratio = \frac{|Positive\ words| - |Negative\ words|}{|Total\ words|} \in [-1, 1] \tag{1}$$

In this work, the source of information sentiment will be inferred from tags and associated texts. More specifically, for a given tag, we access its Wikipedia page and retrieve the first paragraph, since it usually corresponds to a definition of the concept depicted in that page. We make use of Wikipedia's API [6] to make this process automatic. Such process, although only considered for extracting sentiment features for music recommendation, could be applied to other domains beyond music, thanks to the (almost) universal coverage of Wikipedia. Nonetheless, it must be considered that, at least in music, it has been studied that tags do not capture the distinction between emotions evoked and emotions perceived by music [39]; hence, such limitations should be considered when adapting this approach to other domains.

Once we collected the text to be analyzed, we extracted the sentiment features through a custom spaCy[1] NLP pipeline that takes care of several processing tasks, which are essentially trained models that focus on specific goals. In particular, we developed an analyzer that leverages WordNet, a corpus that includes synsets or synonym sets, which allow searching for words of similar meaning derived from their lemma and Part-of-Speech tag. Upon execution, the analyzer loads the NRC-VAD lexicon [41], a collection of over 20,000 words to which fine-grained VAD scores were manually assigned, in the form of emotion-aware word embeddings. To increase analysis coverage, our analyzer incorporates synonym search functionality. When a word is not found in the lexicon, the analyzer makes use of synsets to retrieve words related to the original lemma, effectively expanding the search scope and providing a broader analysis of sentiment nuances. Finally, after processing all tokens in the text, the analyzer generates the final Valence, Arousal, Dominance, and Sentiment ratio (VADS) scores, providing an answer to RQ1 (*How sentiment information can be automatically acquired for music items and modeled to its exploitation by personalized recommender systems?*).

3 INTEGRATING SENTIMENT FEATURES IN FACTORIZATION MACHINES

Factorization machines are Machine Learning algorithms that combine the advantages of Support Vector Machines and factorization models [25]. When applied to recommendation, they allow alleviating sparsity issues since the classical sparse matrix representation is replaced by dense attribute vectors. Moreover, they allow extending user, item or interaction data with any number of attributes in a straightforward way.

Such flexibility is exploited in this work to include the sentiment features defined in the previous section as part of the attribute vectors. Besides the classical Factorization Machine (FM) technique [25], we also consider other related approaches to understand which of them may better model users and items according to the sentiment information, all of them exploiting deep learning architectures in different ways. More specifically, in our empirical study, we include Neural Factorization Machines (NFM) [10], specialized on performing prediction under sparse settings, as it seamlessly combines the linearity of FM in modeling second-order feature interactions and the non-linearity of neural network in modeling higher-order feature interactions. We also consider a method proposed by Google, Wide & Deep Learning (WideDeep) [3], which jointly trains wide linear models and deep neural networks to combine the benefits of memorization and generalization for recommender systems. FM based on Deep Learning (DeepFM) [9] combines the power of factorization machines for recommendation and deep learning for feature learning in a new neural network architecture. Compared to the previous model (WideDeep), DeepFM has a shared input to its "wide" and "deep" parts, with no need of feature engineering besides raw features. An extension of DeepFM, called eXtreme Deep Factorization Machine (xDeepFM) [14], is able to learn certain bounded-degree feature interactions explicitly, while learning arbitrary low- and high-order feature interactions implicitly, into one unified model. An improved version of Deep & Cross

[1]https://spacy.io

Table 1. Statistics of the dataset extracted for the experiments. C stands for count and CwS for count where sentiment is available. NA denotes that count does not apply, as users do not have sentiment attributes associated.

Entity	C	CwS
User	52,829	NA
Track	815,631	361,091
Artist	162,702	122,646
Album	384,995	126,572
Tags	199,608	177,106

Interactions	C
User-track	2,452,162
User-artist	527,051
User-album	526,855

Relationships	C
Tag-track	2,512,453
Tag-artist	980,305
Tag-album	733,505

Network (DCN V2) [36] was also included, as it automatically and efficiently learns bounded-degree predictive feature interactions while being cost-efficient in comparison with its previous version (DCN, [35]).

We also tested Attentional Factorization Machines (AFM) [38], which learn the importance of each feature interaction from data via a neural attention network. Even though Product-based Neural Networks (PNN) [24] might seem a completely different approach, it also learns a distributed representation of the categorical data, a product layer to capture interactive patterns between inter-field categories, and further fully connected layers to explore high-order feature interactions, so it can be categorized as a neural FM. Similarly, Field-aware Factorization Machines (FFM) [13] is a generalization of Personalized Tag Recommendation approach from [26], where the factor model is not only applied to user-item, user-tag, and item-tag pairs, but to any other attribute in the data.

4 EXPERIMENTAL SETTINGS

4.1 Dataset

We collected a new dataset from Last.fm that allowed us to integrate enough relevant sentiment data needed for our study, such as user-item interactions and item features. As shown in Table 1, the dataset includes a large number of user-item interactions, with several relationships that might be exploited in the future. The dataset is available in the paper repository.

To collect our dataset, we started retrieving the top 50 chart tags using Last.fm's API[2], representing the most used tags at the time. For each tag, we obtained the top unique artists and the top-30 listeners. Then, we used the API to gather data from the top listeners, acquiring their top 20 tracks, recent tracks, and loved tracks, each with the corresponding timestamp, and artist and album information. Additionally, we collected the top 10 artists and albums for each listener[3]. Finally, we fetched the top 10 tags assigned by users to each unique track, artist, and album (these last two are considered to avoid data sparsity), and their associated definitions from Wikipedia, as explained in Section 2. From these tags, we removed those making a reference to the item itself (e.g., best 2020 album, trending artist) and those assigned to only one item.

The developed sentiment analyzer was applied over the Wikipedia texts to extract VADS scores, where for each unique track, artist and album, weighted averages of the sentiment features were computed based on the associated tags. The importance of each tag was determined by its rank (in an exponential fashion), with the higher-ranking tags carrying much more weight as they are more representative.

[2]https://www.last.fm/api
[3]We acknowledge this methodology may be prone to overly consider popular music tracks and users who listen to popular tracks; we aim to address this limitation in the future.

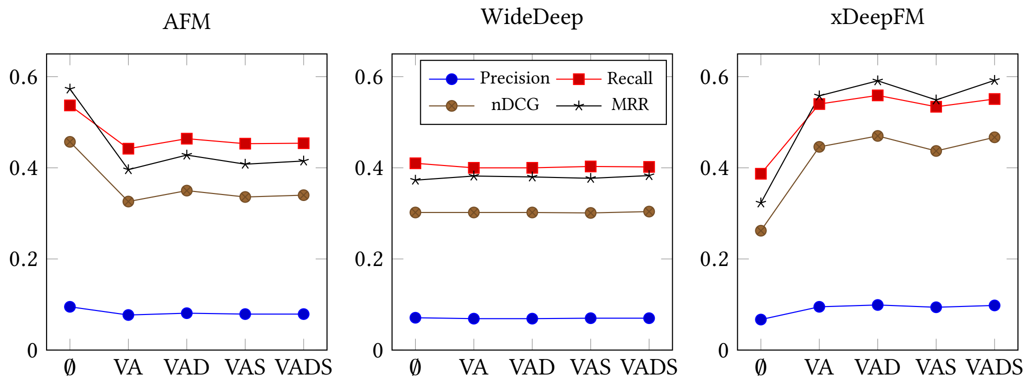

Fig. 1. Performance variation for 3 factorization approaches when no sentiment features are used (∅) or when combinations of Valence, Arousal, Dominance, and Sentiment ratio are considered.

4.2 Recommendation methods

Regarding the recommendation approaches explained in Section 3, we used the implementations provided by RecBole [40]. The training process involved cross-validation with sets for training, validation and testing. After each training epoch, validation was conducted, and if the recommendation performance scores did not improve within a specified number of epochs (10), the training concluded; after which the models were evaluated on the testing set. Additionally, negative sampling was employed to help models distinguish relevant items from irrelevant ones.

4.3 Evaluation metrics

In the conducted experiments, we focused on recommendation accuracy, measuring popular evaluation metrics like nDCG, Recall, Precision, and MRR [8], on an 80-10-10 training-validation-test split. As these metrics were computed on ranked recommendation lists, we only considered the first 20 items of each list, i.e., at cutoff 20, to obtain more robust measurements than with smaller cutoffs [34]. The metric nDCG@20 was used for early training stopping.

5 PERFORMANCE EFFECT OF SENTIMENT FEATURES IN MUSIC RECOMMENDATION

Table 2 reports the performance results for all the factorization techniques included in our experiments, and shows how they are affected by integrating sentiment features. Figure 1 shows 3 of the 9 tested recommenders and how they are affected by integrating sentiment features. We observe that, in terms of nDCG@20, the trend in performance is very similar for the four evaluation metrics reported.

To answer RQ2 (*Can the performance of music recommendation methods be improved through the integration of sentiment features?*), we observe in our results that, for some recommendation approaches, including sentiment features do produce a positive improvement in the performance with respect to the base result. This is true for all cases except AFM and FFM. In those specific two situations, we note that the Dominance feature (VAD) produces better results than Sentiment ratio (VAS) or when the 4 features are combined (VADS). For the rest of the cases, VADS tends to be the best performing combination, although this, again, depends on the recommendation approach (for example, VA is the best one with NFM, VAD with xDeepFM, and VAS with PNN).

More specifically, we could group the tested recommenders according to their sensitivity to the sentiment features:

Table 2. Performance comparison measured with nDCG@20 among all the factorization approaches when no sentiment features are considered (∅) or when combinations of sentiment features are used (notation as in Figure 1). Best value for each approach in bold.

	Sentiment features				
Method	**∅**	**VA**	**VAD**	**VAS**	**VADS**
AFM	**0.457**	0.326	0.350	0.336	0.340
DCN V2	0.264	0.307	0.293	0.277	**0.325**
DeepFM	0.296	0.326	0.315	0.305	**0.335**
FFM	**0.434**	0.395	0.387	0.385	0.383
FM	0.392	0.360	0.390	0.392	**0.399**
NFM	0.256	**0.379**	0.327	0.317	0.319
PNN	0.303	0.461	0.470	**0.478**	0.473
WideDeep	0.302	0.302	0.302	0.301	**0.304**
xDeepFM	0.262	0.446	**0.470**	0.437	0.467

- Negative effect (AFM and FFM): no sentiment features outperform the baseline result where only user-item interactions are considered.
- Neutral effect (FM and WideDeep): slight improvements are observed but they are not consistent, as performance may also deteriorate.
- Positive effect (DCN V2, DeepFM, NFM, PNN, xDeepFM): a clear improvement is obtained when compared the baseline result with any of the situations where two or more of the sentiment features are considered. These improvements could go up to 80%, as for the case of xDeepFM with VAD.

In summary, from our empirical results, we conclude that the performance of music recommendation approaches could be improved by integrating sentiment features in their models. Except for some algorithms, exploiting the four sentiment features analyzed in this work (valence, arousal, dominance, and sentiment ratio) at the same time achieves the best results by producing a more complete overview of the user preferences with respect to the item characteristics, which could be further modeled by the recommendation approach at hand; in this work, limited to factorization machines.

6 RELATED WORK

6.1 Music recommendation

Music recommendation is a specialized domain of recommender systems that focuses on providing personalized music suggestions to users. Modern day advancements in music recommendation involve sophisticated algorithms and machine learning models, as well as rich music data to enhance recommendation accuracy and user satisfaction [31]. It is acknowledged that content (particularly, derived from the items, i.e., tracks) is very important in this domain, either as metadata information or as semantic descriptors [31].

This is why not only collaborative filtering approaches are popular, but content-based methods and hybrid techniques have been proposed and successfully developed. Special attention deserve methods based on deep learning, capable of integrating audio content and collaborative information, together with other metadata, such as [20, 11].

More recently, approaches considering the personality traits of users or their affective state (mood or emotion) aim to provide a more personalized music recommendation experience, by adapting the level of diversity [17], or by integrating the previous affective responses of users to recommended songs in order to adapt for the future [1].

6.2 Sentiment-based recommender systems

The ubiquity of recommender systems open up opportunities to exploit and analyze different sources of information beyond the classical ratings, to identify user-item interactions and their importance. Sentiment, as motivated in this work, is a relevant aspect to consider, and a dimension that is particularly useful, although it might be difficult to capture or infer. In fact, most of the literature requires user comments or reviews to extract sentiment or opinions [33, 32, 21, 7, 12], a data source that is not available in many datasets and domains. In those works, sentiment analysis is used to reduce sparsity of the user-item matrix [21], to enhance the final recommendations [7], to filter out negative items [32], or to produce an expert graph based on item polarity and popularity [12].

Even though sentiment is extracted from items, it has been applied to users in one way or another, in contrast to our proposal. A different, but related concept, is emotion. In this case, as discussed in the introduction, this dimension has been mostly analyzed from the user perspective; however, it is relevant because it is showed prominent in the field of music recommendation. For example, in [4] authors infer the user emotion from microblogs, which are later used in a modified item-based collaborative filtering and to perform a random walk (PageRank) on the user-emotion-music graph. In [23], a general emotion-aware computational model based on affective user profiles is presented. More recently, in [15] a deep learning technique is used to select the most suitable music based on users' mood in previous period and current emotion stimulus; such mood is extracted from low-level features from the music listened by the user.

The closest work to ours may be [5], as it does not consider a user perspective. The authors generate a core emotion lexicon inspired by the Valence-Arousal model, which is used to transform item tags into emotion-oriented item profiles. This approach could complement our proposal in how sentiment features are extracted, while expanding on the current experimental comparison with factorization machines. In [5], however, the authors only considered a binary recommendation problem (i.e., classification) instead of the more typical task nowadays: item ranking.

7 CONCLUSIONS AND FUTURE WORK

In this work, we have tackled the task of music recommendation by studying how to incorporate sentiment features into the items in a system. In this context, we first proposed a simple method that generates sentiment features from tags by exploiting Wikipedia summaries. This method could be applied beyond the music domain or with other music datasets, with none to minimal changes. Then, we experimented with nine recommendation algorithms based on factorization machines, concluding that, except for two cases, integrating sentiment features through our proposal either does not decrease the performance or help to achieve performance improvements, in some cases up to an 80%.

We believe our study could open up further possibilities in the field of user and item modeling by promoting the use of factorization machines for these types of features. Additionally, sentiment analysis could potentially assist in addressing the cold-start problem in recommender systems, where limited user data is available [27].

In the future, we would like to explore other information sources to extract the sentiment features from, such as the track lyrics. However, we anticipate this may reinforce the coverage problems we already observed with respect to track tags. We also plan to test other proposals of factorization machines that match our formulation, such as Field-weighted Factorization Machines [22], and analyze the impact of sentiment features in beyond-accuracy metrics [2].

ACKNOWLEDGMENTS

Work supported by grant PID2022-139131NB-I00 funded by MCIN/AEI/10.13039/ 501100011033 and "ERDF A way of making Europe."

REFERENCES

[1] Deger Ayata, Yusuf Yaslan, and Mustafa E. Kamasak. 2018. Emotion based music recommendation system using wearable physiological sensors. *IEEE Trans. Consumer Electron.*, 64, 2, 196–203. DOI: 10.1109/TCE.2018.2844736.

[2] Pablo Castells, Neil Hurley, and Saúl Vargas. 2022. Novelty and diversity in recommender systems. In *Recommender Systems Handbook*. Francesco Ricci, Lior Rokach, and Bracha Shapira, (Eds.) Springer US, 603–646. DOI: 10.1007/978-1-0716-2197-4_16.

[3] Heng-Tze Cheng et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, DLRS@RecSys 2016, Boston, MA, USA, September 15, 2016*. Alexandros Karatzoglou, Balázs Hidasi, Domonkos Tikk, Oren Sar Shalom, Haggai Roitman, Bracha Shapira, and Lior Rokach, (Eds.) ACM, 7–10. DOI: 10.1145/2988450.2988454.

[4] ShuiGuang Deng, Dongjing Wang, Xitong Li, and Guandong Xu. 2015. Exploring user emotion in microblogs for music recommendation. *Expert Syst. Appl.*, 42, 23, 9284–9293. DOI: 10.1016/J.ESWA.2015.08.029.

[5] Ignacio Fernández-Tobías, Iván Cantador, and Laura Plaza. 2013. An emotion dimensional model based on social tags: crossing folksonomies and enhancing recommendations. In *E-Commerce and Web Technologies - 14th International Conference, EC-Web 2013, Prague, Czech Republic, August 27-28, 2013. Proceedings* (Lecture Notes in Business Information Processing). Christian Huemer and Pasquale Lops, (Eds.) Vol. 152. Springer, 88–100. DOI: 10.1007/978-3-642-39878-0_9.

[6] Jonathan Goldsmith and Wikimedia Foundation. 2013. Wikipedia. https://github.com/goldsmith/Wikipedia. (2013).

[7] Ananth Gouri S, Dr Raghuveer, and Dr Vasanth Kumar S. 2023. Fusion of various sentiment analysis techniques for an effective contextual recommender system. In *Proceedings of the 16th Innovations in Software Engineering Conference*, 1–8.

[8] Asela Gunawardana, Guy Shani, and Sivan Yogev. 2022. Evaluating recommender systems. In *Recommender Systems Handbook*. Francesco Ricci, Lior Rokach, and Bracha Shapira, (Eds.) Springer US, 547–601. DOI: 10.1007/978-1-0716-2197-4_15.

[9] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. Deepfm: A factorization-machine based neural network for CTR prediction. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*. Carles Sierra, (Ed.) ijcai.org, 1725–1731. DOI: 10.24963/IJCAI.2017/239.

[10] Xiangnan He and Tat-Seng Chua. 2017. Neural factorization machines for sparse predictive analytics. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*. Noriko Kando, Tetsuya Sakai, Hideo Joho, Hang Li, Arjen P. de Vries, and Ryen W. White, (Eds.) ACM, 355–364. DOI: 10.1145/3077136.3080777.

[11] Qingqing Huang, Aren Jansen, Li Zhang, Daniel P. W. Ellis, Rif A. Saurous, and John R. Anderson. 2020. Large-scale weakly-supervised content embeddings for music recommendation and tagging. In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2020, Barcelona, Spain, May 4-8, 2020*. IEEE, 8364–8368. DOI: 10.1109/ICASSP40776.2020.9053240.

[12] Rizwana Irfan, Osman Khalid, Muhammad Usman Shahid Khan, Faisal Rehman, Atta Ur Rehman Khan, and Raheel Nawaz. 2019. Socialrec: a context-aware recommendation framework with explicit sentiment analysis. *IEEE Access*, 7, 116295–116308.

[13] Yu-Chin Juan, Yong Zhuang, Wei-Sheng Chin, and Chih-Jen Lin. 2016. Field-aware factorization machines for CTR prediction. In *Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, September 15-19, 2016*. Shilad Sen, Werner Geyer, Jill Freyne, and Pablo Castells, (Eds.) ACM, 43–50. DOI: 10.1145/2959100.2959134.

[14] Jianxun Lian, Xiaohuan Zhou, Fuzheng Zhang, Zhongxia Chen, Xing Xie, and Guangzhong Sun. 2018. Xdeepfm: combining explicit and implicit feature interactions for recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (KDD '18). Association for Computing Machinery, London, United Kingdom, 1754–1763. ISBN: 9781450355520. DOI: 10.1145/3219819.3220023.

[15] Zhiyuan Liu, Wei Xu, Wenping Zhang, and Qiqi Jiang. 2023. An emotion-based personalized music recommendation framework for emotion improvement. *Inf. Process. Manag.*, 60, 3, 103256. DOI: 10.1016/J.IPM.2022.103256.

[16] Adam J. Lonsdale and Adrian C. North. 2011. Why do we listen to music? a uses and gratifications analysis. *British Journal of Psychology*, 102, 1, 108–134. eprint: https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1348/000712610X506831. DOI: https://doi.org/10.1348/000712610X506831.

[17] Feng Lu and Nava Tintarev. 2018. A diversity adjusting strategy with personality for music recommendation. In *Proceedings of the 5th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, IntRS 2018, co-located with ACM Conference on Recommender Systems (RecSys 2018), Vancouver, Canada, October 7, 2018* (CEUR Workshop Proceedings). Peter Brusilovsky, Marco de Gemmis, Alexander Felfernig, Pasquale Lops, John O'Donovan, Giovanni Semeraro, and Martijn C. Willemsen, (Eds.) Vol. 2225. CEUR-WS.org, 7–14. https://ceur-ws.org/Vol-2225/paper2.pdf.

[18] Albert Mehrabian. 1980. Basic dimensions for a general psychological theory: implications for personality, social, environmental, and developmental studies.

[19] Alessandro B. Melchiorre and Markus Schedl. 2020. Personality correlates of music audio preferences for modelling music listeners. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2020, Genoa, Italy, July 12-18, 2020*. Tsvi Kuflik, Ilaria Torre, Robin Burke, and Cristina Gena, (Eds.) ACM, 313–317. DOI: 10.1145/3340631.3394874.

[20] Sergio Oramas, Oriol Nieto, Mohamed Sordo, and Xavier Serra. 2017. A deep multimodal approach for cold-start music recommendation. In *Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems, DLRS@RecSys 2017, Como, Italy, August 27, 2017*. Balázs Hidasi, Alexandros Karatzoglou, Oren Sar Shalom, Sander Dieleman, Bracha Shapira, and Domonkos Tikk, (Eds.) ACM, 32–37. DOI: 10.1145/3125486.3125492.

[21] Nurul Aida Osman, Shahrul Azman Mohd Noah, Mohammad Darwich, and Masnizah Mohd. 2021. Integrating contextual sentiment analysis in collaborative recommender systems. *Plos one*, 16, 3, e0248695.

[22] Junwei Pan, Jian Xu, Alfonso Lobos Ruiz, Wenliang Zhao, Shengjun Pan, Yu Sun, and Quan Lu. 2018. Field-weighted factorization machines for click-through rate prediction in display advertising. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018*. Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis, (Eds.) ACM, 1349–1357. DOI: 10.1145/3178876.3186040.

[23] Marco Polignano, Fedelucio Narducci, Marco de Gemmis, and Giovanni Semeraro. 2021. Towards emotion-aware recommender systems: an affective coherence model based on emotion-driven behaviors. *Expert Syst. Appl.*, 170, 114382. DOI: 10.1016/J.ESWA.2020.114382.

[24] Yanru Qu, Han Cai, Kan Ren, Weinan Zhang, Yong Yu, Ying Wen, and Jun Wang. 2016. Product-based neural networks for user response prediction. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*. (Dec. 2016), 1149–1154. DOI: 10.1109/ICDM.2016.0151.

[25] Steffen Rendle. 2010. Factorization machines. In *ICDM 2010, The 10th IEEE International Conference on Data Mining, Sydney, Australia, 14-17 December 2010*. Geoffrey I. Webb, Bing Liu, Chengqi Zhang, Dimitrios Gunopulos, and Xindong Wu, (Eds.) IEEE Computer Society, 995–1000. DOI: 10.1109/ICDM.2010.127.

[26] Steffen Rendle and Lars Schmidt-Thieme. 2010. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the Third International Conference on Web Search and Web Data Mining, WSDM 2010, New York, NY, USA, February 4-6, 2010*. Brian D. Davison, Torsten Suel, Nick Craswell, and Bing Liu, (Eds.) ACM, 81–90. DOI: 10.1145/1718487.1718498.

[27] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2022. Recommender systems: techniques, applications, and challenges. In *Recommender Systems Handbook*. Francesco Ricci, Lior Rokach, and Bracha Shapira, (Eds.) Springer US, 1–35. DOI: 10.1007/978-1-0716-2197-4_1.

[28] David T. Rubin and Jennifer M. Talarico. 2009. A comparison of dimensional models of emotion: Evidence from emotions, prototypical events, autobiographical memories, and words. *Memory*, 17, 8, (Nov. 2009), 802–808. DOI: 10.1080/09658210903130764.

[29] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39, 6, 1161.

[30] Thomas Schäfer, Peter Sedlmeier, Christine Städtler, and David Huron. 2013. The psychological functions of music listening. *Frontiers in psychology*, 4, 511.

[31] Markus Schedl, Peter Knees, Brian McFee, and Dmitry Bogdanov. 2022. Music recommendation systems: techniques, use cases, and challenges. In *Recommender Systems Handbook*. Francesco Ricci, Lior Rokach, and Bracha Shapira, (Eds.) Springer US, 927–971. DOI: 10.1007/978-1-0716-2197-4_24.

[32] Safa Selmene and Zahra Kodia. 2020. Recommender system based on user's tweets sentiment analysis. In *ICEEG 2020: The 4th International Conference on E-commerce, E-Business and E-Government, Arenthon, France, June, 2020*. ACM, 96–102. DOI: 10.1145/3409929.3414744.

[33] Camila Vaccari Sundermann, Marcos Aurélio Domingues, Roberta Akemi Sinoara, Ricardo Marcondes Marcacini, and Solange Oliveira Rezende. 2019. Using opinion mining in context-aware recommender systems: a systematic review. *Information*, 10, 2, 42.

[34] Daniel Valcarce, Alejandro Bellogín, Javier Parapar, and Pablo Castells. 2020. Assessing ranking metrics in top-n recommendation. *Inf. Retr. J.*, 23, 4, 411–448. DOI: 10.1007/S10791-020-09377-X.

[35] Ruoxi Wang, Bin Fu, Gang Fu, and Mingliang Wang. 2017. Deep & cross network for ad click predictions. In *Proceedings of the ADKDD'17, Halifax, NS, Canada, August 13 - 17, 2017*. ACM, 12:1–12:7. DOI: 10.1145/3124749.3124754.

[36] Ruoxi Wang, Rakesh Shivanna, Derek Cheng, Sagar Jain, Dong Lin, Lichan Hong, and Ed Chi. 2021. Dcn v2: improved deep & cross network and practical lessons for web-scale learning to rank systems. In *Proceedings of the Web Conference 2021* (WWW '21). Association for Computing Machinery, Ljubljana, Slovenia, 1785–1797. ISBN: 9781450383127. DOI: 10.1145/3442381.3450078.

[37] Songhao Wu. 2021. design your own sentiment score. *Towards Data Science*, (May 2021). https://towardsdatascience.com/design-your-own-sentiment-score-e524308cf787.

[38] Jun Xiao, Hao Ye, Xiangnan He, Hanwang Zhang, Fei Wu, and Tat-Seng Chua. 2017. Attentional factorization machines: learning the weight of feature interactions via attention networks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*. Carles Sierra, (Ed.) ijcai.org, 3119–3125. DOI: 10.24963/IJCAI.2017/435.

[39] Marcel Zentner, Didier Grandjean, and Klaus R Scherer. 2008. Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, 8, 4, 494.

[40] Wayne Xin Zhao et al. 2021. Recbole: towards a unified, comprehensive and efficient framework for recommendation algorithms. In *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021*. Gianluca Demartini, Guido Zuccon, J. Shane Culpepper, Zi Huang, and Hanghang Tong, (Eds.) ACM, 4653–4664. DOI: 10.1145/3459637.3482016.

[41] Doris Zhou. 2017. SentimentAnalysis. https://github.com/dwzhou/SentimentAnalysis. (2017).