Context-aware Board Game Recommendations

JOSÉ L. JORRO-ARAGONESES, Universidad Autónoma de Madrid, Spain
IVÁN CANTADOR, Universidad Autónoma de Madrid, Spain
ALEJANDRO BELLOGÍN, Universidad Autónoma de Madrid, Spain

The board game market has experienced rapid growth, driven by the increasing popularity of family and social gatherings and enhanced quality and diversity of games. With a vast array of games catering to various player types and gaming contexts, selecting the right game can be challenging. Factors such as players demographics, playtime, and preferred game mechanics must be considered. Existing online platforms like BoardGameGeek.com and BoardGaming.com offer extensive game information, but lack advanced search and recommendation functionalities. In this paper, we address this gap by presenting preliminary advancements in context-aware recommender systems for board games. Our contributions include defining a game playing context model, building a contextualized ratings database, testing diverse recommendation methods, and proposing context-based evaluation methodology and metrics.

CCS Concepts: • **Information systems → Recommender systems**; • **Human-centered computing → Contextual design**.

Additional Key Words and Phrases: context-aware recommender systems, context modeling, board games, tabletop games

1 INTRODUCTION

In recent years, the board game[1] market has experienced remarkable growth and expansion globally, with a business volume reaching €4.02 billion in 2024 and a compound annual growth rate of 9.19%[2]. This surge can be attributed to several factors, including the enhanced quality and diversity of available games that captivate a wide range of audiences, and the increasing popularity of family and social gatherings. The COVID-19 pandemic accelerated this trend, as lockdowns and social distancing measures led people to seek engaging and interactive indoor activities, making board games become a popular choice for many households, fostering social interaction and entertainment in a time of isolation.

The variety of board games on the market today is huge, catering to different types of players and gaming contexts. For example, "Settlers of Catan" is a strategic game that involves resource management and trading, appealing to players who enjoy tactical decision-making and competitive play. It is particularly suited for groups of friends or relatives looking for an engaging and relatively easy-going gaming experience. On the other hand, "Spirit Island" is a cooperative game where players work together as powerful spirits to defend their island from colonizing invaders. This game is ideal for players who prefer teamwork and enjoy intricate, strategic gameplay, making it perfect for family game nights or collaborative sessions among experienced gamers.

Purchasing a new board game can be a daunting task, not only because of the very large and rapidly increasing number of games, but also due to the significant investment they often represent, with an average game cost of €50-60. Moreover, the selection process must consider various factors, such as the minimum *age* and *number of players*, the available *playing time*, the preferred *game types* and *mechanics*, and the *playing complexity level*. These considerations are crucial to ensure that the chosen game aligns well with the players' preferences and the context in which the game will be played.

To assist in the above selection process, several online platforms provide extensive information on board games. Websites like BoardGameGeek[3] (BGG) and BoardGaming[4] have emerged as essential resources for board game enthusiasts.

[1]Board games are a particular type of tabletop games that involve players interacting with each other and game components on a flat surface. However, both terms are often used interchangeably. In this paper, we use the term "board games" to refer to all types of tabletop games.
[2]Statista: Board Games Market Revenue, https://www.statista.com/outlook/amo/app/games/board-games/worldwide
[3]BoardGameGeek, BGG, https://boardgamegeek.com
[4]BoardGaming, https://boardgaming.com

These platforms provide vast public databases with detailed information about thousands of board games, including game metadata and user reviews and ratings. Additionally, they allow users to create (favorite) game lists, record played games, and participate in an active community that shares opinions and suggestions about games. These platforms, however, have limited information retrieval and filtering functionalities, making the task of finding an ideal game challenging for a user, who may be overwhelmed by the number of games and the overload of content.

There is thus a clear necessity to develop advanced recommender systems tailored to the particular domain of board games. Besides, such systems should ideally be context-aware, meaning that they can account for the specific circumstances associated with distinct playing scenarios. Differently to other domains, in board games, an item (game) is usually consumed (played) in a single context or a limited number of "similar" contexts. Hence, contextual factors such as *who will play* (one person, a couple, young children, a group of friends), *the time available to play* (from quick games of 15-30 minutes to games that can last several hours or even days), and *the players' gaming mood* (e.g., current preferences for easy-to-play games, party games, collaborative games, or strategic competitive games with little room for luck) are crucial for effective recommendations.

Despite their potential benefits, board game recommender systems have barely been explored in the academic literature. Proposed recommendation methods have been quite simple, have not been evaluated with large datasets, and have rarely exploited contextual information about the gaming experience. Addressing these research gaps, in this paper we present initial advancements in defining context models for board games, creating large databases with contextualized ratings, testing diverse context-aware recommendation methods, and proposing context-based evaluation methodology and metrics.

2 RELATED WORK

Although the board game industry has a significant and growing market size and impact, research on board game recommender systems is scarce and preliminary. We next survey published papers, highlighting their contributions and limitations.

Ng and Seaman [7] presented PeGRec, a CombMNZ model that aggregates several rankings to provide recommendations for a given game. These rankings are generated by heuristic strategies that consider diverse game data from BGG, such as *topics* (e.g., fantasy, old west, pirates) and categories (e.g., abstract, cooperative, wargame), *number of players*, *playtime*, *playing complexity*, and *game popularity*. Through a user study in Mechanical Turk, 100 participants evaluated top-3 recommendation lists for a predefined set of 400 games, showing that the proposed model provided better recommendations than Amazon and Barners & Noble baselines, in terms of precision@n, MRR and nDCG.

Zalewski, Ganzha and Paprzycki [10] proposed a user-based kNN collaborative filtering method applied on clusters of games precomputed with BGG data. The authors analyzed pairs of numeric attributes of games, showing high correlations between *average rating* and *rank*, and *playing complexity*, *playing time*, and *minimum age* and *number of players*. Considering non-correlated attributes, and applying the elbow criterion, they used the K-means algorithm to group a set of 3K games (with at least 300 user reviews) into 6 clusters. The authors applied a form of leave-one-out evaluation for 200 users, achieving average HitRatio@10 values around 64%. The obtained clusters were related to relatively short, easy-to-play games, mass-market games, party games, complex games, family-friendly games, and abstract games with a small number of players –two, in most cases.

Ion, Sacharidis and Werthner [4] reported an offline evaluation on a BGG dataset comparing several collaborative filtering (user- and item-based heuristics, matrix factorization, and autoencoders), content-based (Euclidean kNN, and IDF vector space model) methods, as well as a hybrid recommender consisting of an autoencoder fed with both rating and game attribute data. The kNN method exploited game *categories* and *mechanics* (e.g., dice rolling, card drafting, set collection), *playing complexity*, *playing time*, and *minimum age* and *number of players*, whereas the IDF method was restricted to discrete attributes: game *categories* and *mechanics*. Applying an 80-20 training-test split on users with at least 200 ratings, the authors showed that the autoencoder approaches achieved the highest performance in terms of precision@n, recall@n, and nDCG@n (for n = 5, 10, 20, 100), and according to game category- and mechanic-based diversity and novelty metrics. The authors claimed that their dataset originally contained 13M ratings from about 250K users for 80.5K games, but they did not describe the characteristics of the final dataset used.

Finally, Kim et al. [6] evaluated sequential deep learning-based recommendation methods on a rating dataset from BGG. The entire dataset –which is not publicly available at the time of writing– comprised 47.3M ratings assigned by 388.4K users to 87.2% games, entailing a sparsity of 99.86%, higher than well-known rating datasets, such as MovieLens 1M (94.57%). The conducted experiments considered cold-start and non cold-start users, with at least 8 and 200 ratings, respectively. Following an 80-10-20% data split for training, validation and testing, the authors showed that a CNN model outperformed an RNN model, in terms of precision@10, recall@10, nDCG@10, MAP and MRR.

Overall, these studies contributed valuable insights into the development of board game recommender systems, offering diverse strategies to provide personalized recommendations exploiting BGG data. However, their experiments lacked detailed explanations for reproducibility. Moreover, none of them considered contextual factors to tailor recommendations for specific playing scenarios. Contextual factors such as the composition of the players group, the available playtime, and the players' mood (short-term preferences) for certain game types and mechanics are crucial for making effective recommendations. Addressing these gaps, our work focuses on context-aware board game recommender systems, proposing a formal context model, using context-aware evaluation methodology and metrics, and preliminary testing state-of-the-art context-aware recommendation methods.

3 CONTEXTUALIZED RATINGS DATABASE

In this section, we describe the building process and characteristics of the database used in this work. The database was built by crossing public contents from two online websites about board games: BoardGameGeek and Zacatrus[5].

3.1 BoardGameGeek

As previously done in the literature, we built a database using the publicly available contents of the online platform BoardGameGeek. BGG is a well-known repository that hosts a collection of more than 125K board games. It provides a wide variety of information about each game, including the title in different languages, a description, game categories[6], types and families[7], gaming mechanics[8], average playing time, minimum and maximum number of players, and recommended minimum players' age. Additionally, BGG offers user-generated content about the games, including ratings, comments, and playing complexity scores.

Certain game attributes and metadata on BGG are implicitly related to gaming contexts. For example, the attribute "Playing time" indicates the approximate time needed to conclude a game (time context); the mechanic "Cooperative game" corresponds to games suitable for a relaxed, non-competitive environment (gaming mood context); and the category "Children's Games" and type "Party Games" respectively refer to games suitable for playing with children and playing at parties with a relatively large number of people (social companion context). Despite these implicit context signals, in BGG, there is no explicit model or annotations for gaming contexts.

[5]Zacatrus, https://zacatrus.es
[6]BGG game categories, https://boardgamegeek.com/browse/boardgamecategory
[7]BGG game families, https://boardgamegeek.com/browse/boardgamefamily
[8]BGG game mechanics, https://boardgamegeek.com/browse/boardgamemechanic

We started building our database by crawling and scraping the public profiles of all the BGG users from the USA. From these profiles, we obtained their rated (and sometimes reviewed) games. Then, we crawled and scraped the BGG web pages associated to each game, obtaining the game metadata introduced above. At the end of the process, and after discarding the games without categories and types, and the users and games with less than 10 ratings, our database comprised more than 5.2M ratings provided by 67K users to 17.5K games, resulting in a rating sparsity of 99.55%. From this original database, we created a reduced version limited to those items to which we could assign a particular gaming context, as explained in the next subsection.

3.2 Zacatrus

To establish a playing context model for board games, and based on such a model, assign a particular context to a game, we utilized information from Zacatrus, one of the most popular board game stores in Spain.

The Zacatrus website hosts a collection of more than 17.5K games. Similarly to BGG, it provides metadata for each game sold. Notably, it also offers a series of labels for a game, presumably assigned manually. According to some of the available labels, the website provides a simple board game recommender, which consists of a 3-stage filtering process. In each stage, the user selects one criterion from a few options to define characteristics of the game she is interested in. The stages correspond to the "game playing style" (e.g., family, party, travel, quick, cooperative, eurogame[9] and ameritrash[10] games), the "game mechanics" (e.g., card drafting, set collection, exploration and adventure, deduction and investigation, etc.), and the "game themes" (e.g., history, science fiction, pirates, animals, urban, oriental, etc.).

As mentioned before, an interesting fact is that some of these options (labels) are related to gaming contexts. In particular, we consider three different context dimensions:

- *Playing time*, which can distinguishes between quick and long games. In this case, options like "Rápido" (fast) and "Viaje" (travel) are associated with quick games.
- *Gaming mood*, which can differentiate games annotated with "Fiesta" (party), "Cooperativo" (cooperative), "Eurogame" (strategic), "Ameritrash" (thematic), "Narrativo" (narrative or story-based), "Experto" (expert).
- *Social companion*, which can entail games best played by one person ("Solitario"), two people ("Para 2"), children ("Infantil"), and family relatives ("Familiar").

We decided to use these manual annotations to isolate as much as possible the context assigned to a board game from its attributes and metadata in BGG, which will likely be exploited by the recommendation methods. It is clear that some of the BGG game data are related to some of the contexts defined. However, we leave as future work the thorough study on how to map contexts to games according to their attributes and metadata.

With all the above, we crawled and scrapped the public content about board games available in the Zacatrus website. In some cases, the Zacatrus web page of a game included a hyperlink to the game's web page in BGG. This hyperlink contained the numeric identifier of the game in BGG. Thus, we could directly add the Zacatrus labels (and consequently, contexts) to games recorded in our BGG database (Subsection 3.1). For those cases without hyperlinks to BGG, we attempted to map the corresponding Zacatrus games by exact matching of their titles with those of BGG games. At the end, we were able to build a final database with 1,452,768 contextualized ratings from 43,660 users for 901 games, resulting in a rating sparsity of 96.31%. In the database, every user and game has at least 10 ratings, and every game has at least

[9]Eurogames, also known as German-style board games, are characterized by strategic depth, minimal luck, and player interaction that often avoids direct conflict. These games typically focus on resource management, economic themes, and intricate mechanics, with a strong emphasis on strategy and long-term planning.

[10]Ameritrash games are known for their strong themes, immersive narratives, and high levels of player interaction, often involving direct conflict and luck. They usually feature detailed components, such as miniatures and elaborate boards, and emphasize thematic storytelling and gameplay experiences.

one category, one type, and one mechanic; on average: 2.91 categories, 1.21 types, and 4.3 mechanics per game. All games have textual descriptions, and metadata (e.g., the playing complexity score in a [1,5] scale, minimum player age, minimum and maximum numbers of players, average playtime in minutes).

In the next section, we detail the considered board game playing contexts, describing the underlying context model and representation.

4 BOARD GAME PLAYING CONTEXTS

Differently to other domains, such as movie and music, where an item is usually consumed under diverse contextual conditions [1], in the board games domain, we can assume that an item (i.e., a board game) is appropriate for a particular context or a small set of "similar" contexts. For instance, there are quick, easy-to-play games for children, and long, strategic games to be played by several (more than two) buddies.

For this reason, we herein advocate for assigning a representative context for a game, and defining a context representation that allow computing similarities between contexts, so similar contexts may be taken into consideration to generate (and even evaluate) recommendations. Moreover, we propose to explore alternative techniques to set the games' representative contexts, from strict to flexible ones, partially using game attributes and metadata. We dig into these aspects in the next subsections.

4.1 Context Model and Representation

In [5], we presented a conceptual framework for designing and evaluating context-aware recommender systems (CARS). The framework includes a principal component associated to the modeling of the context underlying any CARS. Such a component entails the specification of a formal model to represent the context, distinguishing between context dimensions and factors.

Instantiating the framework for board games, Table 1 shows the elements of the proposed context model. We establish that a game playing context is defined in terms of three dimensions: playing time (c_t), gaming mood (c_m), and social companion (c_s). Each of these dimensions has several factors that can be associated to a given context and are described in different vector representations, depending on whether the factors of a dimension maintain or not certain order among them.

For c_t, a game can be categorized as quick, short, moderate, long, or very long. Quick and short games were those labeled as 'fast' in Zacatrus. Quick games have an average playtime around 15 minutes, and short games, around 30 minutes. Moderate games take less than 2 hours, and long games take between 2 and 3 hours. These factors follow an ordinal, 10-dimensional vector representation that entails that quick and short games are more similar to each other than to moderate/long and very long games.

For c_m, we consider the following contexts: parties in which funny games are appropriate, situations where easy-to-play games are convenient, environments for collaborative games, times for story-based or thematic games, and special occasions in which playing strategic and expert (i.e., complex) games. These factors follow a one-hot, 7-dimensional vector representation that entails null similarity between any two distinct factors.

Finally, for c_s, we first distinguish games appropriate for one or two players. Then, we consider games for toddlers (2-3 years old), preschoolers (4-5 years old), or children (6-12 years old). Other games are suitable for family relatives (likely with diverse ages), and the remaining games are assumed to be played by friends. In this case, the representation is a 7-dimensional vector with one-hot-encoding for all factors, except toddlers, preschoolers, and children, who may enjoy the same games to some extent.

Context dimension	Context factor	Vector representation										Representation type	Similarity
	quick	1											
	short	1	1										
Playing time (c_t)	moderate	1	1	1	1							ordinal	Normalized Euclidean
	long	1	1	1	1	1	1						
	very_long	1	1	1	1	1	1	1	1	1	1		
	party	1											
	easy-to-play		1										
	cooperative			1									
Gaming mood (c_m)	story-based				1							one-hot	Cosine
	thematic					1							
	strategic						1						
	expert							1					
	1-player	1											
	2-players		1										
	toddlers			1									
Social companion (c_s)	preschoolers			1	1							one-hot & ordinal	Cosine
	children			1	1	1							
	family						1						
	friends							1					

Table 1. Proposed model for board game playing contexts, using binary variables for the vectors of the context factors.

With all the above, in our model, a particular context is represented as a 19-dimensional (binary or non-binary) numeric vector, resultant of the concatenation of the three components c_t, c_m, and c_s. The similarity between two contexts takes these three components into account, as explained next.

4.2 Context Similarity

Recent research efforts have been made to formally model and describe board games, their elements, attributes and relationships, and to establish similarity metrics between board games [9]. In alignment with this trend, we define similarity metrics between gaming contexts.

Specifically, we consider that the similarity between two contexts $c^{(x)}$ and $c^{(y)}$ can be expressed in a general form as a weighted sum of the similarities between their contextual dimension components $c_k^{(x)}$ and $c_k^{(y)}$:

$$sim(c^{(x)}, c^{(y)}) = \sum_k \lambda_k \cdot sim(c_k^{(x)}, c_k^{(y)}) \tag{1}$$

where $\lambda_k \in [0,1]$ and $\sum_k \lambda_k = 1$. In our gaming context model, the $k=3$ components correspond to the playing time (c_t), gaming mood (c_m), and social companion (c_s) dimensions. For our experiments, we set $\lambda_k = 1/3$.

Then, as commented in the previous subsection, and shown in Table 1, assuming binary vector representations of context factors, we use different similarity functions for the three contextual components: the normalized Euclidean distance for the ordinal vector representation of c_t, and the cosine similarity for the one-hot vector representation of c_m and c_s. We empirically observed that these metrics return coherent similarity values between pairs of factors in each gaming context dimension.

Another possibility is to use non-binary, numeric vectors, which are valuable in cases where a particular context can be assigned weights for the different context factors; for example, a board game that can be played in short or moderate time periods (playing time context factors), is easy-to-play, cooperative, and story-based (gaming mood context factors), and results appropriate for either children, family relatives, or friends (social companion context factors). For these vectors, we could use any well-known similarity metric, but will focus on the cosine similarity.

In our experiments, we evaluated recommendation methods built upon both binary and non-binary context vectors, and using the above similarity metrics.

5 EXPERIMENTS

In this section, we report results achieved in initial experiments aimed to preliminarily evaluate the performance of some state-of-the-art context-aware recommendation methods on our board game database. The results, which are given in terms of both non-contextual and contextual metrics, were obtained by applying a 5-fold cross-validation strategy in which, for each user, 80% of her ratings were used for training, and the remaining 20% for testing. Besides, for some recommendation methods, 20% of the training ratings per user were used for validation.

5.1 Evaluation Methodology

Following the offline evaluation protocols established in the community [2], we focused our experiments on generating a ranking of items for a given user, which are later used to compute the performance of the recommendation method that produced such ranking. This procedure has to be extended for CARS in general, and for our work in particular, by including the user's current context as input.

Specifically, for each user u, to capture the context to be considered when querying a recommender, we checked every item belonging to the test set of u, and used the context assigned to that item as the current context c. Hence, for each input (u,c) tuple, a ranking $R(u,c)$ was produced by a recommendation method, and the evaluation metrics described in the next section were computed.

Note that this context-based evaluation procedure entails a much greater rating sparsity, and consequently lower recommendation performance values. Moreover, it makes the empirical comparison with non-context-aware recommenders tricky and unfair, since these systems generate item suggestions regardless of the target context.

5.2 Evaluation Metrics

In the conducted experiments, we computed standard ranking-based (and non-context-based) metrics to evaluate the performance of the considered recommendation methods. Specifically, in accordance to the related work described in Section 2, we measured *precision@n*, *recall@n*, *F1@n* and *nDCG@n*, for cutoff values n = 5, 10, 15, 20, 25.

Moreover, as a novel contribution of our work, we also propose and compute a context-based metric that we call *context satisfaction*. Specifically, considering a context formed by a set of binary variables associated to context factors (as exemplified in our model of Subsection 4.1), we define the satisfaction (fulfillment) of a target context c by an item i as the completion of the set of variables of c by the variables of context $c(i)$ associated to i. Hence, to measure context satisfaction, we consider the following set completion metric based on the Jaccard coefficient J:

$$sat_\alpha(c,i) = sat_\alpha(c,c(i)) = \frac{|c \cap c(i)|}{|c \cup c(i)| + \alpha \cdot \frac{|c \setminus c(i)|}{|c|}} \qquad (2)$$

where the second factor of the denominator penalizes the unfulfillment of the target context c by the item's context $c(i)$ variables. The coefficient $\alpha \in [0,1]$ modulates this penalization, such that if $\alpha = 0$, then $sat_\alpha(c,i) = J(c,c(i))$. In the experiments, we set $\alpha = 0.5$.

In the case of non-binary context vectors, the sat_α metric has the same formulation, but the set operations (i.e., intersection, union and difference) are computed by counting the exact matches between the values of the c and $c(i)$

vector components. For instance, if $c = \{0.1,0.2,0.3\}$ and $c(i) = \{0.1,0.3,0.5\}$, then $|c \cap c(i)| = 1$ since the set vectors have the same value 0.1 in their first component, and different values in their second and third components.

Then, we define the satisfaction of c by a ranking R of items (recommended for user u and context c) as follows:

$$sat(c,R) = \frac{1}{|R|}\sum_{i \in R} sat_{\alpha}(c,i) \tag{3}$$

The sat_{α} metric is very strict, since it is based on exact matches between the components of the context vectors. To make the metric tolerant to the partial satisfaction of a context, using binary or non-binary context vectors, we propose an alternative formulation, named as sat_{rel}, which considers relatedness degrees between context factors:

$$sat_{rel}(c,i) = sat_{rel}(c,c(i)) = \frac{1}{S}\sum_{x,y} c_x \cdot c_y(i) \cdot rel(x,y) \tag{4}$$

where x and y refer to context factors (i.e., the 19 game playing factors), c_x is the numeric value of factor x in context c, $rel(x,y)$ is the relatedness degree between factors x and y, and $S = \sum_{x,y} c_x \cdot c_y(i)$ is a normalization term. In our experiments, we set the $rel(x,y)$ values as the similarities between the factor vector representations given in Table 1.

5.3 Recommendation Methods

To preliminary experiment with our board game database, upon the RecBole library[11] [11], we built and evaluated the following context-aware recommendation methods:

- **ContextRnd**, which prefilters the items that satisfy the target context, and sort them randomly.
- **ContextPop**, which prefilters the items that satisfy the target context, and rank them according to their rating popularity.
- **Factorization Machine** (FM), which combines the advantages of Support Vector Machines and factorization models [8].
- **Deep Learning-based FM** (DeepFM) [3], which builds a factorization machine through a deep neural network architecture.

The two prefiltering methods are expected to perform very well considering that the followed evaluation methodology takes the contexts of the test items as target contexts, and the number of items belonging to a particular context is small for the proposed 19-dimensional vector representation. The FM methods were chosen since they are simple and flexible to integrate user, item and interaction (e.g., contextual) data as attributes from which obtaining latent factors. For *FM* and *DeepFM*, we used the default parameter configurations of RecBole[12].

5.4 Empirical Results

Table 2 shows the performance results achieved by the tested recommenders on the raw contexts of the games, that is, on their 19-dimensional contextual vectors. As expected, the prefiltering *ContextRnd* and *ContextPop* methods outperform the FM methods for all metrics. They filter out the games that do not belong to the target test contexts, and rank the remainder games, randomly or by popularity. Thus, they are able to fully satisfy the target context (i.e., achieving 100% of sat_{α} and sat_{rel}), regardless of the cutoff. Also, since they generate shorter recommendation lists, the chances are higher to include the relevant games of the test sets. Hence, their ranking-based metrics (precision, recall, F1 and nDCG) values are higher.

[11] We extended the RecBole implementations to support (user, context) pairs.

[12] *FM*: 300 epochs and early stopping, training and test batch sizes of 2048 and 4096, learning rate 0.001, and embedding size (number of latent factors) of 10. *DeepFM*: 500 epochs and early stopping, training and test batch sizes of 512 and 1024, learning rate 0.001, and AUC as validation metric.

	precision					recall					F1				
	5	**10**	**15**	**20**	**25**	**5**	**10**	**15**	**20**	**25**	**5**	**10**	**15**	**20**	**25**
ContextRnd	0.050	0.045	0.042	0.039	0.034	0.052	0.088	0.117	0.139	0.152	0.045	0.053	0.056	0.055	0.052
ContextPop	0.063	0.052	0.046	0.041	0.036	0.070	0.104	0.129	0.150	0.162	0.058	0.061	0.061	0.059	0.055
FM	0.002	0.004	0.005	0.006	0.006	0.001	0.004	0.007	0.013	0.017	0.001	0.004	0.005	0.008	0.009
DeepFM	0.002	0.004	0.006	0.007	0.007	0.001	0.004	0.010	0.014	0.018	0.001	0.003	0.007	0.008	0.009
	nDCG					**sat$_{\alpha}$**					**sat$_{rel}$**				
	5	**10**	**15**	**20**	**25**	**5**	**10**	**15**	**20**	**25**	**5**	**10**	**15**	**20**	**25**
ContextRnd	0.063	0.075	0.086	0.095	0.099	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ContextPop	0.086	0.095	0.105	0.113	0.118	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
FM	0.002	0.003	0.005	0.007	0.009	0.165	0.170	0.180	0.188	0.192	0.371	0.382	0.411	0.427	0.436
DeepFM	0.002	0.003	0.006	0.008	0.010	0.162	0.169	0.181	0.187	0.186	0.365	0.374	0.396	0.411	0.415

Table 2. Results achieved by the recommendation methods on the raw contexts of the games.

	precision					recall					F1				
	5	**10**	**15**	**20**	**25**	**5**	**10**	**15**	**20**	**25**	**5**	**10**	**15**	**20**	**25**
ContextRnd	0.020	0.020	0.020	0.020	0.020	0.016	0.033	0.050	0.068	0.086	0.016	0.022	0.026	0.028	0.030
ContextPop	0.064	0.055	0.049	0.044	0.039	0.071	0.109	0.138	0.158	0.172	0.060	0.065	0.065	0.063	0.059
FM	0.014	0.015	0.016	0.017	0.018	0.009	0.020	0.033	0.047	0.062	0.010	0.016	0.020	0.023	0.025
DeepFM	0.002	0.004	0.005	0.007	0.008	0.001	0.004	0.009	0.017	0.022	0.001	0.003	0.006	0.009	0.011
	nDCG					**sat$_{\alpha}$**					**sat$_{rel}$**				
	5	**10**	**15**	**20**	**25**	**5**	**10**	**15**	**20**	**25**	**5**	**10**	**15**	**20**	**25**
ContextRnd	0.022	0.027	0.034	0.040	0.047	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ContextPop	0.082	0.095	0.106	0.114	0.120	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
FM	0.014	0.017	0.023	0.028	0.034	0.670	0.609	0.575	0.551	0.530	0.655	0.628	0.612	0.601	0.592
DeepFM	0.002	0.003	0.006	0.009	0.012	0.083	0.085	0.084	0.084	0.083	0.389	0.373	0.366	0.364	0.364

Table 3. Results achieved by the recommendation methods on the cluster-based contexts of the games.

ContextPop performs slightly better than *ContextRnd*, evidencing the potential importance of considering the rating popularity (bias) in BGG, as done in the literature (see Section 2). On the other hand, according to our experiments, there are no significant differences in the performance of *FM* and *DeepFM*. This may be due to the lack of enough data for each context, or the need for more evaluations to find better parameter values, among other reasons.

In any case, according to the reported ranking-based metric values, there is room for improvement, and open research lines in the exploration of more sophisticated personalized recommendations methods and other definitions and representations of game playing contexts. In particular, we built and evaluated the methods in a second experiment with a limited number of contexts. We applied the K-means clustering algorithm on the raw 19-dimensional contextual vectors of the games in our database, considered the centroids of the obtained clusters as representative contexts, and assigned to each game the centroid closest to its raw contextual vector. The number of clusters was 13, which was automatically set by applying the elbow criterion technique. Evaluating other numbers of clusters is left as future work.

Table 3 shows the performance results achieved by the tested recommenders on the cluster-based contexts of the games. Obviously, the context satisfaction values achieved by the prefiltering *ContextRnd* and *ContextPop* methods remain at 1.0. By contrast, their ranking-based performance decreases. Since the number of clusters is much smaller than in the first experiment, the number of items belonging to each context is much larger, and consequently the accuracy of random and rating popularity-based recommendations is lower.

On the other hand, as one may expect, with a few contexts, the performance of *FM* and *DeepFM* increases in terms of both ranking-based and context satisfaction metrics. This increment is more evident for the *FM* method. Nonetheless, note that we did not conduct any parameter optimization, so better results for these two methods could be achieved. Moreover, other number of contexts (clusters) may entail further improvements.

6 CONCLUSIONS

In this paper, we have motivated the potential of recommender systems in general, and context-aware recommenders in particular, in the domain of board games. As seminal work on context-aware board game recommendations, we have proposed a multidimensional vector model of game playing contexts, have built a contextualized rating database with data from the well-known BGG platform, and have proposed novel context-based evaluation methodology and metrics. Moreover, through preliminary experiments, we have shown that simple prefiltering recommendation methods are able to achieve positive performance results in terms of both ranking-based quality and context satisfaction.

Nonetheless, there are much room for improvement and need for evaluating in depth more sophisticated personalized and context-aware recommendation approaches [12]. Besides, we believe that many research lines and opportunities arise from the presented ongoing work. Exploring alternative notions and representations of contexts, and experimenting with additional context-based metrics, are two major research issues that are worth to be further explored and could be considered in other recommendation domains.

ACKNOWLEDGMENTS

This work was supported by Grant PID2022-139131NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF, a way of making Europe."

REFERENCES

[1] Gediminas Adomavicius, Konstantin Bauman, Alexander Tuzhilin, and Moshe Unger. 2022. Context-aware Recommender Systems: From Foundations to Recent Developments. In *Recommender Systems Handbook*. Springer, 211–250. https://doi.org/10.1007/978-1-0716-2197-4_6

[2] Asela Gunawardana, Guy Shani, and Sivan Yogev. 2022. Evaluating Recommender Systems. In *Recommender Systems Handbook*. Springer, 547–601. https://doi.org/10.1007/978-1-0716-2197-4_15

[3] 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 26th International Joint Conference on Artificial Intelligence*. IJCAI, 1725–1731. https://doi.org/10.24963/IJCAI.2017/239

[4] Michael Ion, Dimitris Sacharidis, and Hannes Werthner. 2020. Designing a Recommender System for Board Games. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing*. ACM, 1465–1467. https://doi.org/10.1145/3341105.3375780

[5] Jose L Jorro-Aragoneses, Iván Cantador, and Alejandro Bellogín. 2024. On the Formalization of the Context-aware Recommender Systems Design, Building and Evaluation Processes. In *Proceedings of the 1st Spanish Conference on Recommender Systems*. Proceedings of the 20th Conference of the Spanish Association for Artificial Intelligence, ISBN 978-84-09-62724-0, 617–618.

[6] JaeWon Kim, JeongA Wi, SooJin Jang, and YoungBin Kim. 2020. Sequential Recommendations on Board-game Platforms. *Symmetry* 12, 2 (2020), 210:1–210:15. https://doi.org/10.3390/sym12020210

[7] Yiu-Kai Ng and Iris Seaman. 2017. Personalized Table-top Game Recommendations. In *Proceedings of the 29th International Conference on Tools with Artificial Intelligence*. IEEE, 396–403. https://doi.org/10.1109/ICTAI.2017.00068

[8] Steffen Rendle. 2010. Factorization Machines. In *Proceedings of the 10th IEEE International Conference on Data Mining*. IEEE, 995–1000. https://doi.org/10.1109/ICDM.2010.127

[9] Matthew Stephenson, Dennis JNJ Soemers, Éric Piette, and Cameron Browne. 2022. Measuring Board Game Distance. In *Proceedings of the International Conference on Computers and Games*. Springer, 121–130. https://doi.org/10.1007/978-3-031-34017-8_11

[10] Jan Zalewski, Maria Ganzha, and Marcin Paprzycki. 2019. Recommender System for Board Games. In *Proceedings of the 23rd International Conference on System Theory, Control and Computing*. IEEE, 249–254. https://doi.org/10.1109/ICSTCC.2019.8885455

[11] Wayne Xin Zhao, Yupeng Hou, Xingyu Pan, Chen Yang, Zeyu Zhang, Zihan Lin, Jingsen Zhang, Shuqing Bian, Jiakai Tang, Wenqi Sun, et al. 2022. RecBole 2.0: Towards a More Up-to-Date Recommendation Library. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. ACM, 4722–4726. https://doi.org/10.1145/3511808.3557680

[12] Yong Zheng. 2022. Context-aware Collaborative Filtering using Context Similarity: An Empirical Comparison. *Information* 13, 1 (2022), 42:1–42:18. https://doi.org/10.3390/info13010042