Conjunto	Eventos	Sessões	Itens	Δ	σ	Janela de tempo
Treinamento	90.801	26.649	1.702	3,3	2,0	01/01/2023 a 03/12/2023
Teste	11.265	3.106	589	3,6	2,0	02/08/2023 a 18/12/2023

Tabela 3.6: Informações sobre o *dataset* utilizado na abordagem *single*. Δ e σ são a média e o desvio padrão da quantidade de usuários por sessão.

3.4 Experimentos

3.4.1 Abordagem *single-split*, *session-based*, *next-item*

HIDASI *et al.* [70] utiliza a abordagem *single-split* para comparar a avaliação dos modelos, de forma que o conjunto de treinamento é composto pelos últimos seis meses, enquanto que o conjunto de teste é composto pelo último dia do conjunto completo [15]. Essa é a abordagem mais próxima de um modelo em produção.

O *dataset* é dividido segundo a abordagem *single split*, em um único conjunto de treinamento e teste, tal como na tabela 3.6. Nesses experimentos, os modelos de base de comparação desconsideram o usuário dono da sessão. Dessa forma, não há modelos *session-aware*.

Apenas sessões com um mínimo de dois usuários estão presentes, como ilustrado na figura 3.1. Similarmente, o suporte mínimo dos itens é igual a 2. Dessa forma, há apenas usuários que constam em pelo menos duas sessões publicadas. Essa abordagem é distinta da maioria dos comparativos, que utilizam suporte mínimo igual a 3. Essa decisão foi tomada pela menor quantidade de dados disponíveis, em comparação com as demais bases de dados. Essa distanção é necessária ser destacada, em razão do eventual uso desses dados para comparativos com outras bases. Uma vez que o foco do presente trabalho é a avaliação dentro do contexto da aplicação, essa foi a decisão mais adequada para manter a maior quantidade de dados possível.

Na abordagem apresentada, apenas a primeira interação de cada usuário na sessão consta na base, de forma que nenhum usuário apareça mais de uma vez na mesma sessão. Dessa forma, a tarefa do modelo é prever qual será o próximo usuário inédito a contribuir na sessão.

Modelos não-personalizados

Pop é o modelo de popularidade, com pontuações proporcionais ao suporte dos itens, limitado aos 100 itens mais populares. Random é o modelo aleatório, retornando uma pontuação aleatória para cada item. SPop é o modelo de popularidade de sessão, em que as pontuações são proporcionais à maior frequência dos itens em cada sessão, também limitado aos 100 itens mais populares. Finalmente, RPop é o modelo de popularidade para sessões recentes, em que apenas itens do último dia são

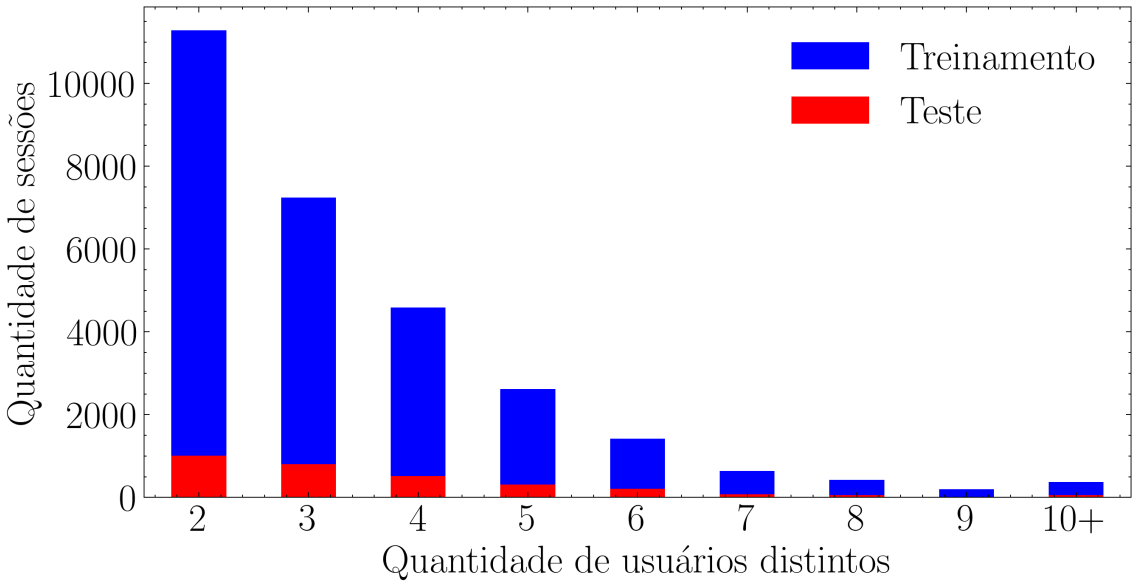

Figura 3.1: Distribuição da quantidade de usuários por sessão na abordagem *single*. A distribuição de treinamento está empilhada acima da distribuição de teste.

considerados por padrão. Foi utilizada a avaliação sobre o próximo item da sessão.

Modelo	HR@5	HR@10	MRR@5	MRR@10	NDCG@10	Cov@10	Pop@10
Pop	0,132	0,268	0,092	0,110	0,155	0,006	0,531
Random	0,003	0,006	0,001	0,002	0,003	**1,000**	**0,013**
RPop	**0,204**	**0,294**	**0,125**	**0,137**	**0,204**	0,010	0,321
SPop	0,109	0,221	0,045	0,058	0,114	0,301	0,473

De forma esperada, o modelo Random maximiza a cobertura, minimizando o índice de popularidade. O modelo de popularidade para sessões recentes (RPop) apresenta resultados superiores aos demais modelos de popularidade. Em seguida, na tabela 3.7, são apresentados os resultados para os modelos de mineração de padrões e vizinhança, fatoração de matrizes e redes neurais.

Modelos por mineração de padrões e vizinhança

Inclui regras de associação, cadeias de Markov, regras de sequência. Também é realizada para métodos de vizinhança: kNN, vskNN, STAN e VSTAN. Os modelos SR, skNN e vsKNN são otimizados com seus respectivos hiperparâmetros, obtidos a partir de uma otimização do MRR@10.

Os modelos de regras de sequência e baseados em cadeias de Markov, inclusive a árvore de contexto, apresentam resultados superiores, considerando as métricas HR@5, HR@10 e MRR@5 e MRR@10. Apesar dos modelos de vizinhança apresentarem bons resultados para o HR, o MRR decai consideravelmente nesses modelos.

Modelo	HR@5	HR@10	MRR@5	MRR@10	Cov@10	Pop@10	$\Delta t_{treino}[s]$
CT	**0,541**	**0,631**	**0,392**	**0,404**	0,518	0,358	8,3
SR_2	0,468	0,593	0,292	**0,310**	0,507	0,260	0,1
Markov	0,465	0,583	0,292	0,308	0,488	0,247	0,1
SR_1	0,461	0,583	0,291	0,308	0,492	0,272	0,1
$skNN_2$	0,444	0,604	0,157	0,179	0,603	0,217	0,1
$VSTAN_1$	0,419	0,573	0,161	0,182	0,553	0,230	0,1
$VSTAN_2$	0,412	0,582	0,148	0,171	0,571	0,236	0,1
$skNN_1$	0,406	0,564	0,151	0,172	**0,636**	**0,187**	0,1
AR	0,403	0,531	0,238	0,255	0,488	0,284	0.1
$STAN_1$	0,403	0,570	0,147	0,170	0,516	0,271	0.1
$STAN_2$	0,366	0,551	0,126	0,151	0,562	0,196	0.1
$vsKNN_1$	0,365	0,527	0,136	0,158	0,312	0,222	0.1
smf_1	**0,521**	**0,637**	**0,339**	**0,354**	0,613	0,228	1146,6
smf_2	0,506	0,616	0,332	0,346	0,321	0,256	987,4
FPMC	0,289	0,421	0,122	0,140	0,840	**0,226**	921,7
FISM	0,280	0,414	0,126	0,144	0,824	0,264	918,7
BPRMF	0,253	0,397	0,107	0,125	0,834	0,233	918,3
Fossil	0,243	0,399	0,100	0,121	**0,848**	0,253	917,7
GNN_1	**0,594**	**0,666**	**0,439**	**0,450**	0,749	0,221	483,7
GNN_2	0,588	**0,666**	0,425	0,435	0,713	0,220	464,6
$STAMP_1$	0,543	0,639	0,385	0,398	0,672	0,244	106,6
$STAMP_2$	0,539	0,638	0,384	0,397	0,635	0,243	106,6
$NextItNet_1$	0,426	0,526	0,277	0,290	0,336	0,278	1205,7
$NextItNet_2$	0,410	0,518	0,278	0,293	0,312	0,287	966,4
$NARM_1$	0,293	0,394	0,173	0,186	**0,718**	0,220	6633,7
$NARM_2$	0,273	0,377	0,165	0,179	0,699	**0,213**	2598,8
GRU4Rec	0,262	0,371	0,144	0,145	0,643	0,205	542,3
$CSRM_1$	0.229	0.315	0.134	0.146	0.623	0.148	128,5
$CSRM_2$	0.237	0.328	0.143	0.155	0.640	0.158	127,4

Tabela 3.7: Resultado para os modelos avaliando o próximo item da sessão, agrupados por abordagem: mineração de padrões e vizinhança, fatoração e redes neurais. Os hiperparâmetros utilizados para cada modelo estão descritos no apêndice.

Quando comparados aos demais agrupamentos, os modelos por mineração de padrões e vizinhança obtiveram bons resultados para HR@5 e HR@10. Entre os modelos de vizinhança, o modelo skNN_2 apresentou os melhores resultados, seguido pelo modelo VSTAN_1, que é modelo mais complexo dentre os modelos de vizinhança.

Métodos baseados em fatoração

Avalia-se os modelos FPMC, FISM, Fossil, BPRMF e smf. Os modelos smf apresentam resultados bem superiores aos demais, tanto no HitRate quanto no MRR. Vale observar que a duração do treinamento de cada um dos modelos baseados em fatoração é consideravelmente maior do que os modelos de mineração de padrões e vizinhança.

Modelos baseados em redes neurais

Inclui os modelos GRU4Rec, NARM, NextItNet, STAMP, GNN e CSRM. Os modelos GNN e STAMP apresentam os melhores resultados globais para as medidas de HR. Os modelos STAMP e NextItNet apresentam bons resultados, enquanto que os modelos NARM, GRU4Rec e CSRM apresentam resultados inferiores, inclusive quando comparados a modelos mais simples, tais como regras de sequência ou baseados em vizinhança.

3.4.2 Abordagem *windowed*, *session-based*, *next-item*

Algumas das limitações da abordagem *single-split* envolvem a maior suscetibilidade a efeitos aleatórios e a particularidades dos dados. Uma alternativa é a abordagem *windowed*, que minimiza os riscos de os resultados serem influenciados por uma única configuração de treino e teste. A abordagem *windowed* equivale a uma validação cruzada, com a limitação de que os dados são ordenados cronologicamente. LUDEWIG e JANNACH [15] dividem os dados em cinco janelas de um mês, em que o último dia de cada janela é reservado para teste. O resultado final para as métricas é a média aritmética dos métricas obtidas em cada janela.

Na abordagem *windowed*, uma janela deslizante é aplicada por todo o período, separando em cinco pares distintos de treinamento e teste, descritos na tabela 3.8. Nota-se o aumento gradual da quantidade de eventos, sessões e itens ao longo do tempo. Os resultados sob abordagem *windowed* são obtidos a partir da média aritmética dos resultados obtidos em cada janela. Os resultados constam na tabela 3.9.

Para cada agrupamento de modelos, os modelos smf, ct, SR, GNN e STAMP novamente apresentam os melhores resultados, tal que a GNN novamente apresenta

Conjunto	Índice	Eventos	Sessões	Itens	Data
Treinamento	1	3190	1098	145	08/01/2023 a 09/03/2023
Teste	1	205	72	40	09/03/2023 a 13/03/2023
Treinamento	2	3976	1346	197	14/03/2023 a 13/05/2023
Teste	2	244	68	53	13/05/2023 a 17/05/2023
Treinamento	3	7407	2191	301	18/05/2023 a 17/07/2023
Teste	3	466	162	100	17/07/2023 a 21/07/2023
Treinamento	4	21683	6267	754	22/07/2023 a 20/09/2023
Teste	4	1779	454	201	20/09/2023 a 24/09/2023
Treinamento	5	43993	12852	1681	25/09/2023 a 24/11/2023
Teste	5	2978	830	351	24/11/2023 a 28/11/2023

Tabela 3.8: Conjuntos de treino e teste separados em cinco janelas para abordagem *windowed*, *session-based*.

o melhor resultado global para as métricas HR@5, HR@10. Novamente, os modelos CSRM e GRU4Rec apresentam os piores resultados globais para o HR@5 e HR@10.

Na figura 3.2, é possível observar que os modelos de redes neurais começam com resultados inferiores aos demais modelos, mas passam a superar os demais conforme a quantidade de eventos por *split* aumenta. Essa observação condiz com a característica de modelos de aprendizado profundo, em geral, dependerem de uma grande quantidade de dados para apresentarem resultados superiores. A figura 3.3 apresenta a progressão do HR e MRR conforme a quantidade de valores preditos para cada métrica.

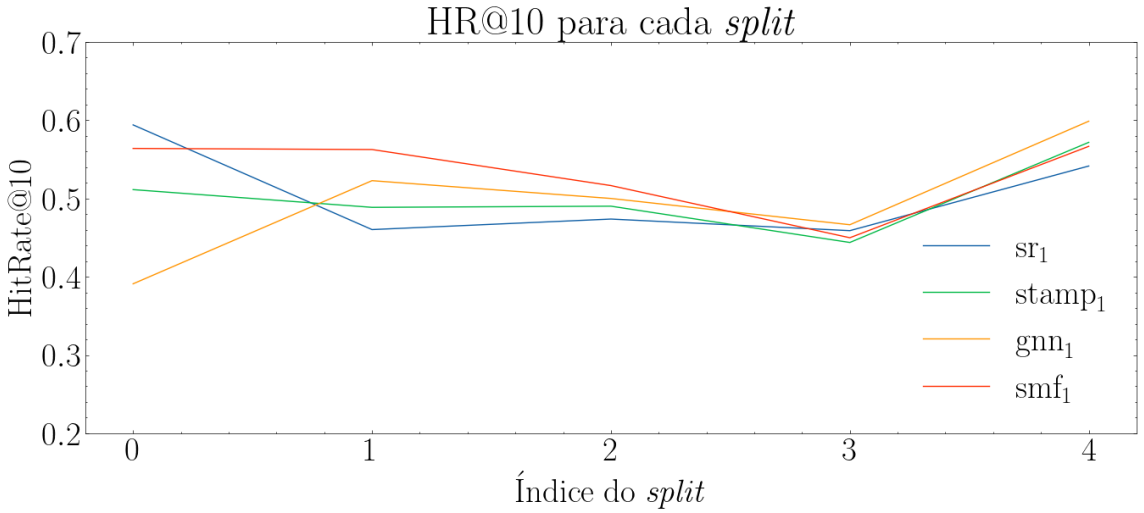

Figura 3.2: HR@10 em cada janela para alguns modelos da tabela 3.9.

Modelo	HR@5	HR@10	MRR@5	MRR@10	Cov@10	Pop@10	$\Delta t[s]$
RPop	**0,222**±0,061	**0,356**±0,079	**0,116**±0,045	**0,13**3±0,046	0,065±0,045	0,355±0,073	0,006
Pop	0,202±0,056	0,325±0,063	0,114±0,047	0,130±0,045	0,034±0,023	0,493±0,099	0,001
SPop	0,168±0,073	0,298±0,072	0,072±0,030	0,089±0,029	0,235±0,036	0,456±0,094	0,002
Random	0,012±0,006	0,032±0,019	0,007±0,004	0,009±0,005	**1,000**±0,000	**0,045**±0,024	0,001
smf_1	**0,401**±0,040	**0,532**±0,045	**0,247**±0,032	**0,264**±0,030	0,492±0,020	**0,281**±0,082	193
smf_2	0,364±0,030	0,495±0,039	0,224±0,024	0,240±0,024	0,271±0,044	0,328±0,101	162
FISM	0,291±0,051	0,443±0,081	0,144±0,038	0,164±0,041	**0,625**±0,041	0,346±0,099	892
FPMC	0,291±0,063	0,438±0,088	0,140±0,040	0,159±0,042	0,613±0,041	0,350±0,098	897
BPRMF	0,290±0,049	0,421±0,054	0,149±0,036	0,166±0,036	0,624±0,044	0,343±0,095	893
Fossil	0,288±0,033	0,417±0,041	0,147±0,028	0,164±0,027	0,605±0,058	0,343±0,091	900
ct	**0,413**±0,058	0,525±0,053	**0,274**±0,041	**0,289**±0,041	0,375±0,009	0,396±0,098	1,198
SR_1	0,378±0,053	0,506±0,054	0,225±0,021	0,242±0,022	0,443±0,055	0,281±0,061	0,056
SR_2	0,375±0,058	0,508±0,042	0,224±0,023	0,241±0,023	0,453±0,053	0,273±0,062	0,065
AR	0,363±0,055	0,476±0,041	0,207±0,041	0,222±0,039	0,455±0,067	0,300±0,073	0,117
$VSTAN_1$	0,363±0,048	0,505±0,061	0,145±0,021	0,164±0,022	0,487±0,067	0,291±0,062	0,291
Markov	0,360±0,054	0,477±0,040	0,218±0,020	0,233±0,020	0,438±0,047	0,257±0,056	0,047
$VSTAN_2$	0,358±0,061	0,493±0,058	0,137±0,024	0,155±0,023	0,500±0,071	0,293±0,063	0,293
$skNN_1$	0,351±0,075	**0,526**±0,061	0,140±0,029	0,164±0,026	**0,534**±0,058	0,258±0,062	0,079
$STAN_2$	0,346±0,065	0,493±0,057	0,129±0,021	0,149±0,020	0,506±0,071	0,262±0,059	0,037
$vsKNN_2$	0,340±0,051	0,477±0,060	0,132±0,021	0,150±0,018	**0,534**±0,056	**0,233**±0,052	0,077
$STAN_1$	0,339±0,067	0,494±0,072	0,132±0,025	0,153±0,025	0,463±0,071	0,324±0,068	0,324
$skNN_2$	0,334±0,070	0,507±0,068	0,132±0,024	0,156±0,024	0,491±0,060	0,282±0,065	0,021
$vsKNN_1$	0,299±0,051	0,458±0,040	0,119±0,020	0,140±0,019	0,511±0,075	0,267±0,060	0,029
GNN_2	**0,421**±0,049	**0,551**±0,056	**0,264**±0,062	**0,280**±0,061	0,517±0,055	0,296±0,089	122
$STAMP_2$	0,395±0,039	0,515±0,035	0,249±0,037	0,266±0,035	0,592±0,041	0,268±0,071	31,1
GNN_1	0,393±0,062	0,496±0,068	0,254±0,056	0,268±0,056	0,544±0,060	0,291±0,079	120
$STAMP_1$	0,382±0,047	0,501±0,042	0,253±0,047	0,269±0,046	0,599±0,059	0,269±0,059	32,0
$NextItNet_2$	0,352±0,045	0,461±0,039	0,235±0,042	0,250±0,041	0,381±0,088	0,316±0,097	83,8
$NextItNet_1$	0,336±0,076	0,451±0,059	0,221±0,059	0,236±0,056	0,326±0,107	0,327±0,109	107,3
$NARM_2$	0,295±0,048	0,426±0,067	0,173±0,034	0,190±0,037	0,586±0,049	0,264±0,076	193,9
$NARM_1$	0,274±0,037	0,418±0,059	0,167±0,024	0,187±0,026	0,576±0,029	0,267±0,085	372,2
$CSRM_1$	0,201±0,034	0,300±0,030	0,117±0,017	0,130±0,016	0,492±0,063	0,267±0,085	19,8
$CSRM_2$	0,191±0,029	0,306±0,050	0,104±0,027	0,119±0,028	0,439±0,046	0,278±0,085	19,9
GRU4Rec	0,176±0,055	0,235±0,073	0,099±0,016	0,107±0,018	**0,731**±0,088	**0,079**±0,051	63,3

Tabela 3.9: Resultado dos modelos *session-based* na abordagem *windowed*, avaliando o próximo item da sessão. Valores exibidos são a média dos cinco *splits* acompanhada da variância. Valores agrupados por abordagem e ordenados internamente por HR@5. Maiores valores para cada medida de desempenho estão em negrito nos agrupamentos.

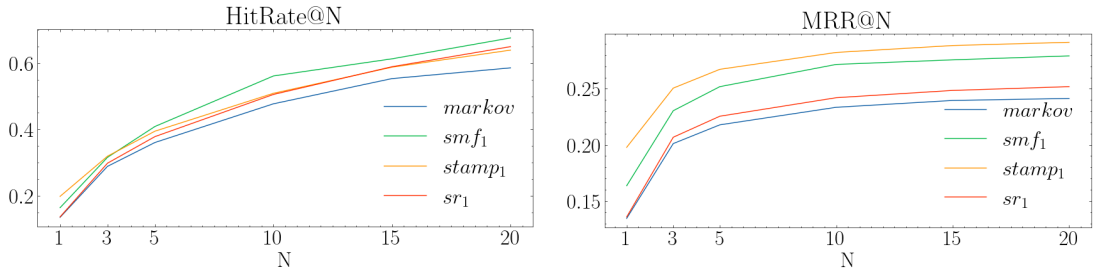

Figura 3.3: Progressão do HR e MRR conforme a quantidade de valores preditos.

3.4.3 Abordagem *windowed*, *session-based*, *remaining-items*

Para a abordagem *remaining-items*, foi realizada uma nova otimização considerando a métrica MAP@10, uma vez que essa métrica substitui o HR@10 para avaliações de listas ou de conjuntos de itens. Dessa forma, os modelos com nome subscrito aqui presentes possuem hiperparâmetros distintos da abordagem anterior, constando no apêndice.

Os modelos que apresentaram melhores resultados para precisão, *recall*, NDCG e MAP foram os mesmos dos experimentos anteriores: smf, ct, SR, GNN e STAMP. Uma diferença observada foi o pior inferior do modelo baseado em cadeia de Markov de primeira ordem dentro de seu agrupamento, apesar do modelo de árvore de contexto permenecer com bons resultados. Também é possível observar que os modelos do agrupamento de regras de sequência e baseados em vizinhança obtiveram bons resultados, quando comparados aos demais agrupamentos.

Modelos	P@10	R@10	NDCG@10	MAP@10	Cov@10	Pop@10
rpop	**0,086**±0,011	**0,361**±0,092	**0,240**±0,060	**0,042**±0,011	0,065±0,045	0,355±0,073
pop	0,075±0,025	0,326±0,063	0,220±0,047	0,038±0,010	0,034±0,023	0,493±0,099
spop	0,065±0,017	0,298±0,053	0,179±0,036	0,032±0,009	0,235±0,036	0,456±0,094
random	0,007±0,004	0,033±0,023	0,018±0,011	0,003±0,002	**1,000**±0,000	**0,046**±0,024
smf_2	**0,100**±0,017	**0,459**±0,035	**0,338**±0,027	**0,055**±0,006	0,343±0,044	**0,321**±0,103
smf_1	**0,100**±0,017	0,450±0,035	0,333±0,023	**0,055**±0,005	0,242±0,043	0,347±0,109
fpmc	0,086±0,014	0,403±0,079	0,267±0,049	0,044±0,007	0,607±0,069	0,345±0,094
bprmf	0,084±0,013	0,389±0,066	0,260±0,049	0,044±0,007	0,612±0,051	0,347±0,093
fism	0,083±0,011	0,396±0,084	0,264±0,055	0,045±0,009	**0,616**±0,047	0,346±0,101
fossil	0,081±0,014	0,373±0,048	0,252±0,042	0,042±0,005	0,606±0,042	0,346±0,093
SR_1	**0,104**±0,018	0,457±0,055	0,337±0,030	0,057±0,007	0,434±0,061	0,286±0,066
ct	0,103±0,015	0,479±0,047	**0,367**±0,039	0,056±0,006	0,375±0,009	0,396±0,098
$vsknn_1$	0,101±0,015	0,471±0,054	0,301±0,036	0,057±0,008	0,477±0,063	0,295±0,067
SR_2	0,101±0,014	0,455±0,051	0,332±0,035	0,056±0,006	0,453±0,053	0,273±0,062
$STAN_1$	0,100±0,011	**0,482**±0,068	0,303±0,042	**0,059**±0,008	0,464±0,067	0,304±0,064
$STAN_2$	0,099±0,012	0,469±0,072	0,296±0,043	0,057±0,008	0,467±0,071	0,316±0,065
ar	0,098±0,014	0,442±0,047	0,331±0,040	0,054±0,006	0,455±0,067	0,300±0,073
$VSTAN_1$	0,098±0,013	0,463±0,070	0,296±0,047	0,057±0,008	0,468±0,072	0,319±0,068
$skNN_1$	0,097±0,012	0,471±0,072	0,294±0,042	0,056±0,009	0,468±0,068	0,318±0,075
$skNN_2$	0,095±0.014	0,487±0,065	0,300±0,045	0,057±0,009	0,540±0,055	0,260±0,059
$vsKNN_2$	0,091±0,019	0,426±0,051	0,268±0,033	0,051±0,010	0,542±0,063	0,216±0,047
Markov	0,091±0,013	0,420±0,041	0,310±0,031	0,050±0,004	0,438±0,047	0,257±0,056
$vstan_2$	0,088±0,012	0,441±0,052	0,266±0,033	0,052±0,007	**0,571**±0,056	**0,212**±0,050
GNN	**0,097**±0,020	**0,455**±0,056	**0,359**±0,044	**0,056**±0,009	0,603±0,060	0,272±0,075
$STAMP_2$	**0,097**±0,018	0,453±0,033	0,341±0,041	0,052±0,006	0,608±0,038	0,285±0,072
$STAMP_1$	0,096±0,017	0,445±0,042	0,331±0,043	0,052±0,006	0,560±0,056	0,286±0,080
$NextItNet_2$	0,095±0,014	0,435±0,032	0,310±0,011	0,051±0,003	0,531±0,034	0,293±0,078
$NextItNet_1$	0,090±0,019	0,427±0,036	0,314±0,049	0,048±0,007	0,321±0,115	0,340±0,112
$NARM_2$	0,079±0,008	0,392±0,045	0,286±0,035	0,043±0,004	0,641±0,037	0,256±0,071
$NARM_1$	0,079±0,010	0,381±0,045	0,266±0,017	0,042±0,004	0,610±0,036	0,256±0,068
$CSRM_2$	0,067±0,013	0,310±0,061	0,209±0,043	0,034±0,007	0,459±0,070	0,278±0,099
$CSRM_1$	0,064±0,011	0,314±0,054	0,210±0,030	0,033±0,004	0,473±0,052	0,279±0,089
$GRU4Rec_2$	0,017±0,015	0,081±0,060	0,053±0,048	0,008±0,007	0,416±0,234	0,026±0,021
$GRU4Rec_1$	0,008±0,004	0,035 ±0,015	0,022±0,012	0,003±0,002	**0,707**±0,092	**0,022**±0,006

Tabela 3.10: Resultados para abordagem *remaining-items*. Em ordem, modelos de popularidade, fatoração de matrizes, modelos baseados em vizinhança, regras de associação e redes neurais.

Conjunto	Índice	Eventos	Usuários	Sessões	Itens	Data – 2023
Treino	1	2705	76	721	124	01/08 a 13/03
Teste	1	245	73	73	75	16/03 a 13/03
Validação – Treino	1	2427	76	645	119	01/08 a 12/03
Validação – Teste	1	269	74	74	75	16/01 a 13/03
Treino	2	3463	74	898	160	14/03 a 17/05
Teste	2	256	73	73	87	20/03 a 17/05
Validação – Treino	2	3186	74	824	158	14/03 a 17/05
Validação – Teste	2	275	74	74	93	20/03 a 17/05
Treino	3	6620	109	1520	270	18/05 a 21/06
Teste	3	368	108	108	123	01/06 a 21/07
Validação – Treino	3	6226	109	1411	265	18/05 a 21/07
Validação – Teste	3	386	106	106	130	23/05 a 21/07
Treino	4	15180	308	3143	581	21/07 a 24/09
Teste	4	1614	305	305	318	23/07 a 24/09
Validação – Treino	4	13483	308	2835	556	21/07 a 24/09
Validação – Teste	4	1630	298	298	305	22/07 a 24/09
Treinamento	5	23244	646	4369	1092	24/09 a 28/11
Teste	5	3317	629	629	670	30/09 a 28/11
Validação – Treino	5	19820	646	3723	1027	24/09 a 28/11
Validação – Teste	5	3288	625	625	676	26/09 a 28/11

Tabela 3.11: Conjuntos de treino, teste e validação separados em cinco janelas para abordagem *session-aware*.

3.4.4 Abordagem *windowed*, *session-aware*, *next-item*

Em seguida, são apresentados os resultados obtidos com a abordagem *windowed* e *session-aware*. Os modelos passaram por otimização de seus hiperparâmetros, cujos resultados constam no apêndice.

A tabela 3.11 apresenta os conjuntos de treino e teste separados em cinco janelas. Novamente, é possível observar o aumento gradual da quantidade de eventos, sessões e itens ao longo dos *splits*.

Metade dos modelos *session-aware* apresentam resultados que superam todos os modelos da abordagem *session-based*. Os maiores valores para as medidas de avaliação entre todos os experimentos foram obtidos na presente abordagem. O modelo de regras de sequência *session-aware* é o único modelo com MRR@5 acima de 0,3. Os modelos NARM e iiRNN também obtiveram MRR@10 superior a 0,3. O modelo vsKNN *session-aware* obteve um incremento considerável no HR@10 quando comparado ao seu equivalente *session-based*. Os modelos *session-aware* baseados no GRU4REC obtiveram resultados inferiores, o que também foi observado na abordagem *session-based*.

Modelo	HR@5	HR@10	MRR@5	MRR@10	Cov@10	Pop@10
USR	**0,487**±0,046	0,627±0,026	**0,325**±0,031	**0,344**±0,028	0,812±0,062	0,242±0,050
USTAN	0,484±0,047	0,626±0,059	0,256±0,032	0,275±0,031	0,724±0,050	0,298±0,060
UVSKNN	0,474±0,068	**0,648**±0,057	0,243±0,035	0,267±0,033	0,697±0,080	0,300±0,059
UNARM	0,465±0,037	0,610±0,049	0,296±0,018	0,315±0,017	0,832±0,072	0,236±0,051
iiRNN	0,442±0,058	0,554±0,047	0,293±0,054	0,308±0,052	0,721±0,073	0,230±0,030
NSAR	0,393±0,056	0,522±0,067	0,253±0,045	0,271±0,047	0,586±0,102	0,279±0,069
NCFS	0,368±0,077	0,493±0,060	0,225±0,064	0,241±0,061	0,452±0,0238	0,370±0,139
UGRU4Rec	0,358±0,045	0,471±0,051	0,238±0,031	0,253±0,030	**0, 910**±0,035	0,116±0,037
SHAN	0,343±0,052	0,471±0,065	0,202±0,029	0,219±0,030	0,470±0,077	0,309±0,050
HGRU4Rec	0,296±0,028	0,377±0,045	0,190±0,018	0,201±0,020	0,762±0,075	**0, 113**±0,35

Tabela 3.12: Resultado dos modelos *session-aware* na abordagem *windowed*, avaliando o próximo item da sessão. Valores exibidos são a média dos cinco *splits* acompanhada da variância. Modelos ordenados por HR@5.

Modelo	$\Delta t[s]$
USR	0,112
USTAN	108,7
UVSKNN	0,079
UNARM	244,6
iiRNN	145,7
NSAR	77,2
NCFS	21,6
UGRU4Rec	44,4
SHAN	642,3
HGRU4Rec	14,5

Tabela 3.13: Duração média do treinamento de cada modelo da tabela 3.12

	Posição	HR@5	HR@10	MRR@5	MRR@10	COV@10	POP@10
RSC15	2^o	**0,457**	**0,569**	0,283	0,298	0,592	0,073
ZALANDO	1^o	0,429	0,462	**0,298**	**0,302**	0,433	0,066
INDABAND	7^o	0,378	0,506	0,225	0,242	0,443	0,281
CLEF	6^o	0,337	0,513	0,189	0,212	**0,608**	0,123
RETAILROCKET	7^o	0,337	0,386	0,236	0,243	0,458	0,050
30MUSIC	1^o	0,2845	0,2326	0,3120	0,2363	0,2913	0,0273
TMALL	8^o	0,173	0,209	0,124	0,129	0,507	**0,020**
NOWPLAYING	1^o	0,1395	0,1712	0,0988	0,1031	0,3605	0,0284
AOTM	3^o	0,0107	0,0149	0,0068	0,0073	0,4481	0,0599
8TRACKS	4^o	0,0090	0,0125	0,0056	0,0061	0,1076	0,0916

Tabela 3.14: Regras de sequência, *windowed*, *session-based*. Ordenados por HR@5.

	HR@5	HR@10	MRR@5	MRR@10
RETAIL ROCKET	**0,833**	**0,886**	**0,621**	**0,629**
DIGINETICA	0,469	0,591	0,305	0,321
RSC15 (1/64)	0,450	0,563	0,313	0,328
INDABAND	0,421	0,551	0,264	0,280
RSC15 (1/12)	0,384	0,497	0,230	0,245

Tabela 3.15: sr-GNN, *windowed*, *session-based*. Ordenados por HR@5.

3.4.5 Medidas de avaliação nas demais bases de dados

As tabelas 3.14, 3.15 e 3.16 trazem as medidas de avaliação obtidas para os modelos que se destacaram no comparativo em relação às demais bases. A tabela 3.14 compara os resultados obtidos para regras de sequência, a tabela 3.15 compara os resultados obtidos para sr-GNN e a tabela 3.16 compara os resultados obtidos para smf. Os modelos estão ordenados por HR@5. Para as tabelas 3.14 e 3.16, foi adicionada a posição de cada modelo nos comparativos de suas respectivas bases de dados. Os valores foram retirados de LUDEWIG [17] e SHEHZAD e JANNACH [78].

Observa-se que os valores obtidos com a base do presente trabalho estão na mesma ordem de grandeza de algumas das demais bases de dados, a exemplo de RSC15, ZALANDO, INDABAND, CLEF e RETAILROCKET. Algumas bases obtém valores baixos para as métricas de avaliação, mesmo nos modelos com melhores resultados para essas bases.

3.4.6 Base restrita a faixas inéditas

Uma característica particular das sessões do Indaband é a capacidade de criar uma sessão a partir de outra já existente, modificá-la, adicionar novas gravações e publicá-la como uma nova iteração a partir da funcionalidade de *fork*.

	Posição	HR@5	HR@10	MRR@5	MRR@10	COV@10	POP@10
RSC15	3º	**0,459**	**0,575**	**0,280**	**0,295**	**0,486**	0,073
INDA	3º	0,401	0,532	0,247	0,264	0,492	0,281
ZALANDO	8º	0,380	0,418	0,259	0,265	0,239	0,103
CLEF	3º	0,354	0,529	0,198	0,222	0,582	0,097
RETAILROCKET	1º	0,322	0,393	0,211	0,221	0,360	0,092
30MUSIC	5º	0,2223	0,2547	0,1712	0,1756	0,1117	0,1062
TMALL	9º	0,168	0,213	0,112	0,118	0,193	**0,039**
8TRACKS	3º	0,0092	0,0148	0,0051	0,058	0,1076	0,0916
NOWPLAYING	6º	0,1181	0,1484	0,0818	0,0859	0,1847	0,0960
AOTM	1º	0,0149	0,0205	0,0097	0,105	0,1795	0,2085

Tabela 3.16: smf, *windowed*, *session-based*. Ordenados por HR@5.

Essa funcionalidade é muito utilizada pelos usuários. A tabela 3.1 mostra que 76% das faixas criadas são geradas via *fork*, independentemente se foram publicadas ou não. Dessa forma, usuários distintos podem contribuir para uma mesma sessão em momentos distintos, ou um usuário pode contribuir para uma mesma sessão de outro usuário que era desconhecido até então.

Uma vez que o objetivo é recomendar usuários prováveis de contribuir gravando uma faixa inédita para uma determinada sessão, é razoável gerar um cenário em que o recomendador seja treinado e avaliado estritamente para prever a próxima faixa inédita gravada.

As abordagens *next-item* e *remaining-items* são as mais utilizadas nos comparativos publicados porque aproveitam todos os itens de cada sessão para o treinamento ou para a avaliação. No caso de uma aplicação em que haja redundância entre as sessões, em que usuários possam salvar, repetir e compartilhar sessões entre si, essas abordagens acabam por facilitar o trabalho do recomendador, uma vez que cada iteração a mais no processo de treinamento significa uma nova oportunidade para minimizar o erro do modelo. O mesmo vale para a avaliação, em que identificar faixas *forkadas* seria em tese uma tarefa mais fácil. O cenário mais desafiador é aquele em que o item a ser avaliado é necessariamente um item inédito no contexto daquela sessão.

Em contraponto, a abordagem *last-item* é a que mais se aproxima do cenário mencionado, uma vez que mantém inalterada a sequência de itens das sessões. O que muda é a forma de avaliação, que passa a considerar apenas o último item da sequência como o item a ser previsto, sem que os demais itens anteriores sejam avaliados, minimizando a influência de itens *forkados* na avaliação.

Com essa finalidade, a base de treinamento do presente trabalho é filtrada: Nesse último experimento, constam apenas sessões em que o último item da sequência é a adição de uma faixa inédita. Essa identificação é feita a partir de um metadado

disponível, informando se a faixa foi gerada por um *fork* ou foi criada, seja por gravação, por importação de uma faixa ou por separação de fontes.

Para essa abordagem, é utilizado o avaliador *last-item*. Esse avaliador considera apenas o último item da sequência como o item a ser previsto, sem que os demais itens advindos de *fork* sejam recomendados.

Modelo	HR@5	HR@10	MRR@5	MRR@10	Cov@10	Pop@10
GNN	**0,551**	**0,625**	**0,437**	**0,446**	0,660	0,202
$STAMP_1$	0,519	0,584	0,377	0,386	0,610	0,217
sknn2	0,518	0,647	0,213	0,230	0,643	0,169
sknn1	0,515	0,642	0,209	0,226	0,592	0,197
ct	0,509	0,579	0,390	0,399	0,477	0,349
$STAMP_2$	0,506	0,590	0,366	0,377	0,609	0,201
$NextItNet_2$	0,465	0,563	0,341	0,354	0,566	0,249
$STAN_2$	0,464	0,596	0,178	0,196	0,547	0,179
$STAN_1$	0,451	0,572	0,186	0,202	0,529	0,177
$NextItNet_1$	0,449	0,536	0,337	0,349	0,525	0,257
vsknn1	0,447	0,580	0,1193	0,211	0,544	0,208
$VSTAN_1$	0,441	0,580	0,180	0,198	0,507	0,221
SMF_2	0,439	0,553	0,291	0,306	0,306	0,240
SMF_1	0,431	0,530	0,298	0,311	0,240	0,257
vsknn2	0,427	0,564	0,172	0,190	0,618	0,163
SR_2	0,423	0,534	0,277	0,291	0,471	0,238
$VSTAN_2$	0,418	0,546	0,171	0,189	0,588	0,142
SR_1	0,410	0,512	0,279	0,292	0,445	0,259
ar	0,395	0,498	0,248	0,262	0,454	0,265
$NARM_1$	0,361	0,455	0,229	0,241	0,658	0,199
$NARM_2$	0,324	0,440	0,211	0,226	0,665	0,188
FOSSIL	0,287	0,420	0,137	0,154	0,776	0,249
FPMC	0,281	0,419	0,124	0,143	0,764	0,241
FISM	0,275	0,434	0,120	0,141	0,749	0,240
BPRMF	0,271	0,407	0,122	0,141	0,744	0,250
CSRM	0,264	0,343	0,158	0,168	0,596	0,137
$GRU4Rec_2$	0,208	0,278	0,130	0,139	0,873	**0,030**
$GRU4Rec_1$	0,195	0,271	0,123	0,134	**0,858**	0,040
rpop	0,165	0,225	0,106	0,113	0,009	0,319
pop	0,127	0,242	0,083	0,098	0,006	0,530
spop	0,125	0,219	0,059	0,071	0,299	0,471
random	0,003	0,005	0,001	0,001	1,000	0,014

Tabela 3.17: Resultado dos modelos *session-based* na abordagem *single item*, avaliando o último item da sessão. Modelos ordenados por HR@5.

A tabela 3.4.6 mostra os resultados obtidos pela abordagem *last-item*. As métricas obtiveram valores na mesma faixa da abordagem *next-item*, demonstrando que, por mais que a avaliação e o treinamento seja reservados apenas ao último item da

sequência, os modelos mantém desempenho equivalente.

Novamente, o modelo GNN obteve os melhores resultados. Modelos mais simples, como o skNN e a árvore de contexto também obtiveram bons resultados.