Benchmarking Session-based and Session-aware Recommender Systems for Jusbrasil

Marcos Aurélio Domingues
madomingues@uem.br
State University of Maringá and Federal University of Amazonas and Jusbrasil
Maringá, Brazil

Edleno Silva de Moura
edleno@icomp.ufam.edu.br
Federal University of Amazonas and Jusbrasil
Manaus, Brazil

Leandro Balby Marinho
lbmarinho@computacao.ufcg.edu.br
Federal University of Campina Grande
Campina Grande, Brazil

Altigran da Silva
alti@icomp.ufam.edu.br
Federal University of Amazonas
Manaus, Brazil

Abstract

In this paper, we present a benchmark of several session-based, session-based with reminders and session-aware recommender systems that can be used to improve legal document recommendation in Jusbrasil, the largest legal search engine in Brazil. We focus this benchmark on the logged users, and the results show that some recommender systems can achieve gains of accuracy of around 19% with respect to the current recommender system adopted by Jusbrasil.

Keywords: legal document recommendation, session-based recommender systems, reminders, session-aware recommender systems

1 Introduction

Jusbrasil[1] is known as the largest legal search engine in Brazil. With the goal of combining law and technology so that justice crosses the borders of the courts and reaches the homes of any citizen, it provides an on-line platform where users can find the legal documents that best match their information needs [18].

Millions of people currently access the company's platform. On the other hand, its database has billions of documents containing different artifacts related to law in Brazil. The two main ways of finding legal documents in Jusbrasil are by using the search engine and the recommender system provided by the company. In this work, we present a benchmark of several session-based, session-based with reminders, and session-aware recommender systems that can be used to improve legal document recommendation in Jusbrasil. The benchmark is focused on the logged users, and our experiments show that of evaluated some recommender systems can achieve gains of accuracy up to 19% compared to the current recommender system in Jusbrasil. With this work, we expect to provide new directions and insights into the performance of different recommender systems for the legal domain.

2 Recommender Systems

A recommender system is an information filtering technology that can be used to recommend items that may be of interest to users [23]. In our case, an item is a recommendable legal document. In this benchmark, we explore three paradigms of recommender systems: session-based, session-based with reminders, and session-aware.

Session-based recommenders: For these systems, the input corresponds to logs of recorded user-item interactions, where the interactions are grouped into anonymous sessions. A session is a sequence of interactions (e.g. clicks) with a clear boundary. With such an input, the system is able to recommend the next item (or all subsequent items) of interest, given only the interactions of a current session of the user. In this paradigm, we explore five categories of recommendation models:

- *Non-personalized Models.* A non-personalized model usually provides the most popular or random items available as recommendations. We have explored four non-personalized models for Jusbrasil [13]: **random**, **pop**, **rpop** and **spop**;
- *Models Based on Pattern Mining.* These models assess the strength of simple two items co-occurrence patterns, and we have included three models in our benchmark: **ar** [16], **markov** [19] and **sr** [12]. It is worth to say that the **ar** model is a proxy for the current recommender system in Jusbrasil. This model is a simplified version of the association rule mining

[1] https://www.jusbrasil.com.br

In: WebMedia in Practice, Curitiba, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2022.

ISSN 2596-1683

technique [1], and the recommendations are generated by returning those items that most frequently co-occurred with the last item of the current session [16];

- *Nearest Neighbors Models.* The models in this category try to identify sessions with the most similar interactions (nearest neighbours) in order to suggest items that are the most popular among these neighbours. In our benchmark we have included five models: **iknn** [5], **sknn** [9], **vsknn** [16], **stan** [3] and **vstan** [17];
- *Factorization Models.* In its natural form, the *factorization models* characterize items and users using vectors of factors (aka embeddings) inferred from item rating patterns. However, a number of *factorization models* were proposed in recent years for session-based recommender systems, and we have included five existing models in our analysis: **bprmf** [16], **fpmc** [22], **fism** [11], **fossil** [4] and **smf** [16];
- *Neural Network Models.* Models based on neural networks represent the most recently explored family of techniques for session-based recommender systems. Four different deep neural network models are included in our benchmark: **gru4rec** [5], **narm** [14], **stamp** [15] and **sgnn** [24].

Session-based recommenders with reminders: The reminders can be defined as a technique to emphasize items that the user has browsed before in previous sessions by placing some of them into the recommendation list built by the session-based model [10]. In this work, we use a hybrid reminder to combine aspects of interaction recency, session similarity, and item relevance score [13]. We have applied such a technique in the following session-based models [13]: **sr**, **vsknn**, **stan**, **vstan**, **gru4rec** and **narm**.

Session-aware recommenders: The systems are similar to session-based ones but the users are not anonymous, i.e., the recommender system has access to to the current and previous sessions of users. In our benchmark, we have included four neural network models proposed in the literature: **hgru4rec** [21], **ncfs** [7], **nsar** [20] and **shan** [25].

All recommendation models used in this benchmark are available in the recommendation framework **session-rec**[2], and were ran in our empirical evaluation with their default parameter values, providing promising results.

3 Empirical Evaluation

In general, the main task of session-based, session-based with reminders, and session-aware recommender systems is to generate a ranked list of recommendations given the current session. The performance of such systems can be measured by resorting to assess the capability of a model to predict the withheld entries of a session [16]. Thus, to evaluate the models, we withhold all items and iteratively reveal one item after the other as it reflects the user journey throughout a session in the best way [6].

Technically, we predict the *immediate next item* given the first n items in the current session. For each session, we iteratively increment n, measure the Hit Rate and the Mean Reciprocal Rank (MRR), and then calculate the average values for all sessions for the different list lengths. Besides the accuracy measurements, we also made two additional quality measurements in this work [9]: Coverage and Popularity bias. Coverage can be defined as the number of different items in the catalog of available items that ever appear in the top-k recommendations. Popularity bias can be used to measure if high accuracy values are correlated with the tendency of a model to recommend highly popular items [8].

As our goal is to benchmark recommender systems for the logged users in Jusbrasil, we have used a dataset that we called **jusbrasilrec_logged_users**, which contains 30 days of data from only the logged users. In this dataset, each session is defined for each 30 minutes of user inactivity [2]. The dataset was preprocessed to keep a minimum of 3 sessions per user, where each session contains a minimum of 2 and a maximum of 50 items. Thus, the dataset used in this work contains 13308981 interactions from 673580 users to 3083495 items, generating a total of 2788281 sessions.

We carried out the experiments by applying a sliding window protocol [16], where the data sessions are split into 5 slices of 6 days, where we have 5 days of training and 1 day of test data. Table 1 reports the average of the results for all slices at a top-10 (i.e. length 10) recommendation list for the task of predicting the *immediate next item* in a session.

Regarding the session-based paradigm, we can see in Table 1 that the lowest accuracy values are obtained with the *non-personalized models.* The *models based on pattern mining* exhibited a competitive performance mostly occupying the middle places in the table. In the *nearest neighbors* category, we have the highest accuracy models (i.e. **stan** with a HitRate of 0.752 and **vstan** with a MRR of 0.602). Following, the models in the *neural network* category reached values of HitRate and MRR very close to the best models in the *nearest neighbors* category, however, the *neural network* models are much more time training demanding than the *nearest neighbors* ones. Finally, the results for the *factorization* category were not so consistently.

With respect to the session-based with remainders paradigm, we can see that by using the hybrid reminder technique with its default parameter values, the variant models do not provides expressive gains compared to the respective original models. In some cases, the variants were even worse than the original models (i.e. without the reminders). However, this fact does not mean that the usage of reminders is not

[2] https://github.com/rn5l/session-rec

Table 1. Hit Rate, Mean Reciprocal Rank (MRR), Coverage, and Popularity bias for a top-10 obtained for the *immediate next item* recommendation task in the jusbrasilrec_logged_users dataset. The highest values for each paradigm are highlighted in boldface.

Paradigms	Models	HitRate@10	MRR@10	Coverage@10	Popularity@10
Session-based	random	0.000 ± 0.000	0.000 ± 0.000	**1.000 ± 0.000**	0.014 ± 0.001
	pop	0.017 ± 0.001	0.006 ± 0.000	0.000 ± 0.000	**0.735 ± 0.065**
	rpop	0.017 ± 0.001	0.006 ± 0.000	0.000 ± 0.000	0.700 ± 0.090
	spop	0.602 ± 0.002	0.520 ± 0.002	0.575 ± 0.009	0.630 ± 0.053
	ar	0.671 ± 0.002	0.506 ± 0.002	0.780 ± 0.008	0.077 ± 0.006
	markov	0.650 ± 0.004	0.535 ± 0.001	0.724 ± 0.010	0.065 ± 0.005
	sr	0.665 ± 0.003	0.529 ± 0.002	0.747 ± 0.009	0.071 ± 0.005
	iknn	0.183 ± 0.004	0.108 ± 0.003	0.714 ± 0.008	0.043 ± 0.003
	sknn	0.740 ± 0.003	0.530 ± 0.002	0.773 ± 0.010	0.083 ± 0.007
	vsknn	0.751 ± 0.003	0.573 ± 0.001	0.770 ± 0.009	0.086 ± 0.007
	stan	**0.752 ± 0.002**	0.590 ± 0.001	0.769 ± 0.010	0.082 ± 0.006
	vstan	0.746 ± 0.003	**0.602 ± 0.001**	0.779 ± 0.009	0.067 ± 0.005
	bprmf	0.565 ± 0.003	0.504 ± 0.002	0.927 ± 0.006	0.094 ± 0.007
	fpmc	0.558 ± 0.002	0.512 ± 0.002	0.990 ± 0.002	0.032 ± 0.002
	fism	0.277 ± 0.010	0.215 ± 0.009	0.956 ± 0.004	0.022 ± 0.002
	fossil	0.061 ± 0.038	0.033 ± 0.024	0.941 ± 0.017	0.136 ± 0.089
	smf	0.239 ± 0.016	0.159 ± 0.012	0.036 ± 0.002	0.535 ± 0.051
	gru4rec	0.682 ± 0.003	0.525 ± 0.002	0.909 ± 0.006	0.042 ± 0.002
	narm	0.719 ± 0.004	0.548 ± 0.004	0.928 ± 0.007	0.080 ± 0.006
	stamp	0.688 ± 0.007	0.507 ± 0.010	0.845 ± 0.034	0.079 ± 0.004
	sgnn	0.722 ± 0.005	0.541 ± 0.004	0.824 ± 0.011	0.093 ± 0.007
Session-based with reminders	sr-reminders	0.665 ± 0.003	0.529 ± 0.002	0.747 ± 0.009	0.071 ± 0.006
	vsknn-reminders	0.750 ± 0.003	0.574 ± 0.001	0.771 ± 0.009	**0.086 ± 0.007**
	stan-reminders	**0.751 ± 0.002**	0.591 ± 0.001	0.769 ± 0.010	0.081 ± 0.006
	vstan-reminders	0.746 ± 0.003	**0.602 ± 0.001**	0.779 ± 0.009	0.067 ± 0.005
	gru4rec-reminders	0.682 ± 0.003	0.525 ± 0.002	0.909 ± 0.006	0.042 ± 0.002
	narm-reminders	0.718 ± 0.005	0.547 ± 0.005	**0.929 ± 0.010**	0.080 ± 0.005
Session-aware	hgru4rec	0.576 ± 0.009	**0.419 ± 0.006**	**0.876 ± 0.007**	0.048 ± 0.003
	ncfs	**0.615 ± 0.005**	0.400 ± 0.002	0.741 ± 0.005	0.081 ± 0.006
	nsar	0.550 ± 0.014	0.379 ± 0.014	0.853 ± 0.015	0.076 ± 0.005
	shan	0.313 ± 0.012	0.183 ± 0.008	0.342 ± 0.027	**0.165 ± 0.014**

useful for Jusbrasil, but that more experiments are necessary, in particular tuning the reminders parameters, to see some improvements in the recommender systems.

For the session-aware paradigm, we can see that 3 out of 4 neural network session-aware models presented results that place them in the middle of the table. However, it is worth to mention that the models in this paradigm do not overcome the *neural network* models in the session-based paradigm.

Finally, taking the **ar** model (which is a proxy for the current recommender system in Jusbrasil) as a reference, we can see that the **stan** model provides a HitRate value 12% higher, and that the **vstan** model reached a MRR value 19% higher than the **ar** model. With respect to Coverage, we can see that the models mostly cover between 70% and 90% of the items. In terms of Popularity bias, the values are low for mostly of the models (i.e. less than 8%), which means that they are not focusing on recommending the most popular items from the dataset.

4 Final Remarks

In this work, we presented a benchmark of several session-based, session-based with reminders and session-aware recommender systems that can be used to improve legal document recommendation in Jusbrasil. The benchmark was focused on the logged users, and the results showed that some recommender systems can achieve gains of accuracy of around 19% with respect to the current recommender system adopted by Jusbrasil. Thus, this benchmark can be seen as a guideline for the next generation of legal document recommender systems in Jusbrasil.

Acknowledgments

To Jusbrasil (https://www.jusbrasil.com.br), for the financial support.

References

[1] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. 1993. Mining Association Rules between Sets of Items in Large Databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data* (Washington, D.C., USA) *(SIGMOD '93)*. Association for Computing Machinery, New York, NY, USA, 207–216. https://doi.org/10.1145/170035.170072

[2] Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava. 1999. Data Preparation for Mining World Wide Web Browsing Patterns. *Knowl. Inf. Syst.* 1, 1 (1999), 5–32. https://doi.org/10.1007/BF03325089

[3] Diksha Garg, Priyanka Gupta, Pankaj Malhotra, Lovekesh Vig, and Gautam Shroff. 2019. Sequence and Time Aware Neighborhood for Session-Based Recommendations: STAN. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) *(SIGIR'19)*. Association for Computing Machinery, New York, NY, USA, 1069–1072. https://doi.org/10.1145/3331184.3331322

[4] Ruining He and Julian McAuley. 2016. Fusing Similarity Models with Markov Chains for Sparse Sequential Recommendation. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*. 191–200. https://doi.org/10.1109/ICDM.2016.0030

[5] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based Recommendations with Recurrent Neural Networks. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). http://arxiv.org/abs/1511.06939

[6] Balázs Hidasi, Massimo Quadrana, Alexandros Karatzoglou, and Domonkos Tikk. 2016. Parallel Recurrent Neural Network Architectures for Feature-Rich Session-Based Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) *(RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 241–248. https://doi.org/10.1145/2959100.2959167

[7] Liang Hu, Qingkui Chen, Haiyan Zhao, Songlei Jian, Longbing Cao, and Jian Cao. 2018. Neural Cross-Session Filtering: Next-Item Prediction Under Intra- and Inter-Session Context. *IEEE Intelligent Systems* 33, 6 (2018), 57–67. https://doi.org/10.1109/MIS.2018.2881516

[8] Dietmar Jannach, Lukas Lerche, Iman Kamehkhosh, and Michael Jugovac. 2015. What Recommenders Recommend: An Analysis of Recommendation Biases and Possible Countermeasures. *User Modeling and User-Adapted Interaction* 25, 5 (2015), 427–491. https://doi.org/10.1007/s11257-015-9165-3

[9] Dietmar Jannach and Malte Ludewig. 2017. When Recurrent Neural Networks Meet the Neighborhood for Session-Based Recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (Como, Italy) *(RecSys '17)*. Association for Computing Machinery, New York, NY, USA, 306–310. https://doi.org/10.1145/3109859.3109872

[10] Dietmar Jannach, Malte Ludewig, and Lukas Lerche. 2017. Session-based item recommendation in e-commerce: on short-term intents, reminders, trends and discounts. *User Model. User Adapt. Interact.* 27, 3-5 (2017), 351–392. https://doi.org/10.1007/s11257-017-9194-1

[11] Santosh Kabbur, Xia Ning, and George Karypis. 2013. FISM: Factored Item Similarity Models for Top-N Recommender Systems. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Chicago, Illinois, USA) *(KDD '13)*. Association for Computing Machinery, New York, NY, USA, 659–667. https://doi.org/10.1145/2487575.2487589

[12] Iman Kamehkhosh, D. Jannach, and Malte Ludewig. 2017. A Comparison of Frequent Pattern Techniques and a Deep Learning Method for Session-Based Recommendation. In *RecTemp@RecSys*.

[13] Sara Latifi, Noemi Mauro, and Dietmar Jannach. 2021. Session-aware recommendation: A surprising quest for the state-of-the-art. *Information Sciences* 573 (2021), 291–315. https://doi.org/10.1016/j.ins.2021.05.048

[14] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. 2017. Neural Attentive Session-Based Recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (Singapore, Singapore) *(CIKM '17)*. Association for Computing Machinery, New York, NY, USA, 1419–1428. https://doi.org/10.1145/3132847.3132926

[15] Qiao Liu, Yifu Zeng, Refuoe Mokhosi, and Haibin Zhang. 2018. STAMP: Short-Term Attention/Memory Priority Model for Session-Based Recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (London, United Kingdom) *(KDD '18)*. Association for Computing Machinery, New York, NY, USA, 1831–1839. https://doi.org/10.1145/3219819.3219950

[16] Malte Ludewig and Dietmar Jannach. 2018. Evaluation of Session-Based Recommendation Algorithms. *User Modeling and User-Adapted Interaction* 28, 4–5 (2018), 331–390. https://doi.org/10.1007/s11257-018-9209-6

[17] Malte Ludewig, Noemi Mauro, Sara Latifi, and Dietmar Jannach. 2021. Empirical analysis of session-based recommendation algorithms. *User Model. User Adapt. Interact.* 31, 1 (2021), 149–181. https://doi.org/10.1007/s11257-020-09277-1

[18] Edleno Moura, Rafael Costa, Gabriel Jordão, and Gustavo Maia. 2021. Jusbrasil e os Desafios Tecnológicos para Facilitar e Aprimorar o Acesso à Justiça. In *Anais do XLVIII Seminário Integrado de Software e Hardware* (Evento Online). SBC, Porto Alegre, RS, Brasil, 207–213. https://doi.org/10.5753/semish.2021.15824

[19] J. R. Norris. 1997. *Markov Chains*. Cambridge University Press. https://doi.org/10.1017/CBO9780511810633

[20] Tu Minh Phuong, Tran Cong Thanh, and Ngo Xuan Bach. 2019. Neural Session-Aware Recommendation. *IEEE Access* 7 (2019), 86884–86896. https://doi.org/10.1109/ACCESS.2019.2926074

[21] Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. 2017. Personalizing Session-Based Recommendations with Hierarchical Recurrent Neural Networks. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (Como, Italy) *(RecSys '17)*. Association for Computing Machinery, New York, NY, USA, 130–137. https://doi.org/10.1145/3109859.3109896

[22] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing Personalized Markov Chains for Next-Basket Recommendation. In *Proceedings of the 19th International Conference on World Wide Web* (Raleigh, North Carolina, USA) *(WWW '10)*. Association for Computing Machinery, New York, NY, USA, 811–820. https://doi.org/10.1145/1772690.1772773

[23] Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor (Eds.). 2011. *Recommender Systems Handbook*. Springer. https://doi.org/10.1007/978-0-387-85820-3

[24] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-Based Recommendation with Graph Neural Networks. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence* (Honolulu, Hawaii, USA) *(AAAI'19/IAAI'19/EAAI'19)*. AAAI Press, Article 43, 8 pages. https://doi.org/10.1609/aaai.v33i01.3301346

[25] Haochao Ying, Fuzhen Zhuang, Fuzheng Zhang, Yanchi Liu, Guandong Xu, Xing Xie, Hui Xiong, and Jian Wu. 2018. Sequential Recommender System based on Hierarchical Attention Networks. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, 3926–3932. https://doi.org/10.24963/ijcai.2018/546