ON THE EFFECTIVENESS OF AN AI-DRIVEN EDUCATIONAL RESOURCE RECOMMENDATION SYSTEM FOR HIGHER EDUCATION

Johannes Schrumpf
Institute of Cognitive Science, University of Osnabrück
Wachsbleiche 27, 49074, Osnabrück, Germany

ABSTRACT

Digital resources offer a vast assortment of educational opportunities for students in higher education. From 2018 to 2022, a digital study assistant (DSA), named SIDDATA, was developed at three German universities and consequently field-tested. One of the DSA's features is an AI-driven natural language interface for educational resource recommendation. This paper performs an analysis of the effectiveness of recommendations, by analyzing data generated over the course of two years of DSA usage. We find that although initial user interest is high, only a small percentage of users engage with the recommendation feature. Furthermore, we find that quality of recommendations was perceived as mixed to negative.

KEYWORDS

Artificial Intelligence, Digital Study Assistant, Recommendation Engine, Higher Education, Evaluation

1. INTRODUCTION

The continuous effort of merging digital technologies into the domain of German higher education poses unique challenges and opportunities for students, lecturers and higher education institutions. Algorithms from the domain of Artificial Intelligence (AI) in particular, have been identified as vital for personalized education (Florea and Radu, 2019).

Within the scope of project SIDDATA (Studienindividualisierung durch digitale, datengestützte Assistenten, eng: *Study individualization through digital, data-driven assistants*), a digital study assistant (DSA) system was developed. The system aims at assisting students to discover, reflect upon and ultimately pursue their individual educational goals beyond established study program curricula. The system achieves this through a utilization of multiple, independent features, each of which providing assistance in a particular study matter and over different temporal scopes.

One feature for example provides the user with a visual aid to organize their personal learning goals in a hierarchical fashion, allowing for an increased understanding of personal goals. Another feature supports users in organizing and taking a semester abroad, enabling students to further individualize their higher education experience. These features can be individually activated or deactivated within a combined web-based platform, the digital study assistant. The feature of interest to this study is the so-called "professional interests" module, a Neural Network driven Natural Language Processing system. This feature allowed students to discover educational resources (Courses, single events in these courses such as single lectures, MOOCs and OERs) that match their interests, formulated in natural language. We have reported on the technological aspects utilized in the feature in a previous publication (Schrumpf *et al.*, 2021).

Over the course of two years of total run-time, the DSA underwent rigorous live-testing by being integrated into the learning management system (Stud.IP) of three German universities. Here, users were able to use the DSA in a prototype stage, all the while passively generating data on user-system interaction.

A previous study (Schurz *et al.*, 2021) investigated the effectiveness of features based on their activation/deactivation/non-interaction ratio. The study showed that the "professional interests" feature enjoyed a high popularity in terms of activation rate. However, these early results relied on a subset of data available today and only investigated feature activation ratios.

The present study seeks to investigate the effectiveness of the AI-driven "professional interests" feature within the DSA by expanding the scope of analysis to all data generated from two years of DSA testing. This data is analyzed for user-feature interaction on a finely-granular level, meaning that individual interactions between user and feature and taken into consideration. The information gathered from these individual interactions then serves as the basis for an assessment of feature utility. We close this study by highlighting key findings and by contextualizing our insights within literature.

2. DATA ANALYSIS

We analyze data from two datasets: The first dataset (P2) was generated over the run-time of the DSA's second prototype and is publicly available (Schrumpf *et al.*, 2022). The second dataset (P3) was generated over the run-time of the DSA's third and final prototype. The first prototype of the DSA generated no data and is therefore not considered in this study. Due to changes in the software architecture of the DSA between prototype 2 and 3, the datasets analyzed in this work partially contain different and incompatible information. We therefore highlight the source of data for every analysis whenever applicable.

To analyze the effectiveness of the "professional interests" feature, we first highlight activation/deactivation and non-interaction ratios between features. Next, we analyze how many interests were entered by users who activated the feature, giving insight into the actual usage of the feature. Finally, we analyze the effectiveness of resource recommendations by highlighting user feedback statistics and the number of resource recommendations that were accepted (clicked on) by users.

We begin our analysis with an overview of the "professional interests" feature use between the two datasets. When users engage with the DSA, they are able to activate features they are interested in. As reported in (Schurz et al., 2021), the "professional interests" feature enjoyed a high relative activation rate for prototype 2 of the SIDDATA DSA. Analyzing data from P3, we observe a similar phenomenon. Figure 1 shows the total number of activations per feature present in P3.

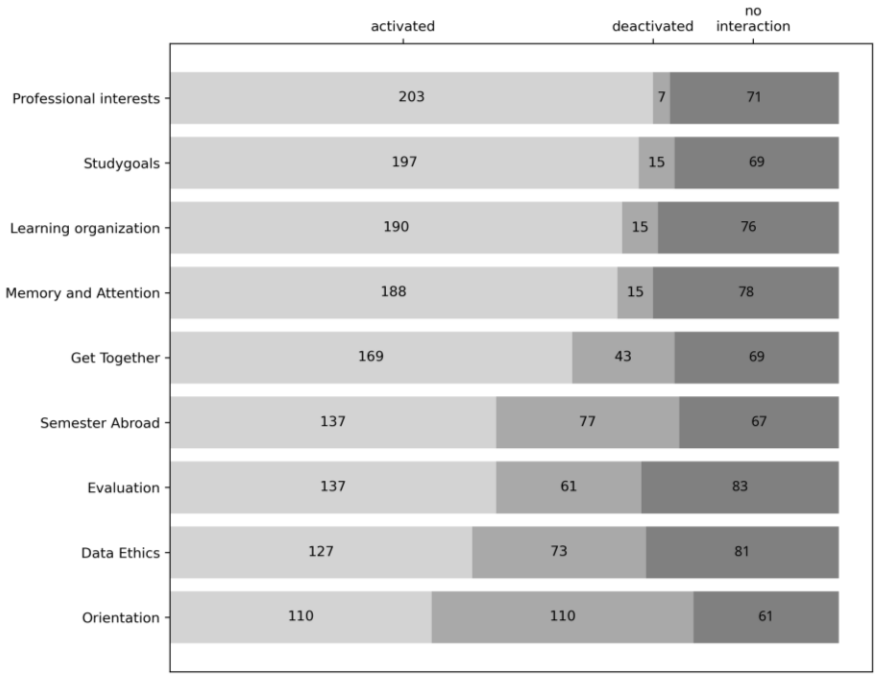

Figure 1. Activation, deactivation and no-interaction cases for P3

While non-interactions stay within the mid to high double-digit range, the "professional interests" feature exhibits a high relative activation rate for users in P3, while maintaining a low deactivation rate. Probing further into the user-feature interaction for the "professional interests" feature, we count the number of instances where a user entered at least one interest (query) and compare the results between P2 and P3.

Table 1 shows the number of users that activated the feature for their account and consequently interacted with it by writing at least one query.

Table 1. Total and relative user-feature interaction count for both P2 and P3 datasets

	P2 total	P2 percentile	P3 total	P3 percentile
DSA users	735	1.00	281	1.00
"Professional interests" feature activations	595	0.81	203	0.72
Users with at least one entered query	106	0.14	52	0.19

With an activation rate of 81% and 72% between prototypes, a high number of DSA users activated the feature for their account. However, only 14% and 19% respectively entered at least one query for educational resource recommendation. Among such users who entered at least one query, we further investigate the number of times users entered professional interests. Figure 2 shows the number of queries entered per number of users.

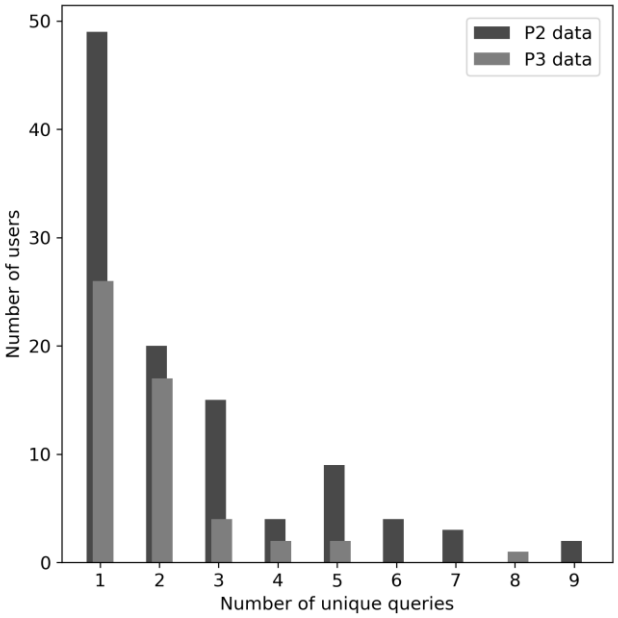

Figure 2. Number of users who entered n queries over the testing period. Data from the P2 dataset is depicted in dark while P3 data is depicted in lighter color

The majority of users enter one query. Even though P3 captured less user-feature interaction data, the number of users who entered two queries for P3 is only slightly lower than in P2.

We further analyze the perceived quality of recommendations generated by the system. For prototype 2, users were able to rate educational resource recommendations on a four-item scale (displayed as emojis), where 1 corresponds to with a low rating and 4 with a high rating. Of 1807 total recommendations generated, 86 (4,7%) were evaluated with the feedback function. We display the total and relative rating of recommendations for P2 in table 2.

Table 2. Total and relative ratings of recommended resources for P2. 1 corresponds to negative feedback (sad emoji), while 4 corresponds to positive feedback (smiling emoji)

	1	2	3	4
Number of rated recommendations	51	14	7	14
Percentile of rated recommendations	0.59	0.16	0.08	0.16

With 59%, the rating given to recommended resources are overwhelmingly negative. Only 16% of recommendations were rated positively, while 16% and 8% of rating were in between.

Between prototype 2 and 3, the rating function was removed and replaced by a passive logging of users clicking on a link associated with the recommended resources. We postulate that a click on a resource's associated link indicated the recommendation to be of sufficient interesting for users and hence a good recommendation (Oard and Kim, 1998). Of a total of 537 recommended resources, 36 (6.7%) were clicked on by users.

Another change between prototype 2 and 3 was the inclusion of external educational resources such as MOOCs, OERs and courses from external universities into the list of recommendable educational resources. Because an interaction between perceived usefulness of recommended resource and the type of resource cannot be ruled out, we show the number of times these resources were included in the search filter for queries relative to the number of times these resources were clicked on by users in table 3.

Table 3. Inclusion of educational resource type relative to number of clicks for resources of this type for P3

	Local courses	External courses	MOOC	OER
Number of inclusions in search queries	54	49	59	38
Number of clicks	14	8	4	10
Relative frequency	0.26	0.16	0.07	0.26

With 26%, courses and OERs were clicked on most frequently when a click occurred and when they were included in the filter options. Even though MOOCs were included most frequently in the query filtering options, they were clicked on the least.

3. DISCUSSION & CONCLUSION

With a high activation and low deactivation rate between P2 and P3, the "professional interests" feature of the SIDDATA DSA appears to have attracted a large amount of initial interest from users. However, as our analysis shows, the number of users engaging with the feature after initial activation is low. Among those users who entered a query, only a small percentage does so for at least two queries, indicating that after a novelty effect, the usefulness of the feature is not perceived as sufficiently high to warrant further interaction. One possible explanation for this is found in ratings and clicks on recommended resources respectively: Even though only a small percentage of users rated recommended resources, the gathered feedback is overwhelmingly negative. While a bias towards giving a rating if the presented resource did not fit the interest cannot be ruled out, these results suggest that users were on average not satisfied with resource recommendations. This is reflected in the low number of clicks on recommended resources from our P3 data. A small success can be seen in the changes implemented between prototype 2 and 3 of the DSA, leading to MOOCs and OERs being added to the list of recommendable resources: Even though click-rates remained low, at least a small number of these resources were clicked on by users, making OER and MOOCs more accessible to students.

With the analysis being the results of data gathered between two prototype iterations, we ultimately conclude that the application of a natural language processing based educational resource recommendation system as a feature in a digital study assistant software did not contribute to self-regulated and self-determined learning in the ways we hoped it to: Low rating scores and small numbers of clicks on recommended resources suggest poor recommendation performance of the underlying AI technology. While a small number of recommendations were rated positively in the P2 dataset, and a small number of educational resources were clicked on in the P3 dataset, most item recommendations received negative ratings or were ignored by users. A low user engagement with the feature may be the result of a low recommendation performance by the underlying AI technology. Compared to similar proposed systems such as MCRS (Zhang *et al.*, 2018), our system appears to fall behind in terms of recommendation fidelity. This suggests that our approach of implementation falls short in key requirements of recommendation systems.

A future qualitative assessment of the difference in technology and data availability between both approaches therefore may shed insight into the technological reasons for the perceived poor recommendation performance of our system.

In parallel, future studies will have to investigate the impact of overall DSA utility perception on single features: With a high activation rate and low user engagement, we hypothesize that this effect may be at least partially derivable from a low perceived usefulness of the DSA system as a whole. Future studies will need to investigate this hypothesize by extending the scope of analysis beyond the "professional interests" feature.

REFERENCES

Florea, A.M. and Radu, S. (2019), "Artificial Intelligence and Education", *2019 22nd International Conference on Control Systems and Computer Science (CSCS)*, pp. 381–382.

Oard, D.W. and Kim, J. (1998), "Implicit Feedback for Recommender Systems", *Proceedings of the AAAI Workshop on Recommender Systems*, pp. 81–83.

Schrumpf, J., Weber, F., Schurz, K., Dettmer, N. and Thelen, T. (2022), "A Free and Open Dataset from a Prototypical Data-driven Study Assistant in Higher Education", *Proceedings of the 14th International Conference on Computer Supported Education*, pp. 155–162.

Schrumpf, J., Weber, F. and Thelen, T. (2021), "A Neural Natural Language Processing System for Educational Resource Knowledge Domain Classification", in Kienle, A., Harrer, A., Haake, J.M. and Lingnau, A. (Eds.), *DELFI 2021*, Gesellschaft für Informatik e.V., Bonn, pp. 283–288.

Schurz, K., Schrumpf, J., Weber, F., Seyfeli, F. and Wannemacher, K. (2021), "Towards a User Focused Development of a digital Study Assistant Through a Mixed Methods Design", in Sampson, D.G., Ifenthaler, D. and Isaías, P. (Eds.), *18th International Conference on Cognition and Exploratory Learning in the Digital Age, CELDA 2021*, IADIS Press, pp. 45–52.

Zhang, H., Huang, T., Lv, Z., Liu, S.Y. and Zhou, Z. (2018), "MCRS: A course recommendation system for MOOCs", *Multimedia Tools and Applications*, Multimedia Tools and Applications, Vol. 77 No. 6, pp. 7051–7069.