

Classifeye: Classification of Personal Characteristics Based on Eye Tracking Data in a Recommender System Interface

Martijn Millecamp^a, Cristina Conati^b and Katrien Verbert^a

^aDepartment of computer science, KU Leuven, Celestijnenlaan 200A bus 2402, Leuven, Belgium

^bDepartment of computer Science, ICICS/CS 107, 2366 Main Mall, Vancouver, BC, Canada

Abstract

Due to the increasing importance of recommender systems in our life, the call to make these systems more transparent becomes louder. However, providing explanations is not as easy as it seems, as research has shown that different users have varying reactions to explanations. So not only the recommendations, but also the explanations should be personalised. As a first step towards these personalised explanations, we explore the possibility to classify users based on their gaze pattern during the interaction with a music recommender system. More specifically, we classify three personal characteristics that have been shown to play a role in the interaction with music recommendations: need for cognition, openness and musical sophistication. Our results show that classification based on eye tracking has potential for need for cognition and openness, as we are able to do better than random, but not for musical sophistication as no classifier did better than a uniform random baseline.

Keywords

eye tracking, classification, recommender system, openness, need for cognition, musical sophistication

1. Introduction

In the field of recommender systems (RS), researchers are increasingly aware that optimizing accuracy is not enough to reach the full potential of recommender systems (RS) [1, 2]. For example, users will not choose a recommended item unless they have trust in the system [3]. One possible way to increase this trust is providing explanations which reveal (a part of) the internal reasoning of the

RS to the user [4, 5]. Especially the combination of these explanations with control can help users not only to understand the RS, but also to steer the RS with input and feedback [5]. Despite the increased interest in explanations for RS, it is still not clear how to implement explanations in practice as users have varying reactions to them which shows the need to personalize explanation to the user [6].

However, before the system could adapt explanations to personal characteristics (PCs), it needs to be aware of the PCs of the user. A possible way to obtain these characteristics is by explicitly asking the users to fill in questionnaires [7] or by implicitly inferring PCs through an analysis of the social media of the user [8]. Nonetheless, asking users to fill in questionnaires or to give access to their social media is often not desirable. Moreover, to personalize explanations it is not necessary

HUMANIZE: Joint Proceedings of the ACM IUI 2021 Workshops, April 13–17, 2021, College Station, USA

✉ martijn.millecamp@kuleuven.be (M. Millecamp);

conati@cs.ubc.ca (C. Conati);

katrien.verbert@kuleuven.be (K. Verbert)

🆔 0000-0002-5542-0067 (M. Millecamp);

0000-0002-8434-9335 (C. Conati); 0000-0001-6699-7710

(K. Verbert)



© 2021 Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings
(CEUR-WS.org)

to obtain a fine-grained result, but a classification into two categories suffices [9].

For this reason, we explore in this paper whether it is possible to classify users' personality traits during the interaction with a music RS with explanations by analyzing their gaze. We will focus on three different PCs: openness, need for cognition (NFC) and musical sophistication (MS) [9, 10]. These PCs will be explained in detail in Section 2.

Openness is one of the Big Five personality traits which measures how open a person is to new experiences. Millicamp et al. [9] showed that there was a significant difference in the gaze pattern between low and high openness users. This is the reason we hypothesize that classifying openness based on gaze might be possible.

Similarly, we hypothesize that inferring MS, which is a measure of domain knowledge in the music domain, from gaze data might be possible as the study of Millicamp et al. [9] also found significant differences in gaze pattern between low and high MS.

NFC is a cognitive style which influences the way a person prefers to process information and thus looks at information. Previous studies already showed that NFC moderates the perception of explanations in a music recommender system, which was the motivation to explore whether inferring NFC from gaze would be possible.

Next to exploring the general accuracy, we also want to explore how much data we need to infer these PCs.

The contribution of this paper is twofold. First, to our knowledge, we are the first to explore whether it is possible to infer PCs during the interaction with a RS in the presence of explanations. Second, we make the gathered dataset publicly available to support the research in this area. This dataset is unique because it provides both gaze data and data about PCs.

2. Related work

With the increasing role of RS in our daily lives, the call for explainable, transparent RS also becomes louder so that users can make better informed decisions whether or not to follow the recommendations [11, 6]. In combination with controls, this transparency also enables users to correct the RS whenever they feel it makes wrong assumptions [5]. However, research has shown that different users have different reactions to explanations [6, 12, 13]. In the field of music RS, recent research has shown that there are three PCs that could influence the way users perceive explanations: openness, NFC and MS [10, 9].

Openness is one of the five factors of the Five Factor Model, also known as the Big 5 model [14]. This model describes personality in five different traits and it has been used in several studies which showed the positive impact of considering personality in RS [15]. The factor openness describes *the breadth, depth and complexity of an individual's mental and experiential life* [16]. It has been shown that openness is related to the preferred amount of diversity in RS and to the willingness to use a system with explanations [17, 18, 9].

Need for cognition has been shown to influence the success of a RS [13, 12, 19, 20] and is defined as *"a measure of the tendency for an individual to engage in, and enjoy, effortful cognitive activities"* [21]. NFC has been shown to have an impact on the willingness of users to rely on a RS [12], on the confidence in a playlist created in a music RS with explanations [10], on preference matching [22], on the style of explanations they prefer [13] and on the reason why users need a transparent RS [23].

Musical sophistication is defined by Mülensiefen et al. [24] as a concept to describe the multi-faceted nature of musical expertise. In the music domain, Millicamp et al. [9] showed that users with high MS feel more

supported to make a decision in a RS interface that provided explanations than an interface without such explanations, while this made no difference for users with low MS. Another study showed that users with high domain experience perceive a higher diversity in a scatter plot than in a simpler bubble chart [25].

To acquire the PCs of users, the most common way is to ask users to fill in validated questionnaire [7], but there exist also other approaches such as inferring PCs by analyzing the social media of the user [26, 8], by analyzing a conversation with a chatbot [27] or by analyzing the physical signals such as brain activity [28] and gaze data [7].

The previously mentioned works rely on fine-grained personality scores. In contrast, in our work we focus on adapting interfaces to users for which we only need a classification in two groups. We aim to base this classification on the gaze pattern during the interaction with a music RS interface instead of asking users to watch carefully selected stimuli, to fill in questionnaires or to share their social media profile. Previous studies which classified users based on their gaze pattern during normal activities are almost all only focused on cognitive abilities and visualization experience [29, 30, 31, 32]. One exception is the study of Hoppe et al. [33] which inferred the Big Five personality traits by studying the gaze of a walk through a campus. This study is different from our work, as we investigate if it is possible to infer PCs while interacting with a music RS and also focus on different PCs.

3. Data

The gaze data that is used in this study was generated in a user study by Millecamp et al. [9]. We will provide a brief summary of this experiment, but a more elaborate description

Table 1

An overview of personal characteristics measured, together with their highest and lowest possible scores and summary statistics for the scores of the participants

PC	Possible Range	Median Score
Age	18-65	24
MS	18-126	64
NFC	0-100	68.75
Openness	0-100	55

can be found in [9]. As mentioned in Section 2, we focus in this study on openness, NFC and MS as previous research has found that these PCs could affect the perception of explanations in a music RS [9, 10] and the study of Millecamp et al. [9] already showed that openness and MS change the gaze pattern between an interface in the presence or absence of explanations. To measure these three characteristics, users were asked to fill out three questionnaires before the experiment started. To measure openness, we used the 44-item Big Five Inventory [34] and selected afterwards the questions related to openness. For NFC, we used the 18-items questionnaire of Cacioppo et al. [21] and for MS the Goldsmiths Musical Sophistication Index¹ was used. The dataset we used in this study consists of the gaze data of 30 participants (21 male). For the three PCs, the participants were divided into a high and low group based on a median split. This resulted in equally distributed groups for MS and NFC and almost equally groups for openness (16 in the low and 14 in the high openness group). An short overview of the characteristics of the participants can be found in Table 1.

The gaze data was recorded with a Tobii 4C remote eye tracker at a sampling rate of 90Hz. Each sample contained information about

¹<https://www.gold.ac.uk/music-mind-brain/gold-msi/> May 2020

the focus point on the screen denoted as an x and y coordinate, the distance between the participant and the screen, and the validity of these measures. To calibrate the eye tracker, the experiment started with a standard calibration procedure provided by Tobii Core Software. After the calibration, users were asked to explore the interface of a music RS in the presence of feature-based explanations until they understood all functionalities. A screenshot of the interface is shown in Figure 1.

As shown in Part A of this figure, users first can search for an artist they like through a search bar in the top left corner. When they add the artist, this artist is shown in Part B. Based on this artist, the system starts to generate recommendations which were listed in a two-column format as shown in Part F. When users hover over the cover of the picture of a recommended song, they can click a play button to listen to a 30s preview of the song. On the right side of each explanation, they can click on the thumb-up icon to add the song to their playlist. Through the sliders shown in Part D of Figure 1, users can modify several audio features² such as popularity, energy and danceability which are also taken into account in the recommendation process. To help users steer these sliders, the minimum and the maximum for each audio feature is shown for each artist.

After the user explored all the options of the interface, the recording of the gaze started. As shown in Part E of Figure 1, users were asked to create a playlist of five songs. To create this playlist, they could use all functionalities without any restriction. When they added the fifth song to their playlist, we stopped the recording of the gaze. On average, users took 4 minutes 26 seconds to complete their playlist. As part of this paper's contribution, this data is publicly available³.

²<https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/>

³augment.cs.kuleuven.be/datasets/classifeye

4. Classifiers

4.1. Features

The Tobii 4C does not come with software to detect fixations and saccades so we identified fixations and saccades using an implementation of the ID-T algorithm [35] with a dispersion threshold of one degree and a duration threshold of 100ms [35]. This means that in this study a fixation is identified as a circle on the screen in which the user keeps focusing for at least 100ms without moving their eyes more than one degree. All other movements are then identified as saccades, i.e. quick movements of gaze from one fixation to another [30].

Based on these saccades and fixations, we generated a set of eye-tracking features as listed in Table 2. Most of these features are selected because they are widely used in previous eye tracking studies [7, 30, 36]. In addition to these features, we included *Most frequent saccade direction* and fixations in a *4x4 heatmap* as the study of Hoppe et al. [33] indicated that these features are important in the extraction of personality. We did not include features that contain explicit information about the content of the interface, so called areas of interest (AOI) even as previous work has shown that these features could have more predictive power [30]. The reason for this is that this information is already partially captured in a more general way by *Most frequent saccade direction* and fixations in a *4x4 heatmap*. Thus, at this stage we chose to investigate how far we can go with display-independent features, which also have the advantage of possibly being more generalizable to other interfaces.

4.2. Data windows

To explore whether classification of the three PCs would be possible with only a partial

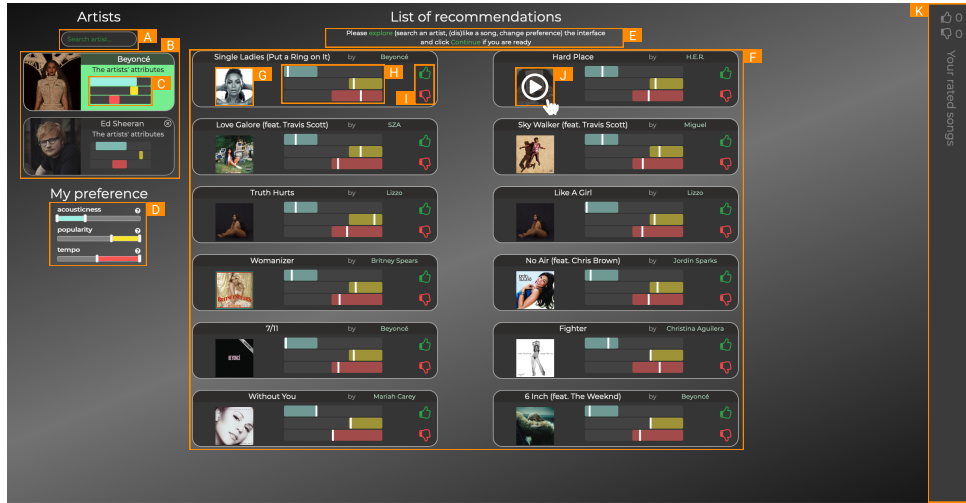


Figure 1: The interface with the different parts highlighted in orange. A: Searchbox, B: Artist, C: Attributes of the artist, D: Preference of the user, E: Task, F: Recommendations, G: Cover of a song, H: Explanations I: (dis)like buttons, J: Play button K: list of (dis)liked songs

Table 2
Description of eye tracking features

Features	Description
Saccade rate	Number of saccades divided by segment duration
Avg. saccade length	Average distance between the two fixations delimiting the saccade
Avg. saccade amplitude	Average size of saccade in degrees of visual angle
Avg. saccade velocity	Average velocity (saccade amplitude / saccade duration) of saccades
Peak saccade velocity	Maximum saccade velocity in segment
Most frequent saccade direction	Most frequent saccade direction (segments of 45°)
Fixation rate	Number of fixations divided by segment duration
Avg. fixation duration	Average duration of fixation in ms
Ratio Fixations/Saccades	Ratio of total nb of fixations divided by total nb of saccades
4x4 Heatmap	Percentage of fixations in 16 raster areas
Avg. pupil size	Average pupil size of both eyes

amount of data, we generated three different data windows to simulate partial observations of gaze data during the task similar to Steichen et al. [30] and Conati et al. [31]. Each window consists of a partial observation of each participant based on relative duration: the first window consisted of the first 30% of data, the second window of the first

60% and the last window consisted of the first 90% of data. Despite the fact that this approach requires a task to be fully completed to determine what 100% of the data constitutes, it still allows to provide valuable insights into trends and patterns about inferring PCs from gaze data [30]. Each of these windows consist of three different measure-

ments and for each of these measurements the data was divided in ten different segments of equal length. For each of these segments, we generated the mentioned set of eye-tracking features resulting in a feature vector of 260 features for each measurement.

The reasoning behind creating these different datasets is to verify whether we would be able to adapt the RS interface to the needs of the user during the task. As such we did not include a window with 100% of the data as the adaptation would be too late. Additionally, previous research [30] showed already that after a certain amount of data, the accuracy started to converge or even that the accuracy decreases after a certain amount of data. In this study, we want to explore whether we would notice similar trends for different PCs.

4.3. Classification methods

To classify users in a low and high category, we used *scikit-learn* to train five different classifiers and a baseline [37]. To evaluate the performance of the classifiers, we applied a leave-one-out methodology. Because of this evaluation methodology and the uniform groups, we could not use the most common majority class baseline which predicts the most likely class (this would lead to 0% accuracy) [30, 31, 33]. As a consequence, we choose a random uniform baseline which has a theoretical accuracy of 50%. To classify the characteristics, we trained Logistic Regression, Random Forest, Gaussian Naive Bayes, Linear Support Vector Machines and Gradient Boosting. The reasoning behind the implementation of all these classifiers is that in previous research there is no consensus about which classifiers work the best. Steichen et al. [30] found that Logistic Regression performed better than Decision Trees, Support Vector Machines and Neural Networks. Lallé et al. [38] and Hoppe et al. [33] found

Table 3

Description of parameters of the different classifiers

Classifier	Parameter
Baseline	strategy: uniform
Logistic Regression	solver: liblinear
Random Forest	estimators: 100
Gaussian Naive Bayes	na
Linear Support Vector Machines	gamma: scale probability: True
Gradient Boosting	maximum depth: 4

that Random forests worked the best. However, Berkovsky et al. [7] conclude that Naive Bayes and Support Vector Machines are the best. Additionally, Gradient Boosting performed well in the study of Barral et al. [39]. Because of the small sample size, we chose not to use deep learning methods. For each of these classifiers we tried to optimize the accuracy. The resulting parameters can be found in Table 3.

To strengthen the stability of the results, we ran this evaluation 10 times with different random seeds. We calculated the average accuracy over all participants, and all runs to measure performance of the classifier.

5. Results

To examine whether it is possible to classify users in the correct personality group and whether this classification works better on specific windows, we ran for each PC a two-way repeated measures ANOVA with accuracy as the dependent variable and both classifier and window as independent variables. As we run multiple ANOVA's and pairwise comparisons, the reported p-values are adjusted using the Benjamini and Hoghberg procedure [40] to control for the family-wise false discovery rate. The main results of this analysis are shown in Figure 2 and we will report the results for

each of the PCs in detail in the next paragraphs.

Need for cognition. The results of the two-way repeated measures ANOVA revealed a significant main effect of classifier on accuracy ($F(7,14) = 18.8, p < .001$). To investigate this main effect, we ran post-hoc pairwise comparisons which showed that the mean accuracy of the logistic regression classifier (0.59) performed statistically better than the baseline ($p = .0491$) which is shown in Figure 2a. This figure also shows the accuracy in the three different windows and that the peak accuracy (0.67) is reached in the last window.

Musical sophistication. The results of the two-way repeated measures ANOVA revealed that no classifier could outperform the baseline and that most of the classifiers performed even worse.

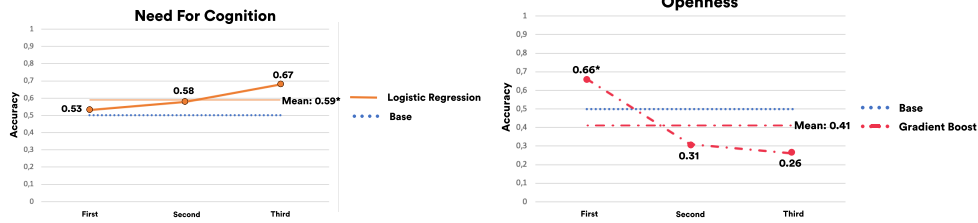
Openness. The results of the two-way repeated measures ANOVA revealed a significant interaction effect of classifier with window on accuracy ($F(14,28) = 4.88, p < .001$). An analysis of the effect of classifier showed a significant effect for the first window ($F(7,16) = 4.512, p = .006$) and a post-hoc test revealed that in this window the Gradient Boost performed significantly better than the baseline ($p = .020$). The analysis of the effect of window showed a significant effect for the Gradient Boost classifier ($F(2,6) = 8.12, p = .020$) and a post-hoc analysis showed that the gradient boost classifier performed significantly better in the first window than in the second ($p = .028$) and the third window ($p = .029$). Figure 2b shows that the highest accuracy of Gradient Boost is reached in the first window (0.66). This accuracy is significantly higher than the accuracy of the baseline and the accuracy of Gradient Boost in the other windows.

6. Discussion

Our results show that we have a higher accuracy than the random baseline for NFC and for openness in the first window, but that we were not able to do beat the random baseline classifier for MS.

For the classification of openness, it is interesting that we are able to outperform the baseline while openness was one of the few traits of which Hoppe et al. [33] could not outperform the baseline. This might be due to a different classification technique as Hoppe et al. only used a Random Forest classifier while we outperformed the baseline with a Gradient Boost classifier. Another possible reason could be that this difference is due to the fact that we trained the classifiers on different data windows and that our results show that the performance to classify openness is only significantly better than the baseline in the first window. As far as we know, no other studies formally showed that classifying PCs on early stages of the task can outperform classifiers trained on more data. However, other studies such as the study of Steichen et al. [30] already discussed this trend for perceptual speed, verbal working memory and visual working memory. They argued that these characteristics most strongly affect the gaze pattern of the user during the initial phase of a task and that other factors dilute the gaze pattern as the task continues. This is probably also the reason why we are only able to classify openness in the beginning of the task. However, this is not necessary a problem as we want to adapt an interface to the openness of a user as early on as possible. Nevertheless, the obtained accuracy is still too low to be used to adapt the explanations. Also, more research is needed to verify that openness will always affect the gaze during the beginning of a task or only when they see a new interface.

To classify NFC, our results show a signif-



(a) Accuracy of Logistic Regression for NFC.

(b) Accuracy of Gradient Boost for openness.

Figure 2: Accuracy of classifiers that perform significantly better than the baseline.

icant main effect of Logistic Regression on accuracy. The reason that we do not see a significant difference between the windows could be that NFC is correlated with decision-making processes [12] and creating a playlist in a music RS constantly involves making decisions. Despite the significant main effect, the accuracy to classify NFC seems not high enough to adapt the interface, especially not in the first two windows. As a consequence, this means that further research needs to focus on reaching a higher accuracy in the beginning of the interaction to be able to adapt explanations early on in the process or on adapting the interface if the user re-visits the application. Additionally, further research should investigate why Logistic Regression performed the best to classify NFC as this is similar to previous studies in which Logistic Regression performed well to classify PCs, but we do not have an explanation why logistic regression outperforms other algorithms [30, 41].

As a previous study in the field of music RS showed that MS influences the way users look to a music RS interface and previous studies in the field of information retrieval also showed the potential of predicting domain knowledge based on eye tracking [42, 43, 9], we expected to be able to classify MS based on gaze data. However, our results show that we could not outperform the baseline. A pos-

sible reason for this could be that we did not include AOI related features which were included in the above-mentioned studies. An interesting further line of research is to verify whether including these AOI features can improve accuracy.

7. Conclusion

In this paper, we explored whether it would be possible to adapt the explanations in a music RS interface based on personal characteristics. To do so, we investigated whether a classification of personal characteristics could be inferred by studying the gaze pattern during the creation of a playlist in this system. More concretely, we classified musical sophistication, need for cognition and openness because these characteristics have shown to impact the user experience of explanation in a RS [9]. We trained the classifiers on different windows to detect whether the classification would already work with only a partial observation of the creation of a playlist.

Our results show that even as our accuracy is not yet high enough for practical use, we are able to outperform a baseline to classify need for cognition with Logistic Regression. If we only consider the first third of the data, our results show that the classification of openness with Gradient Boost beats

the baseline. Despite the limitations in terms of accuracy, this finding is important because it shows the potential to adapt explanations during the interaction with a music RS interface. In a next step, we want to increase the accuracy of the classifiers particularly in the beginning of the interaction which we plan to do by gathering more training data and by using different features such as AOI related features. Additionally, more research is needed to verify whether the results of this study could be generalized to different tasks and interfaces which we also plan to address in future research.

Acknowledgments

Part of this research has been supported by the KU Leuven Research Council (grant agreement C24/16/017) and the Research Foundation Flanders (FWO).

References

- [1] R. R. Sinha, K. Swearingen, et al., Comparing recommendations made by online systems and friends., *DELOS 106* (2001).
- [2] C. He, D. Parra, K. Verbert, Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities, *Expert Systems with Applications* 56 (2016) 9–27.
- [3] J. Kunkel, T. Donkers, L. Michael, C.-M. Barbu, J. Ziegler, Let me explain: Impact of personal and impersonal explanations on trust in recommender systems, in: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–12.
- [4] A. Springer, S. Whittaker, Making transparency clear, in: *Algorithmic Transparency for Emerging Technologies Workshop*, 2019, p. 5.
- [5] N. Tintarev, J. Masthoff, A survey of explanations in recommender systems, in: *2007 IEEE 23rd international conference on data engineering workshop*, IEEE, 2007, pp. 801–810.
- [6] A. Springer, S. Whittaker, Progressive disclosure: empirically motivated approaches to designing effective transparency, in: *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 2019, pp. 107–120.
- [7] S. Berkovsky, R. Taib, I. Koprinska, E. Wang, Y. Zeng, J. Li, S. Kleitman, Detecting personality traits using eye-tracking data, in: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–12.
- [8] G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, M. Kosinski, D. J. Stillwell, L. H. Ungar, M. E. Seligman, Automatic personality assessment through social media language., *Journal of personality and social psychology* 108 (2015) 934.
- [9] M. Millecamp, N. N. Htun, C. Conati, K. Verbert, What’s in a user? towards personalising transparency for music recommender interfaces, in: *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 2020, pp. 173–182.
- [10] M. Millecamp, N. N. Htun, C. Conati, K. Verbert, To explain or not to explain: the effects of personal characteristics when explaining music recommendations, in: *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 2019, pp. 397–407.
- [11] M. Naiseh, N. Jiang, J. Ma, R. Ali, Explainable recommendations in intelligent systems: delivery methods, modalities and risks, in: *International Conference on Research Challenges in Infor-*

- mation Science, Springer, 2020, pp. 212–228.
- [12] S. T. Tong, E. F. Corriero, R. G. Matheny, J. T. Hancock, Online daters’ willingness to use recommender technology for mate selection decisions., in: *InRS@ RecSys*, 2018, pp. 45–52.
- [13] S. Naveed, T. Donkers, J. Ziegler, Argumentation-based explanations in recommender systems: Conceptual framework and empirical results, in: *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, 2018, pp. 293–298.
- [14] L. R. Goldberg, The structure of phenotypic personality traits., *American psychologist* 48 (1993) 26.
- [15] R. Hu, P. Pu, Enhancing collaborative filtering systems with personality information, in: *Proceedings of the fifth ACM conference on Recommender systems*, 2011, pp. 197–204.
- [16] V. Benet-Martinez, O. P. John, Los cinco grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the big five in spanish and english., *Journal of personality and social psychology* 75 (1998) 729.
- [17] N. Tintarev, M. Dennis, J. Masthoff, Adapting recommendation diversity to openness to experience: A study of human behaviour, in: *International Conference on User Modeling, Adaptation, and Personalization*, Springer, 2013, pp. 190–202.
- [18] L. Chen, W. Wu, L. He, How personality influences users’ needs for recommendation diversity?, in: *CHI’13 extended abstracts on human factors in computing systems*, 2013, pp. 829–834.
- [19] U. Gretzel, D. R. Fesenmaier, Persuasion in recommender systems, *International Journal of Electronic Commerce* 11 (2006) 81–100.
- [20] A. Felfernig, R. Burke, Constraint-based recommender systems: technologies and research issues, in: *Proceedings of the 10th international conference on Electronic commerce*, 2008, pp. 1–10.
- [21] J. T. Cacioppo, R. E. Petty, C. Feng Kao, The efficient assessment of need for cognition, *Journal of personality assessment* 48 (1984) 306–307.
- [22] K. Y. Tam, S. Y. Ho, Web personalization as a persuasion strategy: An elaboration likelihood model perspective, *Information systems research* 16 (2005) 271–291.
- [23] M. Milencamp, R. Haveneers, K. Verbert, Cogito ergo quid? the effect of cognitive style in a transparent mobile music recommender system, in: *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 2020, pp. 323–327.
- [24] D. Müllensiefen, B. Gingras, J. Musil, L. Stewart, The musicality of non-musicians: an index for assessing musical sophistication in the general population, *PloS one* 9 (2014) e89642.
- [25] Y. Jin, N. Tintarev, K. Verbert, Effects of individual traits on diversity-aware music recommender user interfaces, in: *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, 2018, pp. 291–299.
- [26] J. Golbeck, C. Robles, M. Edmondson, K. Turner, Predicting personality from twitter, in: *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, IEEE, 2011, pp. 149–156.
- [27] M. X. Zhou, G. Mark, J. Li, H. Yang, Trusting virtual agents: the effect of personality, *ACM Transactions on Interactive Intelligent Systems (TiiS)* 9 (2019) 1–36.
- [28] J. Wache, R. Subramanian, M. K. Abadi,

- R.-L. Vieriu, N. Sebe, S. Winkler, Implicit user-centric personality recognition based on physiological responses to emotional videos, in: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, 2015, pp. 239–246.
- [29] D. Toker, C. Conati, B. Steichen, G. Carenini, Individual user characteristics and information visualization: connecting the dots through eye tracking, in: proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2013, pp. 295–304.
- [30] B. Steichen, C. Conati, G. Carenini, Inferring visualization task properties, user performance, and user cognitive abilities from eye gaze data, *ACM Transactions on Interactive Intelligent Systems (TiIS)* 4 (2014) 1–29.
- [31] C. Conati, S. Lallé, A. Rahman, D. Toker, Further results on predicting cognitive abilities for adaptive visualizations, *IJCAI International Joint Conference on Artificial Intelligence (2017)* 1568–1574. doi:10.24963/ijcai.2017/217.
- [32] M. Gingerich, C. Conati, Constructing models of user and task characteristics from eye gaze data for user-adaptive information highlighting, in: Proceedings of the AAAI Conference on Artificial Intelligence, 1, 2015.
- [33] S. Hoppe, T. Loetscher, S. A. Morey, A. Bulling, Eye movements during everyday behavior predict personality traits, *Frontiers in Human Neuroscience* 12 (2018) 1–8. doi:10.3389/fnhum.2018.00105.
- [34] O. P. John, E. M. Donahue, R. L. Kentle, The big five inventory—versions 4a and 54, 1991.
- [35] D. D. Salvucci, J. H. Goldberg, Identifying fixations and saccades in eye-tracking protocols, in: Proceedings of the 2000 symposium on Eye tracking research & applications, ACM, 2000, pp. 71–78.
- [36] J. H. Goldberg, X. P. Kotval, Computer interface evaluation using eye movements: methods and constructs, *International journal of industrial ergonomics* 24 (1999) 631–645.
- [37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* 12 (2011) 2825–2830.
- [38] S. Lallé, C. Conati, G. Carenini, Prediction of individual learning curves across information visualizations, *User Modeling and User-Adapted Interaction* 26 (2016) 307–345.
- [39] O. Barral, S. Lallé, G. Guz, A. Iranpour, C. Conati, Eye-tracking to predict user cognitive abilities and performance for user-adaptive narrative visualizations, in: Proceedings of the 2020 International Conference on Multimodal Interaction, 2020, pp. 163–173.
- [40] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal statistical society: series B (Methodological)* 57 (1995) 289–300.
- [41] S. Kardan, C. Conati, Exploring gaze data for determining user learning with an interactive simulation, in: *International Conference on User Modeling, Adaptation, and Personalization*, Springer, 2012, pp. 126–138.
- [42] M. J. Cole, J. Gwizdka, C. Liu, N. J. Belkin, X. Zhang, Inferring user knowledge level from eye movement patterns, *Information Processing & Management* 49 (2013) 1075–1091.
- [43] X. Zhang, M. Cole, N. Belkin, Predicting

users' domain knowledge from search behaviors, in: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, 2011, pp. 1225–1226.