

Extending the Evaluation Paradigm for Personalisation: A Categorisation Use Case

John Dowell
Computer Science
University College London
London, UK
j.dowell@cs.ucl.ac.uk

Marios Constantinides
Computer Science
University College London
London, UK
m.constantinides@cs.ucl.ac.uk

ABSTRACT

The Conlan whitepaper makes a compelling case for a community-scale means of evaluating algorithms for user modeling, adaptation and personalisation (UMAP). The authors propose an evaluation paradigm focused on a personalisation use case within an open modeling environment. Their use case is one where a mobile interface learns a user's preferences for different notifications in different contexts. Additional use cases could be incorporated in this paradigm to provide contrasting kinds of challenge for personalisation evaluation. In this position paper we summarise a use case where personalisation is achieved by categorising a user and switching their interface to a variant. The user interface is a mobile news app within a platform that also comprises a user modeling function and an interface personalisation service. Comparison of the use cases helps to map the space for the evaluation paradigm.

KEYWORDS

Evaluation, mobile news reading, personalisation, user category, interface variant.

ACM Reference format:

John Dowell and Marios Constantinides. 2019. Extending the Evaluation Paradigm for Personalisation: A Categorization Use Case. In *27th Conference on User Modeling, Adaptation and Personalisation Adjunct (UMAP'19 Adjunct)*, June 9–12, 2019, Larnaca, Cyprus. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3314183.3323682>

1. UMAP EVALUATION USE CASES

Adaptivity so changes the compass of interactive systems that its evaluation requires dedicated frameworks and methods [1]. Several frameworks have been developed and typically involve a layered approach to separately assess the input data, the inference method, the adaptation decisions, system effectiveness, and the user experience. The Conlan whitepaper makes the

contrasting and welcome proposal for an evaluation paradigm in the form of a use case, a data set, an environment for creating personalisation by reinforcement learning, and metrics for evaluating the personalisation. Their use case is smart notification for mobile users, where the system learns a user's preferences for particular kinds of notification in particular kinds of context.

The basic aim of the paradigm is to enable the comparison of personalisation evaluation methods. If the evaluation paradigm contained additional use cases representing the variety of evaluations, a more comprehensive comparison of methods would be possible.

One such alternative use case is outlined in this position statement; it is the personalisation of a mobile news app interface where the system determines a user's category on the basis of their interactive behaviour and context, and switches the interface to one of a set of variants. Evaluation of this personalised interface has been carried out both in the lab and in a field trial deployment.

2. MOBILE NEWS EVALUATION USE CASE

Mobile news apps perfectly complement the 24-hour nature of digital news, but if users are now never out of reach of the news, they need more than ever for that access to be personalised. Mobile news reading is a good target for personalisation given both its prevalence and the variation in users' preferences and patterns of use. Intuitively, some mobile users read the news frequently and broadly, trying to 'keep up with the news'; other readers read habitually but less often, trying to 'catch up with the news'; other readers read occasionally and selectively, trying to 'find news that interests them, when it interests them'. This situation should mean that the user interface of mobile news apps is personalised for different users, for example, a feature like pulling out a reading list of browsed headlines would suit one type of user but not another type. Yet personalisation of news apps has been focused on content recommendation and user interface personalisation is limited to user-configurable features. Our framing of the mobile news reading use case treats users as different sub-populations or categories who will want and need different user interfaces or interface variants. The challenge is whether a personalisation system can accurately determine a user's category and determine a suitable interface variant. Its simplest form assumes a small set of mutually exclusive categories and a corresponding set of interface variants. Interface personalisation by hiding or adding layers of features according to ability, expertise or role is a related

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honoured. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

UMAP'19 Adjunct, June 9–12, 2019, Larnaca, Cyprus

© 2019 Copyright is held by the authors. Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6711-0/19/06...\$15.00

<https://doi.org/10.1145/3314183.3323682>

approach [3, 4]. In a more sophisticated form, the challenge assumes that a user could be of several categories and personalizing a user interface then entails creating a composite of variants.

2.1. Platform and App

Habito News is a researchers' platform for investigating personalisation of news user interfaces. It consists of a personalised mobile news app, a user modelling server and a web server. The Habito News app is available through the Playstore and runs native under Android. The app logs both a user's news reading activity and its context using the device's sensors and Google's Activity Recognition API; data is transferred in batches to the user modelling server. News content is obtained from the BBC news API through a lightweight background RSS feeder service. Presentation of headlines and news articles replicates layouts used in the BBC news app, including scrollable carousels for separate news categories, each carousel presenting selectable thumbnail images with attached headlines [5]. The app has an adaptation manager that controls the Android views and interface widgets.

2.2. Data Set and Modelling

A log of users' low level behaviours interacting with Habito News is captured including each scroll through the headlines, each headline selection, each scroll through an article as it is read, and the user's location [6]. This log is passed to the user modelling server where it is translated into a set of low level features, for example, the headlines browsing direction, and the frequency headlines selection by category. Low level features also include the context features of time and location coordinates.

From these low level features six high level reader factors are determined: how often users read the news during the day (several times, once, less than once), where they read it (home, work, travelling/other), when they read it (morning, afternoon, evening), for how long they read it (less than 5, 5 to 10, more than 10 minutes) their characteristic headlines browsing behavior (browsing all categories or browsing only a preferred sub set of categories), and their characteristic article reading behavior (inferred reading speed of word-for-word, or skim reading, or keyword scanning).

From the six high level reader factors, each user is identified with one or more reader categories for which frequency of news reading is a primary discriminant. The categorisation uses supervised learning methods exploiting a training set of labeled data.

2.3. Interface Personalisation and Metrics

Personalisation of the Habito News interface is carried out when the user modeler has decided on a user's category and notifies the adaptation manager which in turn 'switches' the user interface presentation and widgets. Separate prototyping studies have established the preferences of different categories of user for different interface designs [5]. For example, tracker readers prefer news articles to lead with a synoptic paragraph and for updated news items to be highlighted.

The Habito News platform has been evaluated in a field trial with ten participants, each using the app for 6 days beginning with the same baseline interface. By day 3 of usage the user modeler determined the user's category and notified the interface manager to switch the baseline interface for a particular interface variant.

Evaluation has examined three constructs. First, the accuracy of categorisation by the user modeler: whether the user modeler can determine the user's category that the user would ascribe to themselves. Second, whether the system is able to predict a user's preference for the user interface variant compared with the baseline interface. Third, evaluation has attempted to measure the effect on user experience of switching the Habito News user interface to a category-specific variant.

3. DIMENSIONING THE PERSONALISATION EVALUATION PARADIGM

Contrasts between the mobile news use case and the Conlan use case reveal dimensions of the space of the personalisation evaluation paradigm [1]. First, the smart notification use case learns a model of the user in terms of preferences in a given context, ie, preferences for which notifications they receive and how they receive them when the context may include the location. Each user is represented individually with a highly idiosyncratic model. By contrast the Habito News use case learns a model of the user in terms of signature behaviours, including the way in which headlines are browsed and how often news is read. It then attempts to associate the user as a member of a group of similar users who may all share a preference for the same kind of interface. A second contrast raised by the use cases is whether the evaluation paradigm should involve validating the user model's predictions with users, whether it is for particular notification management strategies or for particular formats for presenting news articles in an app. These are just two of the contrasts between the use cases; other contrasts, and contrasts with yet other user cases help map the space of the evaluation paradigm for personalisation.

REFERENCES

- [1] Alexandros Paramythi, Stephan Weibelzahl, and Judith Masthoff. 2010. Layered evaluation of interactive adaptive systems: framework and formative methods. *User Modeling and User-Adapted Interaction* 20.5, 383-453.
- [2] Owen Conlan, Kieran Fraser, Liadh Kelly, Bilal Yousuf. 2018. Proposal for a Shared Challenge in the UMAP Space. <http://evalumap.adaptcentre.ie/whitePaper.pdf>
- [3] Ben Shneiderman. 2002. Promoting universal usability with multi-layer interface design. *Proceedings of ACM SIGCAPH Computers and the Physically Handicapped*, 1-8.
- [4] Francesco Bellotti, Riccardo Berta, Alessandro De Gloria, and Ludovica Primavera. 2009. Adaptive experience engine for serious games. *IEEE Transactions on Computational Intelligence and AI in Games* 1, 4, 264-280.
- [5] Marios Constantinides, John Dowell, David Johnson, and Sylvain Malacria. 2015. Exploring mobile news reading interactions for news app personalisation. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 457-462.
- [6] Marios Constantinides and John Dowell. 2018. A Framework for Interaction-driven User Modeling of Mobile News Reading Behaviour. In *Proceedings of UMAP 2018, the 26th Conference on User Modeling, Adaptation and Personalisation*. ACM, 33-41.