Proposal for a Shared Challenge in the UMAP Space

Task Title: User model generation for notification suggestion using simulated profiles

Authors[*]**:** Owen Conlan[1], Kieran Fraser[1], Liadh Kelly[2], Bilal Yousuf[1]
[1] Department of Computer Science and Statistics, Trinity College Dublin, Ireland
[2] Department of Computer Science, Maynooth University, Ireland
* Authors listed alphabetically.

Summary

This paper intends to advance the idea of shared challenges in the User Modeling, Adaptation and Personalization (UMAP) space by introducing a proposal for such a challenge and by discussing the possible metrics that may be used in such a comparative evaluation. In this proposal we put forward one potential means for shared challenge generation in UMAP. This challenge focuses on user model generation using logged mobile phone data, with an assumed purpose of supporting mobile phone notification suggestion. The simulated dataset consisting of individuals' logged mobile phone data is described, followed by challenge operation details and purpose, and intended evaluation metrics. It is not expected that this proposal is the only possible UMAP shared challenge, rather in putting forward this challenge proposal we seek to open further discussion and progress towards shared challenge generation for the UMAP community.

Background Context

There is currently no established or standardized means for comparative evaluation of algorithms and systems developed by researchers in the User Modeling, Adaptation and Personalization (UMAP) space. The development of such methodologies has proven to be extremely difficult, but would be highly rewarding for the community. Privacy concerns, the challenges of working with interactive scenarios, and the individual differences in behavior between users all must be addressed in order to facilitate repeatable and comparable evaluation and to advance research in this domain.

The EvalUMAP workshop series[1] [1, 2] is a new concerted drive towards the establishment of shared challenges for comparative evaluation within the UMAP community. The first workshop in the series brought the community together to discuss challenges and potential solutions associated with generating shared evaluation challenges in the UMAP space. Building on the success of the first edition of the workshop, the second edition made concrete steps towards identifying datasets and methods that could be exploited for shared UMAP evaluation challenges. It is intended that the third edition of the workshop, running at UMAP 2019, will make final steps towards shared challenge generation.

Ahead of this year's EvalUMAP workshop, in this paper we present a proposed methodology for the first shared challenge in the community, including practical steps to implementation. Our intention is to make this proposed methodology available for critical appraisal by the community, to foster progress and encourage further developments.

[1] http://evalumap.adaptcentre.ie/

Task Description

Use-case

The use-case for the proposed task is *personalized mobile phone notification generation* with the intention of expanding to other use-cases and challenges in the future. Our previous work in this space [3, 4] has explored intercepting incoming mobile notifications, mediating their delivery such that irrelevant or unnecessary notifications do not reach the end-user and generating synthetic notification datasets from real world usage data. The next step toward an improved notification experience is to generate personalised notifications in real-time, removing the need for interception and delivery mediation.

Specifically, assuming individuals' interactions with their mobile phone have been logged, the challenge is to create an approach to generate personalized notifications on individuals' mobile phones, whereby such personalization would consist of deciding what events (SMS received, etc.) to show to the individual and when to show them. Given the number of steps associated with such personalization, the task proposed in this paper will focus on the first step in this process, that of user model generation using the logged mobile phone interactions. For this task, a dataset consisting of several individuals' mobile phone interactions will be provided, described next.

Dataset

The dataset associated with this proposed shared challenge is a simulated dataset that is based on mobile notifications gathered by the WeAreUs Android app [5].

Since this dataset consists of synthetic data, as opposed to real individuals data, the ethical and privacy concerns are negligible as the data cannot be combined or analysed to identify real individuals.

Dataset Generation

The WeAreUs Android app collected mobile data on 15 users over a four month trial period in 2018. This data set consists of 31,239 in-the-wild notifications, 4,940 smartphone general-usage logs and 291 Experience Sample Method questionnaire responses. The WeAreUs app captured several features including the notifying app, category, priority, subject, and time of notifications, the contact, and the action taken by the user upon receiving the notifications, such as dismissing it or viewing it.

Using these captured features as seed and Cialidini's 6 principles [6] (authority, scarcity, liking, social proof, commitment and consistency, and reciprocity) as a guide, a synthetic data set with notifications that were predicted to be persuasive was generated. The Wasserstein Generative Adversarial Network with Gradient-Penalty (WGAN-GP) [7] was combined with Cialidini's 6 principles, ranging in values which indicated persuasiveness, to generate the synthetic data. WGAN-GP was chosen for generating the synthetic data as it shows enhanced training stability and enables categorical, as well as continuous, feature generation. Full details are provided in [8].

Dataset Content

The synthetic data provided in the challenge dataset is comprised of notification, engagement and contextual features. The notification features relate to the event: posting of notification to the user's device. The contextual features describe the user/device context at particular moments of interest such as when a notification is posted and when it is removed. The engagement features describe the reaction the user has to the notification. See Appendix 1, Table 1 below for an outline of the captured data features.

Challenge operation

The challenge will operate with a campaign style format.

Challenge participants will be provided with a copy of the dataset described in the previous section, and will be required to create user models for the individuals in the dataset.

As a means of steering user model creation toward a tangible goal, and hence toward evaluative metrics, two challenges in the domain of mobile notification management are proposed. Challenge 1 is an offline scenario where models are evaluated with a static test set. Challenge 2, in contrast, simulates a live interactive environment in which models must adapt on the fly. Participants can take part in one or both of these challenges.

Challenge 1: participants are given 3 months of historical notification data. The goal is to develop a user model, using the historical data, which takes a context as input and outputs a personalised notification for the given context (see Diagram 1). Once the model is built, participants can test it using an OpenAI Gym [9] environment (described below). This OpenAI Gym test environment contains an additional 3 months of contextual data which can be queried and used as input to the user model. The resulting personalised notifications generated can be returned to the test environment for evaluation.

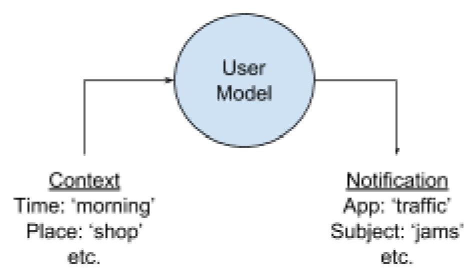

Diagram 1: Input & Output of Model

Challenge 2: participants are asked to create a user model based on the same notification, context and engagement features but without historical notification data to train with. This user model should again take context as input and output a personalised notification. This user model will need to query the OpenAI Gym environment to receive a context feature and subsequently pass it a generated notification for evaluation. As the environment steps through each context item and as engagement history becomes available, it can be used by the user model to improve the generation of personalised notifications. The goal is to develop a model which adapts and learns in real-time how to generate personalised notifications without prior history of the user (cold-start problem).

As previously mentioned, participants will test their generated user models using an OpenAI Gym [9] environment. **OpenAI Gym** is an open source interface to reinforcement learning (RL) tasks. It provides environments for researchers to benchmark RL agents on simulations of real-world problems. Gym-push [10] is a custom OpenAI Gym environment developed for this proposed challenge which attempts to simulate push-notifications arriving to a user's device, the context in which the user receives the notification and the subsequent reward received for engagements made by the user. Gym-push is a simulated environment which will be used to evaluate the performance of participants' user models. The participants will receive context objects from the environment which they can apply as input to their user models to generate personalised notifications. They can then pass these generated notifications to the environment for evaluation. Within the environment, an agent, acting as the user, will engage with the generated notifications and metrics measuring various facets of performance (outlined below) will be tracked. Additionally, if the newly generated notification data sets are saved within the gym-push environment it will give others an opportunity to benchmark their own algorithms in a standard setting as well as provide the RL community with additional flavours of environments on which to train. It is important therefore that the user models created conform to the requirements of the gym-push environment to ensure evaluation can take place (see guidelines below).

The following are a number of guidelines to consider when creating the user model:

- For Challenge 1, the user model should take a context as input and a personalised notification as output.
- Both the context and notification should be (strictly) represented by the following set of features (<u>note</u>: the names of features should not be changed):

  Notification: {postingApp, category, numberOfUpdates, subject, priority, ongoing, visibility}

  Context: {day, time, place, activity, noise, batteryLevel, charging, headphonesIn, musicActive, proximity, ringerMode}

- The values for each feature are also strictly limited to the sets provided (consult [10] for more detail).
- For Challenge 2, engagement history will also be available to the user model wishing to adapt in real-time. Engagement history will be represented by the union of *Notification* and *Context* features (same as above) as well as an additional feature 'action', which conveys whether the notification was 'opened' or 'dismissed'.

Evaluation Approach and Metrics

The gym-push environment will evaluate the user models by deploying an agent to act as a user engaging with the generated personalised notifications. The agent will be trained on historical data of the user and decide, given the context, to open or dismiss the notification generated by the model. The following two metrics will be tracked in both Challenge 1 and 2:

- *Diversity* - This metric will evaluate the diversity of generated personalised notifications which have been accepted by the agent over the 3 months. Notification sets which boast greater diversity will be scored higher.
- *Performance* - This metric will track and compare engagements resulting from the generated personalised notifications with those of the actual notifications. Scenarios which improve end-user engagements are scored higher (see table below).

For a given context		
Actual Notification	Personalised Generated Notification	Reward
opened	opened	+1
dismissed	dismissed	+0
dismissed	opened	+2
opened	dismissed	-1

Challenge 2 tracks additional metrics:

- *Response Time* - This metric evaluates the time it takes the user model to generate a notification once given the context by the environment. Shorter times are scored higher.
- *Learning Rate* - This metric evaluates how quickly the performance metric (above) of the model improves over each step (context item) of the environment.

Appendix 1

Table 1: Dataset content - an outline of the captured data features.

Feature Type	**Feature Name**	**Description**
Notification	posting app	the app from which the notification originates
	category	the category of the notification as selected by the originating app creator
	number of updates	the number of times the notification was updated with new information
	subject	the subject of the notification

		content
	priority	the priority of the notification as selected by the originating app creator
	ongoing	whether or not the notification was a persistent notification
	visibility	the visibility level of the notification e.g. viewable from the lockscreen
Contextual	day posted/removed	the day of the week the notification was posted/removed
	time of day posted/removed	the time of day the notification was poste/removed
	place posted/removed	the category of place the device is located when the notification is posted/removed
	contact significant context / contact significant in general	whether or not the associated sender of the notification is relevant to the given context / in general (to the user)
	activity posted/removed	activity (e.g. walking, still) inferred when the notification is posted/removed
	average noise posted/removed	average noise (amplitude) recorded in 5 second burst from smartphone microphone when notification posted/removed
	battery level posted/removed	battery level of the device when notification posted/removed
	charging posted/removed	whether or not the device was currently charging when notification posted/removed
	headphones-in posted/removed	whether or not headphones were attached to the device when notification posted/removed
	light intensity posted/removed	the ambient light level

		measured by device (lx) when notification posted/removed
	music active posted/removed	whether or not the device was playing music when notification posted/removed
	proximity posted/removed	a measure of how close something is to the front of the device when notification is posted/removed
	ringer mode posted/removed	the state of the ringer mode when notification posted or removed (e.g. vibrate or silent)
Engagement	time app last used	describes how long ago the app which posted the notification was last opened/used by the user
	seen time	the time it takes the user to first notice the notification after it arrives
	decision time	the time it takes the user to act upon the notification, once they have seen it
	response time	the time from the notification being posted to the user acting upon the notification
	action	the action taken by the user on the notification (open or dismiss)

References

[1] Conlan, O., Kelly, L., Koidl, K., Lawless, S., & Staikopoulos, A. EvalUMAP: Towards comparative evaluation in user modeling, adaptation and personalization. In Proceedings of the EvalUMAP Workshop, held in conjunction with the 24th Conference on User Modeling, Adaptation and Personalization (UMAP), July, 2016, Halifax, Canada.

[2] Conlan, O., & Staikopoulos, A. 2nd International EvalUMAP Workshop (EvalUMAP2017) Chairs' Preface & Organization. In Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization (pp. 171-172), ACM, July, 2017.

[3] Fraser, K., Yousuf, B., & Conlan, O. (2016, October). A context-aware, info-bead and fuzzy inference approach to notification management. In Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE Annual (pp. 1-7). IEEE.

[4] Fraser, K., Yousuf, B., & Conlan, O. (2017, July). Synthesis & Evaluation of a Mobile Notification Dataset. In Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization (pp. 179-184). ACM.

[5] Fraser, K. (2018, March). WeAreUs Mobile App. Retrieved from https://www.weareus.eu/.

[6] Cialdini, R.B.: Influence, vol. 3. A. Michel Port Harcourt, 1987.

[7] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. In: Advances in Neural Information Processing Systems. pp. 5767-5777, 2017.

[8] Fraser, K., Yousuf, B. & Conlan, O. “Scrutable & Persuasive Push-Notifications”. Accepted for publication at Persuasive Technology 2019, 14th International Conference on Persuasive Technology.

[9] Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). Openai gym. arXiv preprint arXiv:1606.01540.

[10] Fraser, K. (2019, February). Gym-push. Retrieved from https://github.com/kieranfraser/gym-push.