MANAGEMENT OF USER PERSONA UPDATES BY PREDICTING USER BEHAVIOURAL PATTERNS

B. A. Rathnayaka

158244K

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

February 2019

MANAGEMENT OF USER PERSONA UPDATES BY PREDICTING USER BEHAVIOURAL PATTERNS

B. A. Rathnayaka

158244K

Thesis submitted in partial fulfillment of the requirements for the degree Master of Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

February 2019

DECLARATION

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works.

…………………………………………......   ………………………

Buddhika Amila Rathnayaka   Date

The above candidate has carried out research for the Masters thesis under my supervision.

…………………………………………......   ………………………

Dr. G.I.U.S. Perera   Date

ABSTRACT

User centred design equipped persona concept to provide better and more acceptable user interfaces. This persona concept based on user research which collects information about the user's goals binds with the product. Due to user bases changing without prior notice, personas are vulnerable to become outdated or not productive without any notice. Avoid these situations, need to revise personas with time.

Providing a mechanism to find a point of persona updating is the focus of this research. Predicting user and persona relationship in periodically and analyse user behaviour against personas provide statistics to finding a persona updating point.

Neural network model is created and trained using the data that use for persona creation for predict personas mapped with the users periodically. According to these statistics, it is possible to find the consistency of user's stickiness to persona and whether the system needs a persona revising.

If system owners or maintainers do not monitor the changes of the user base, it is very difficult to identify whether the current status of personas are outdated or not. If there is no mechanism to monitor user base, then persona revising happens unnecessarily. The proposed method for find persona updating point help to manage the frequency of persona updates with avoiding situations that system using outdated personas.

Considered data set within the research shows that 71.33% of persona user consistency. This figure generally shows that the majority of users well served by personas and can consider there is no need to updates personas at this time. It is possible that stakeholders of the product have higher needs of providing highly effective user interfaces with the product, situation (e.g., consistency rate above 80%) like this there is a persona application for the application considered in the product.

Keywords: Neural network, User cantered design, User persona, User persona updates

ACKNOWLEDGEMENT

I would like to thank my supervisor, Dr Indika Perera for the supervision, for providing right leads and guidance and for helpful advice given throughout to make this research a success. My sincere appreciation goes to my family for the continuous support and motivation given to make this thesis a success. Last I also thank my friends and staff of the department of computer science and engineering - the University of Moratuwa who supported me in this whole effort.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

HCI	Human Computer Interaction
PoC	Proof of Concept
UCD	User Centred Design

Chapter 1

INTRODUCTION

Alan Cooper introduces user personas, in his book "The Inmates Are Running the Asylum" that published in 1998. There are two types of personas that use for marketing and development. Byer personas used to define the possible customer who purchases the product and user personas used to defining possible users who interact with the product. User personas are subject to this thesis and refer it as "persona" here onwards within this thesis. Purpose of using personas is understanding the user's expectations and how they like to use the product and group the users into persona clusters. Personas are fictional characters that act as representatives of user groups. According to Alan Cooper, personas are defined by their goals [1]. "Personas create a strong focus on users and work contexts through the fictionalized setting" [4]. With this introduction to the software industry, it provides a better understanding of application users to developers and designers when following the user centred design.

1.1 User Personas

Personas used to understand users of the product. It is difficult to focus on the needs of all users when designing, but if there are few, then it is easy. So creating few personas provide this focus for designers and developers. Even though they are fictional by adding real-life identity, it can consider as a representative of a certain user group. When bringing personas into real life, following information categories are taken into consideration.

- Persona Group
- Fictional name
- Job titles and major responsibilities
- Demographics such as gender, age, education, ethnicity, and family status
- The goals and tasks they are trying to complete using application
- Their physical, social, and technological environment
- Pain points or frustrations
- Casual pictures representing that user group

Can personas use forever? It is possible that when created personas work well for a certain period, but the factors related to the background research which done for persona creation can change with the time. The user base behind the research can change with time. The goals of the users of the product can change with time due to the changes happen to their lives like occupation, marital status, attitude, the technology they deal with in day to day life and new users come into the user base and some of the existing users left from the user base. This type of changes causes the mismatches between the users and personas associated with the system [7].

Personas need to update or revise over the time to avoid situations like mentioned above. Personas can user forever if there are not any changes within the product user base. However, in real-world user bases does not have that much of stable behaviours. The best way to avoid these type of obstacles is to revise personas and update the product according to revised personas.

1.2 Research Questions.

Persona creation involves three different stages, regardless of the approach taken to persona creation.

1. Conducting research or gathering data
2. Analysing data
3. Crafting the persona

Performing these steps take some time and conducting research or gathering data takes most of the time form the entire process [7]. Time consumption causes a problem with the revising personas due to the effort needed to complete a cycle. All cycles need to complete these steps because the changes happen within the user base and need to analyse real user for crate personas.

If possible to revise personas only after a proper period which is not sooner or later, avoid the unnecessary revising cycles and effort, is an advantage. Maintain proper period between two revision cycles is a solution for this and need to come up with a mechanism for finding a proper time of user persona updates.

1.2 Objectives.

As above mentioned, providing a mechanism to find the proper time for revising persona is subject to this thesis and research work conducted, for fulfilling the above requirement with the following objectives.

- Create a neural network model which can identify a user's persona based on user behaviour.
- Predict persona that mapped with the user, based on user's behaviour
- Compare user persona predictions done at several times after persona creation for identifying user behaviour changes.

Chapter 2

LITERATURE REVIEW

2.1 Human Computer Interaction (HCI)

Human Computer Interaction “is a discipline concerned with the design, evaluation, and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them” [9, p. 7]. It covers various fields like computer science, psychology, sociology, industrial design, which gives science, engineering and design qualities to HCI [8, p.20].

When considering goals of HCI it can summarise to safety, utility, effectiveness, efficiency and appeal and these goals focus on make the system easy to learn, easy to use, limiting error frequency and severity. In general, HCI makes the system simple and usable [8, p.24]. For achieving the goals of HCI, some approaches can be utilized.

Those are:

- Involving users who can influence system design.
- Integrating, different kind of knowledge and expertise that can contribute to HCI design.
- Use the iterative process.

Above mentioned shows, HCI is a user-centric design, and it is important to undertake usability evaluations to get feedback regarding negative and positive aspects of the prototype. [8, p.25]

Earlier days HCI experts are involved at the end of the design process and later understood that is a mistake and their consultancy should be provided from the beginning if design process. So it is important to consider how HCI applies to the overall development process. [8, p.21] “The Interface is not something that can be plugged in at the last minute; its design should be developed integrally with the rest of the system. It should not just present a “pretty face”; but should support the tasks that people actually want to do, and forgive the careless mistakes” [10, p.3].

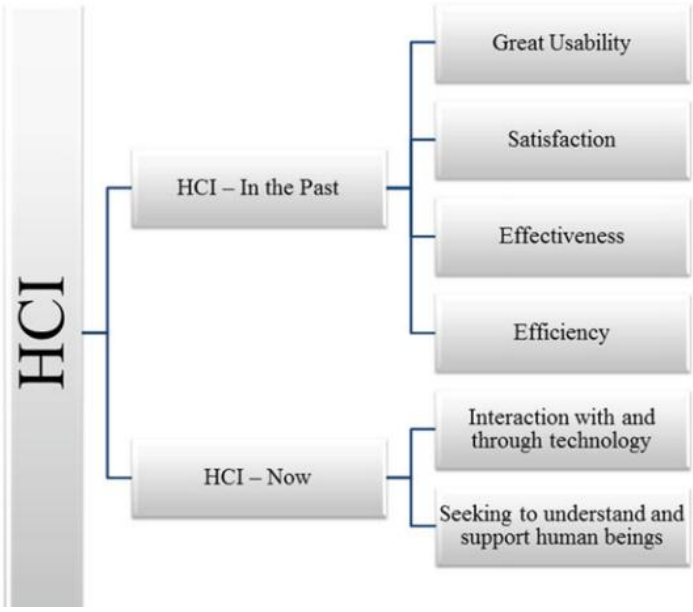

Figure 2.1 HCI Past and Now [12].

2.2 Usability

As mentioned in the background section, it measures the quality of the system, and it measures how easy to use a system. “From the user’s perspective, usability is considered a very important aspect in the development process as it can mean the difference between performing and completing a task in a successful way without any frustration” [8, p.29].

“There are principles need to be followed in order to make systems easy to use and easy to learn.

Those principles are:

- **Learnability**: by which new users can begin effective interaction and achieve maximal performance
- **Flexibility**: the multiplicity of ways the user and system exchange information
- **Robustness**: the level of support provided to the user in determining successful achievement and assessment of goals

- **Efficiency**: once the user learns about the system, the speed with which s/he can perform the tasks
- **Memorability**: how easily the user will remember the system functions, after a period time of not using it
- **Errors**: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
- **Satisfaction**: how enjoyable and pleasant is it to work with the system?

These principles can be applied to the design of an interactive system in order to promote its usability. Therefore, the purposes behind adopting these principles are to give more assistance and knowledge to system developers (and the users) regarding the system design." [8, p.33]

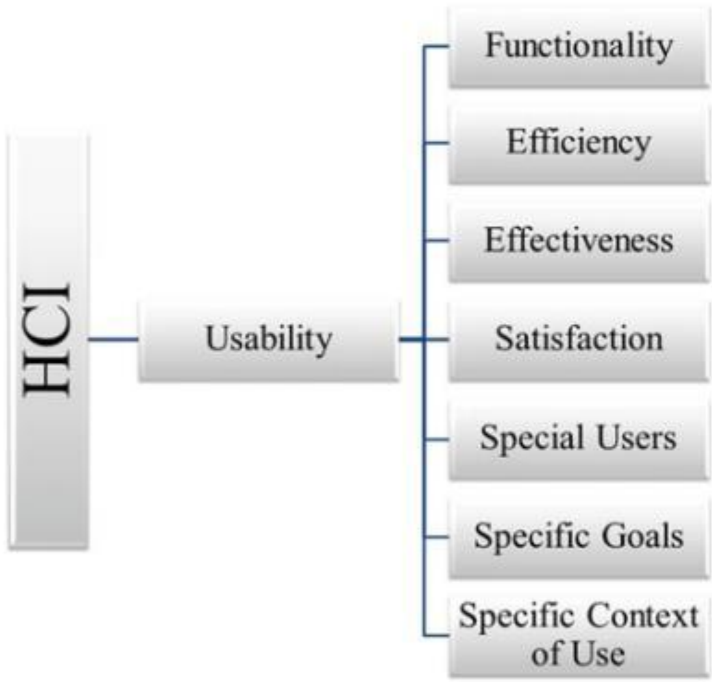

Figure 2.2 Areas that Usability should address [12].

2.4 Persona and User Centred Design.

According to Alan, creating a persona is “Develop a precise description of our user and what he wishes to accomplish” [1]. Persona is a powerful tool that allows developers and designers to bring actual user needs into the design. In the process of persona creation, there are several steps to follow.

- Research on users.

  Data for create persona is gathered through the researching on target users collect a large amount of data, and this is an expensive and time-consuming task.

- Set up user categories

  User’s roles and goals are taken into consideration and group user into categories who have similarities in their behaviour.

- Create skeleton

  Review the user database on a user group and identify important facts. Skeletons that can consider as a prototype of personas are created using identified important facts.

- Prioritise skeletons

  Since a limited number of personas allowed within a system, all skeleton are not able to convert as personas. Most important skeletons prioritise for persona creation.

- Convert skeletons to a persona

  Skeletons only contain abstract data received from the user research, but personas are fictional characters from the real world. While this convention happens, the fictional properties such as name, photo and some other human characteristics are added.

Sample templates to represent a persona.

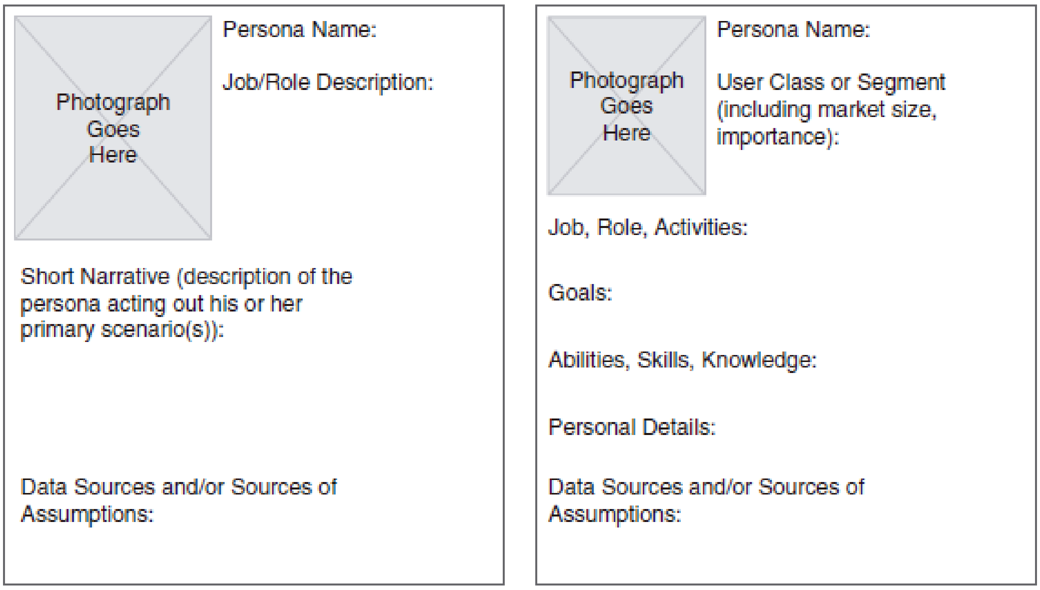

Figure 2.3 Persona templates. Left: One-Page Style. Right: Resume Style [2].

"Personas are very believable. A persona is more likely to be accepted and remembered than other user-centred design methods because a persona owns many natural attributes of human beings" [11].

User-centred design (UCD) process made considerable improvements to developing products that can satisfy individual needs and easy to use. The popularity of UCD allows web sites, systems and products to improve their usability by bringing users or consumers into the design process. According to Tomasz and Kenneth "While preaching the importance of practising user-centred concepts, many organisations fail to consider the consumer needs as the focal point of their design processes" and this implies that still, UCD is not able to eliminate all the problems faced by a customer such as product complexity and difficulties with using features of modern products [5].

In UCD, the understanding of the user is the main task. According to John and Tamara "Our more natural tendency is to be self-centred, which translates to taking an approach to product design based on our own wants and needs, but most of the time the people on your product development team are not representative of the target audience for your product and the result is inadequate product" [2]. If the centre is replaced by a Persona who is created based on actual users, there are many benefits and the improvements that can achieve from incorporating UCD with persona. Following are Few benefits can gain from UCD and personas incorporation listed by Tomasz and Kenneth collected within the literature [5].

- Increase focus on the users and their goals
- Facilitate effective communication about users
- Increase focus on a specific audience
- Challenge assumptions
- Lead to more user-friendly designs
- Lead to better design decisions

2.4 Revising Personas

Even though personas are fictional characters, the creation of personas is entirely base on the real user of a product. Every time that user base changes, personas should reflect them, but it is not happening since the personas, and their sources are not actively connected. Due to this separation, personas are fixed with their properties. With time passes, personas deviated from their actual users. According to Kim, there are several reasons which cause these changes in user bases. Changes mainly fell into the following main categories.

- Changes in business and technology
- Changes user demographics and interaction patterns

When considering business and technology, it possible to evolve the technologies used with their products or competitors with their products, which can change the user's attitudes to expect more usability from the product. Receive a set of new customers for the product and lose some existing customer, and it changes the composition of the user base [7].

These factors highlight the increasing the differences between user and personas and the literature shows survey results about the effectiveness of persona against their refreshing frequency.

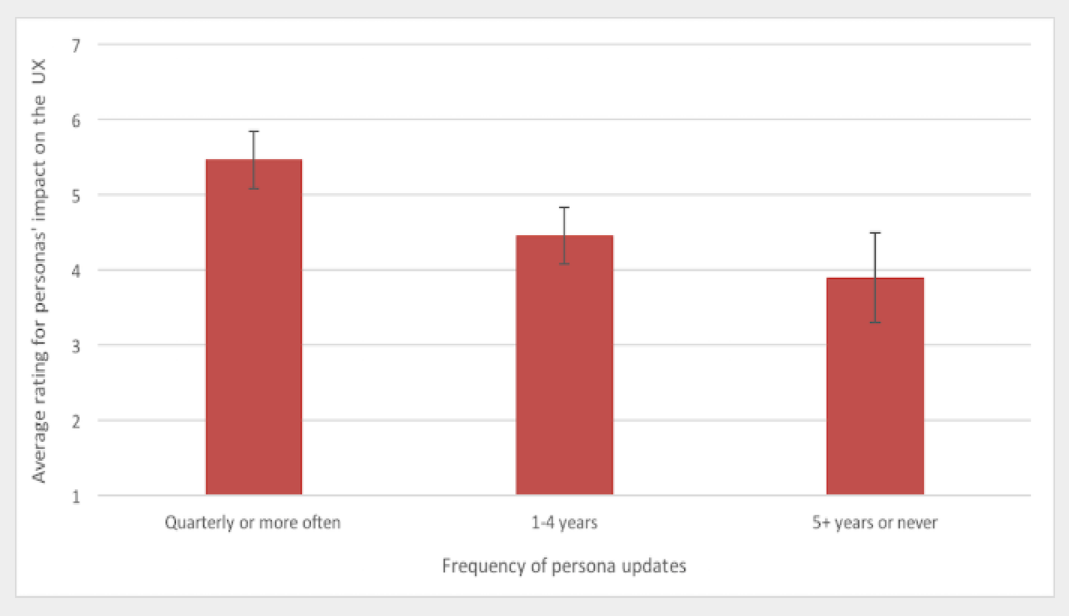

Figure 2.4 Average personas' impact rating segmented by revision frequency [7].

2.6 Behavioural Patterns, Goals and Personas

A persona with well-understood goals and behavioural patterns it is possible to satisfy a larger group of users represent by that persona [12]. Research done with real users allows gathering data about behavioural patterns. Goals are the divers of the real user's behavioural patterns because the user expects to achieve his goal by performing activities. Accurately finding a goal is not possible by directly asking the user's goal as a research question due to the user not be able to answer a question accurately or possible to hide his goals form the third party. Finding the goals of users from behavioural patterns and other data is a task of designers and user base researches [13].

To find out behavioural patterns, need to consider behavioural variables of users and using those variables, try to capture user's activity frequencies, desires, motivations and many other interests. Other than behavioural variables, it is possible to use demographic variables to identify behavioural patterns [13].

2.6 Artificial Neural Networks

"Artificial neural networks are mathematical inventions inspired by observations made in the study of biological systems, though loosely based on the actual biology" [14]. Purpose of artificial neural networks maps set of inputs to the desired output for even unseen set of inputs by making their own decisions. Neural networks consist layer of neurons called input layers, output layers and one or many hidden layers. Neural networks are powerful tools that can capture complex relations and called them as universal classifiers [15,16].

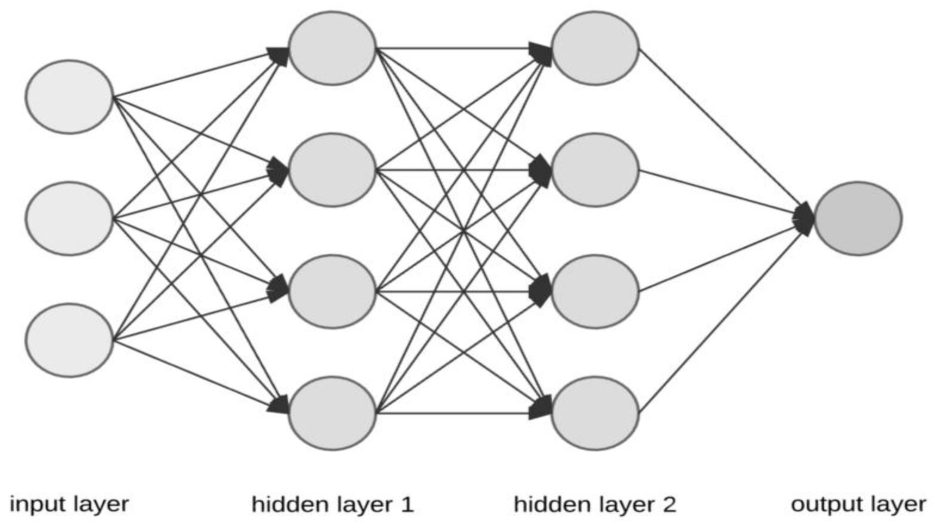

Figure 2.5 Multilayer neural network.

2.6 Training Neural Networks

Neural networks need to training to learning the functionality that they need to apply when the set of inputs received. This learning process called model training. According to Kevin and Paul, there are mainly two methods for training a neural network, and those are

- Supervised Training
- Unsupervised Training

Supervised training conducted by using training data sets that have a set of inputs and a set of desired outputs. Inputs are fed to the neural network and predict outputs then calculate the error of predicted output base on the desired output given with given data set. Considering the error of prediction, the weights of the neural network adjust to minimise the error. This process repeats until neural networks achieve an acceptable level of accuracy of predictions [14].

Unsupervised training conducted by a set of training datasets and set of adaptation rules that control the general behaviour, instead of the desired outputs. Adaptations rules are the calculators of error neural network. Weights of the neural networks adjust base on the responses of input dataset and adaptation rules until the neural network achieve desired performance. Since the unsupervised training method depends on its response, it does not need a set of desired output for certain inputs [14].

2.6 K-means Algorithm and Clustering

Clustering method used with an application like data mining and knowledge discovery, pattern recognition and classification based on patterns [20]. There are many methods for solving clustering problems and most mostly used one is k-means clustering, and there are many variants of this algorithm exists [19].

When the k-means algorithm is capable of clustering data into similar groups that have patterns, it is possible to use creating sample persona clusters. Skit learn k-means algorithm implementation is used within-cluster sum-of-squares create clusters and algorithm need the number of clusters that data set need to split.

Chapter 3

METHODOLOGY

Methodology to find facts that need to decide whether personas need an update or not is discussed in this chapter. Comparison between behavioural patterns initially assigned for a user and behavioural patterns observed while using the product, consider as a base for the decision making mentioned above. Behavioural pattern belongs to persona consider here is a collection of individual user behaviours

3.1 Methodology

Persona creates based on behavioural patterns identified within product user research. Each user has their behaviour and different from others even though they are mapped into a single persona through a behavioural pattern. A person consists of users who have closely matching behaviours and persona consider as a behavioural pattern of all users belongs to it.

Since behaviour uniquely map with an individual user, Considering B as behaviour, then mapping between a behaviour specific to a user (B_{ui}) and the user (U_i) represent as

$$B_{ui} \rightarrow U_i$$

Since personas are defined as a behavioural pattern of a set of user behaviours (Bu). The relationship between the behavioural pattern of persona (PI) and set of user behaviours (Bu) represent as following,

$$B_{ui} \in P_I$$

Since a product can have multiple personas (e.g., P_X, $P_{Y,}$.....P_Z), it is possible to represent all behaviours of a set of user($B_u(x)$) and relationship with the relevant persona as,

$$
\begin{gathered}
B_{ux} \in P_X \\
B_{uy} \in P_Y \\
\vdots \\
B_{uz} \in P_Z
\end{gathered}
$$

$$\{x \mid x \in \mathbb{N}\}, \{y \mid y \in \mathbb{N}\}, \{z \mid z \in \mathbb{N}\}, \forall\,(x, y, z), B_{ux} \neq B_{uy} \neq B_{uz}$$

Considering the unique mapping between users and their behaviours, the relationship between a set of users U(*x*) and relevant persona can represent as following,

$$
\begin{gathered}
U_x \in P_X \\
U_y \in P_Y \\
\vdots \\
U_z \in P_Z
\end{gathered}
$$

Changes in the behaviour of a user can identify comparing two instances of behaviour created with t time gap. Let consider the initial set of behaviour recognition for users at time t_1, and the relationship between corresponding personas represent as follows.

$$
\begin{gathered}
B_{t_1 ux} \in P_X \\
B_{t_1 uy} \in P_Y \\
\vdots \\
B_{t_1 uz} \in P_Z
\end{gathered}
$$

Users set mapped to personas through the behaviours at time t_1 can represent as follows.

$$
\begin{gathered}
U_{t_1 x} \in P_X \\
U_{t_1 y} \in P_Y \\
\vdots \\
U_{t_1 z} \in P_Z
\end{gathered}
$$

These set of behaviours and corresponding personas use to train the neural network to predict personas based on future user behaviour changes. Behaviour set of users considered as inputs and corresponding personas considered as desired outputs for the neural network within the supervised training process. This training provides a neural network model based on behaviours of users at time t_1, called "Persona Prediction Model at t_1" (PPM-T1).

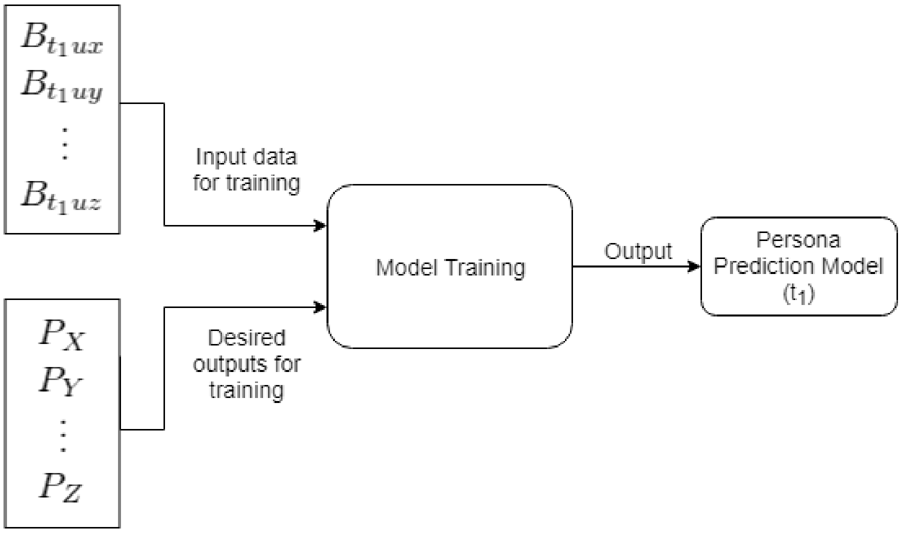

Figure 3.1 Model training process at time t1

The next step is predicting personas for users after t period at $t_2 = (t_1 + t)$. This persona prediction based on behaviours identified after t period using the model trained at time t_1. Running this prediction provides the personas matched with behaviours at time t_2 and relationship can represent as following,

$$\begin{gathered} B_{t_2ux} \in P_X \\ B_{t_2uy} \in P_Y \\ \vdots \\ B_{t_2uz} \in P_Z \end{gathered}$$

Since each behaviour uniquely matched with a user, find the sets of users who belong to a specific persona at time t_2 and mapping between users set and the persona can be represented as follows.

$$
\begin{gathered}
U_{t_2x} \in P_X \\
U_{t_2y} \in P_Y \\
\vdots \\
U_{t_2z} \in P_Z
\end{gathered}
$$

There are two result sets which provide relationships between the set of users and personas at different times t_1 and t_2. Comparison of these user sets against the corresponding persona provides statistics of changes within the user base of the product according to user behavioural patterns.

3.2 Data Collection for Model Training and Persona Predictions

Data need to test the model described in the above section is the behaviour data of users. User activities related to the behavioural variables which are used to create user personas should collect as test data.

Since the user behaviours generate within the process by considering user activities based on periods, collecting data must be able to identify based on relevant periods. The behaviour data related to the period t_1 use for training the neural network model and data related to the period t_2 use to predict behavioural patterns to find relevant personas for users at time t_2.

A product that defines user behaviour based on feature usage within the product. Feature usage measured by considering user visits to features provided within the product. Structure of collecting data may be as represented below.

Table 3.1 Sample structure of expected data collection

User	Time	Activity
U1	T1	Activity 2
U1	T2	Activity 3

U1	T3	Activity 4
U4	T4	Activity 2
U4	T5	Activity 3
U3	T6	Activity 2
U3	T7	Activity 3
U4	T8	Activity 2
U4	T9	Activity 3
U3	T10	Activity 2
…	…	…

3.3 Pre-Processing Data

Collected behavioural data use within the proposed model at two steps of the process based on the period they are gathered. The first occurrence of data usage is training the neural network, and the second is predicting behavioural pattern based on user's behaviour data. For use, these row behavioural data within the process need pre-processing to getting proper results.

Considering above mentioned application which defines user behaviour based on feature usage. Feature usage is the behavioural data that input to the neural network. Calculation of feature usage of a specific user is within considering period,

Usage of specific feature *i* as U_f (i), Number of visits to the feature V(i), Total number of visits to all features TV,

$$U_f(i) = \frac{V(i)}{TV}$$

Table 3.2 Usage data of features for period t1.

User	Feature 1	Feature 2	Feature 3	Feature 4	Feature N
U1	x	x	X	x	x
U2	x	x	x	x	x

…	…	…	…	…	…
U3	x	x	x	x	x

Single record of usage data table shows the behaviour of a user for a period. Based on user and persona mapping, neural network training data preparation may be as following,

Table 3.3 Inputs to neural network training for period t1.

User	Feature 1	Feature 2	Feature 3	Feature 4	Feature N	Persona
U1	x	x	X	x	x	P1
U4	x	x	x	X	x	
…	…	…	…	…	…	…
U3	x	x	x	x	x	P2
U6	X	X	x	x	X	

Only the feature values and persona consider during the neural network training process.

Feature usage calculated using the data collected for period t_2 also arranged as the same structure of Table 3.2. Data use for predict persona and output result of the persona may be as following,

Table 3.4 Predicted output using the trained model at time t2.

User	Persona
U1	P1
U3	P1
…	…
U4	P2
U6	P2

For find, the persona updating point needs to consider the user's movements within personas. When a user behaved differently, the user assigned to a persona based on changed user behaviour. Other than that user should remain within the same persona. Completing several sessions of user persona predictions at different times for different periods and comparing those result provide behaviour changes of users. Comparison result can represent as follows.

Table 3.5 Persona user comparison result

Initial Persona (Assigned with Persona Creation)	Persona at time t1	Persona at time t2	Percentage of users with persona changing pattern (%)
x	x	x	a
x	x	y	c
x	z	z	d
….	…	…	….
y	y	y	f

3.4 Neural Network Model Training and Architecture Configuration

Structure of the data set available for train neural network is shown in above table 3.2, and it provides set of inputs (feature usage data) and mapping classes (personas), shows that input and class relationship is a classification problem. Inputs are consists set of values representing the usage of features and no relation between features. Multilayer perceptron architecture is used to model the classification problem.

Multilayer Perceptron architecture consists of several layers of perceptrons, as shown in image 2.4. The number of perceptrons in the Input layer equal to the number of features in feature vector and number of perceptron in output layer equal to the number of classes defined for output vector. There no specific numbers of hidden layers and the number of nodes for a hidden layer defined, and those values need to decide by experimenting with data related to the relevant question.

To find out the best neural network model needs an experiment on available data and set of experiments conduct as follows.

Table 3.6 Model selecting experiment setup

Number of Hidden	Y

Layers								
Number of Neuron Per Hidden Layer	X		X		X		X	
Number of Research	1	2	1	2	1	2	1	2
Training Result	S/F	S/F	S/F	S/F	S/F	S/F	S/F	S/F
Average Number of epochs	Z		Z		Z		Z	
	S = Success, F = Fail.							

Chapter 4

ARCHITECTURE AND TECHNOLOGIES

The proposed method should have combined with a practical application for test and evaluate. User activity data of an application that user can use features to explore information based on their interest to achieve their goals could provide suitable samples for test the proposed model.

Let consider a web application that promoting and selling applications for pcs or mobiles. Targeting scenario within a web application is find user's interest in certain information categories that provided with selling or promoting applications and create personas based on that information. Information categories considered within these scenarios are,

- Detailed information about application features

  Content: Feature description and screenshots of user interfaces

- User reviews section

  Content: user review comments and ratings by individual users

- Application popularity

  Content: Overall application rating, Number of downloads, Ratings for vender

Personas use for design interfaces to provide a more convenient and usable application.

4.1 Architecture for Proposed Method

Mainly components spread within both client side and server side. Activity data collecting components exist within the client side, and data processing components exist within the server side. Server-side contains a separate server for deploy data processing components due to the resource requirements. Collected data stored in separate storage considering the safety of data collected data for a long period. Components of application and proposed method with interaction within the components are shown in figure 4.1.

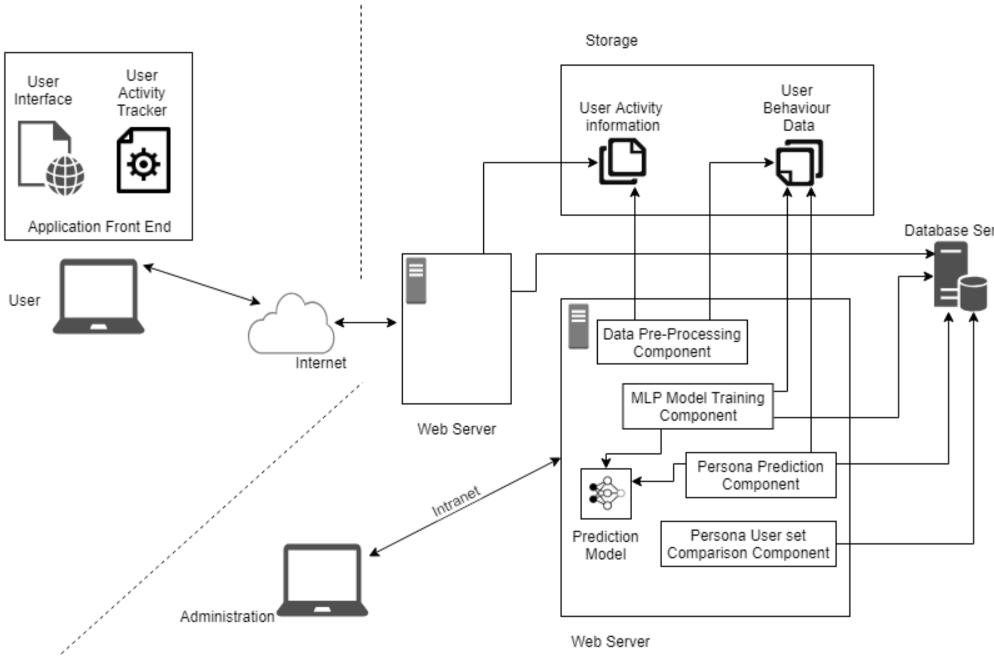

Figure 4.1 High-level architecture diagram for integrating the proposed method with the product.

User Activity tracker runs in client-side and recorded information stored in server side files. Data collected according to the structure presented in table 3.1.

Data pre-processor prepares user behaviour data for each user by calculating feature usage related to the considering a given period. Created behaviour data stored in user behaviour data files according to the structure presented in table 3.2.

MLP Model Training component responsible for train MLP neural network model. Component access the user behaviour data files and database to generate inputs to the neural network training process. Received user behaviour data for a certain period modified by adding relevant persona based on the user to match the structure presented in above table 3.3. The trained model saves for future use.

Persona prediction component is responsible for predict personas based on user behaviour information for a given time. User behaviour information access through user behaviour data files. Predicted persona information store within the database for comparison purposes.

Persona user set comparison responsible for compare users associated with personas and provide statistics about user deviations of user behaviours during a given period.

Components related to the process of this method deployed in a separate server to avoiding disturbances for website activities, since some components need a considerable amount of resources for their execution, e.g., neural network model training.

Database server use for store web application & business related information and processed information of persona updates management system.

4.2 Component Implementation

Purpose of implementing components is demonstrating the feasibility of the proposed method. Planned implementation level is a proof of concept (PoC) and completed up to satisfy the PoC level. Implementation is mainly targeted on components related to the proposed method, and the following components were prioritised while implementing

- Data Pre-Processing
- MLP Model Training
- Persona Prediction
- Persona User Set Comparison

4.3 Technologies

Implementing components of the proposed method using several technologies based on functions carried out by components. Mainly used technologies while implementation, are listed below,

- TensorFlow
- TensorTlow.js
- Node.js
- JavaScript

TensorFlow is a machine Learning framework created by Google for training machine learning or deep learning models, predict future results. It is an open source software, and flexibility of TensorFlow is allowed to deploy on various platforms like Central processing units and graphics processing units. It provides python based, convenient front-end application interface for building applications [18].

TensorFlow.js also an open source JavaScript library that can provide facilities to training and run machine learning models completely on the browser. Further TensorFlow.js is allowed to use TensorFlow as the machine learning service provider, while TensorFlow.js runs in NodeJs.

Node.Js is an open source JavaScript platform based on Google's V8 JavaScript engine that allows client-side programming using JavaScript. Node.js is event-based which using event loop [17].

Chapter 5

EVALUATION

Evaluation of the proposed method discusses under this chapter. A good data set is a vital factor for conduct a proper evaluation for the method proposed. Since there is not a product that has access to collect data, there are another two options can consider creating a dataset for evaluation, which creates a prototype of the product that considering within the research and allow to use some user or find a data set that closely matches the requirements. By creating a prototype and allow the user to use it and collect data is a hard and time-consuming task for fulfil data set requirements.

5.1 Dataset Qualities for Evaluation.

Considering available options and decides to use a data set that has closely matching qualities expected which collected from real users with some other application.

Requirements expect within data collection,

- Enough number of users.

  Few users are not enough to create a behavioural pattern since neural networks need a considerable amount of records to train a proper model
- Enough number of records of user activities

  User behaviour also cannot define looking at two or three activity records
- Should be able to find a sequence of user activities based on time

  Since this method considering periods for extract user behaviour, needs to find user activities in sequence based on time
- Recordset should spread within a considerable period

  To fulfil fist two requirements, need to track activities of users for a certain period and comparison between periods, users need to show their activities within all comparing periods.

A collection of data found on the internet named “Black Friday” [21] which provides user interests about categories of purchasing items and that can fulfil the requirements mentioned above at a considerable level. The data set, taken into consideration, provides the user’s interest in certain categories. Data set does not contain a time value field, but examining the user activity sequences appear within data set it can recognise, activity records appear according to the occurrence of activities. Since it is not providing date time stamp tracking on activity records, have to depend on the activity sequence and measure the time as a percentage based on the total number of activities.

The data set provides details of 5891 users and contains 537577 number of user activity records. The following figure shows there is approximately more than 100 user participation available for each set of 1000 records, most of the time.

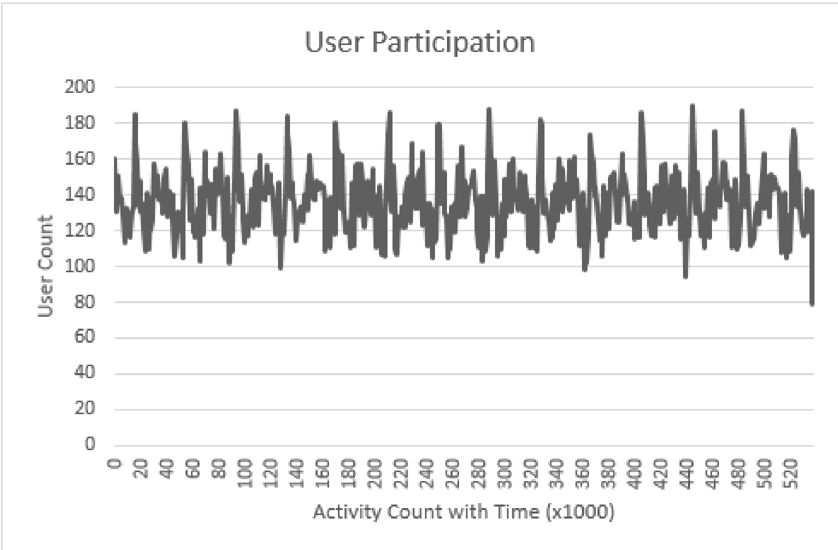

Figure 5.1 User participation for every 1000 records within data set

Information about the user’s interest provided for 18 categories. Due to the expected user behaviour definition based on major three activities within the considered application, only three activities related to suitable categories taken into consideration while mapping dataset users with the users of application considered within the research.

Considered activities of categories mapped with the features of the application. The sequence of activities shows the users have involved with categories with time gaps. Within each time their involvement, they access certain categories, and this behaviour can map with an application user who accesses item to sell (a widget or application) and visit its different information categories.

Selected categories for mapping with application information categories, highly visited by the users and most of the users are visited. Because of the nature of categories and user's behaviour, it is possible to mimic the web application user behaviours using the found data set.

Since the data set contains a sequence of activity occurring order, it is possible to use to predict future activities based on previous activities. This prediction nature of the data set is the most important quality required to test the proposed persona prediction method.

There is no period related to collected data, mentioned with the data set, but it is possible to define a timeline using user activity sequence and a large number of user activities available with the data set.

5.2 Evaluation Plan

Based on the method presented need to fulfil a few prerequisites and came up with the following plan to conduct the evaluation. Relevant entities of data extract from data collection that is the first step of the evaluation plan. Following are the evaluation plan.

1. Choose behaviour data for personas
2. Create personas using behaviour data
3. Find the best MLP model configuration for data.
4. Train MLP model using data chosen for persona creation
5. Choose behaviour data to predict user behaviour patterns.

6. Analyse user behaviour changes.

Within the next sections of this chapter explain and show the results related to each step of the evaluation plan.

5.3 Data Selection for Personas

There is a large number of records and users within the selected dataset. All users are not active at all the time. According to figure 5.1, the number of active users is between 100 and 200 out of 5891 users for any randomly chosen adjacent 1000 activity records. Set of users need to choose who shows most participation for activities and activities also spread within the dataset. Users should frequently active and long term participation. Finding these set of users important due to proposed method compare different time gaps.

Table 5.1 User appearance within intervals based on activity interval lengths.

Activities per interval (x 1000)	Total Number of Intervals	Maximum number of intervals a user appear	Percentage of the user appears in the maximum number of intervals
1	538	14	44.57
10	54	14	44.57
20	27	14	44.57
30	18	14	44.57
40	14	14	25.73
50	11	11	61.02
60	9	9	70.19
70	8	8	76.50
80	7	7	84.21
90	6	6	89.71

100	6	6	81.97
110	5	5	94.77
120	5	5	91.03
130	5	5	40.31
140	4	4	67.67
150	4	4	96.72

According to the above table with activities per interval increases, the appearance of a user within maximum number intervals, increases compared to the total number of intervals. Since the number of activities per interval increase, the number of interval decreases, appearance in the maximum number of intervals increases but the frequency of appearing all over the periods could decrease. With facts of the above table, the best set of users can find with both total numbers of intervals and user percentage with the highest score. Since these factors have inverse proportional behaviour, suitable users set can be identified as follows

Table 5.2 Selection criteria for bets user sets.

Activities for interval	90000
Total Number of Intervals	6
Maximum number of intervals user appear	6
Percentage of the user appears in the maximum number of intervals	89.71

Considering data set provides information about user interests of 18 categories. Mapping these categories with main activities described within chapter Architecture and Technologies can generate user behavioural pattern for users of the application that taken into consideration. According to category user interest data shown below tables 5.3, most interaction happens with categories 1,5 and 8.

Table 5.3 Category user interaction appears in the dataset.

Product Category	1	2	3	4	5	6	7	8	9
Use Count	138353	23499	19849	11567	148592	20164	3668	112132	404

Product Category	10	11	12	13	14	15	16	17	18
Use Count	5032	23960	3875	5440	1500	6203	9697	567	3075

After considering user appearance frequency and interest of categories, three categories (1,5,8) selected for evaluation form the original collection, which contains activities related to 3517 users.

5.4 Persona Creation

Persona creation from users based on their behaviour as well as goals within the application. Considering data set for simulate application users do not provide any information about users other than interests on certain categories, option for creating personas based on researched goals and behaviours is not available. Clustering mechanism for cluster users into a few groups based on their behaviour information. Those user groups can use for persona profile creation within the proposed method evaluation process.

Mapping of information providing sections of application and categories selected from dataset done as following.

Table 5.4 Information section and category mapping.

Information Providing Section	Category
Detailed information on application features (DIAF)	5
User reviews section (URV)	1
Application popularity (APP)	8

Users' behaviour defined based on categories selected in section 5.3. Calculate ratios of interaction between information sections for each section base on total interactions, happen within the time interval that considering according to the explanation on methodology chapter. Calculated ratios for some users are listed below.

Table 5.5 Information section visiting statistics.

User ID	URV	DIAF	APP
1000034	0.0909	0.7273	0.1818
1000035	0.1429	0.2857	0.5714
1000036	0.4091	0.4773	0.1136
1000037	0.1429	0.5714	0.2857
1000039	0.4000	0.2000	0.4000
1000042	0.6667	0.1667	0.1667
1000044	0.5263	0.2105	0.2632
1000045	0.0323	0.6452	0.3226
1000048	0.4828	0.3621	0.1552
1000049	0.6000	0.1000	0.3000
1000050	0.2000	0.0000	0.8000
…	…	…	...

Persona profile creation conduct based on clustering users according to their behaviour information. User clustering using the K-means algorithm. Each user got associated with a cluster after applying K-means algorithm. Users are divided into three groups within clustering. A specific number of groups (3) selected because a lower number of personas always preferred due to the minimising product development complexities.

The output of the K-means algorithm for sample data set as shown below,

Table 5.6 Clusters and the number of users assigned

Cluster	Number of Users Assigned
0	1054
1	1179
2	1284

5.5 Neural Network Model Configuration.

Multilayer Perceptron model train to predict personas of users based on their behaviours. Before train a model to achieving this task, it needs to find the best matching configuration for the model going to build.

While experimenting for the best configurations, the following factors are taken into consideration. Measure the result based on training result successful or failed, and the number of epochs passed to complete the training. Resultant data set with clusters, generated using the K-means algorithm.

- Number of neuron per hidden layer
- Number of hidden layers

While conducting training to find the best configuration expect to achieve a certain level of training accuracy. Those values are

Table 5.7 Accuracy property values need to achieve while training.

Parameter	**Value**
Training accuracy	>0.99
Training loss	<=0.01
Validation accuracy	>0.9
Validation loss	<=0.1

Model training process stopped and saved the trained model when these requirements are met. However, another constraint added due to saving time from configurations that take a long time. Stop model training at the completion of 10000 epochs, and the above requirements are not met.

Table 5.8 Experiments for model configuration.

Number of Hidden Layers	2							
Number of Neuron Per Hidden Layer	30/10		30/20		30/30		30/40	
Number of Research	1	2	1	2	1	2	1	2
Training Result	S	S	S	S	S	S	S	S
Average Number of epochs	1155		1122		2277		693	

Number of Hidden Layers	2							
Number of Neuron Per Hidden Layer	30/50		30/60		20/50		40/50	
Number of Research	1	2	1	2	1	2	1	2
Training Result	S	S	S	S	F	F	S	S
Average Number of epochs	561		5346		Stop at 10000 epochs		2673	

According to the experiments results all the single hidden layer results failed. Only some models with two hidden layers give successful results. Initially found that two-layer network with nodes 30/40 shows good performance. Since configuration lies within the edge of the experiment fields, decides to conduct another set of experiments surrounding the 30/40 result such as 30/50, 30/60, 20/50,40/50 and fund that two-layer network with neurons 30/50 gives the best performance. Experiment results related decision showed above, and the rest of the results included in the Appendix I.

Neural network model created using Tensorflow.js. Three nodes for each output and input layers and the number of hidden layers and neurons for a layer decide using experiments. Neural network model builds using Tensorflow.js sequential structure. Output layer used “softmax” activation function due to the neural network model used to solve a classification problem. All other layers use “relu” activation function. Learning rate is a considerable factor when training an accurate model, but in the tensorflow.js sequential model, it is an optional parameter and tensorflow.js automatically adjust as it needs during model training. Because of that learning rate is not considered as a variable while experimenting.

5.6 Training Neural Network and Predict Personas

An essential factor for training a neural network is data collection related to solving the problem. The research uses a dataset already collected, need to choose a portion of data for training of neural network and the rest of the data to predict user behaviours. Since the data set does not contain the time and only have a sequence of activities based on occurrences of activates, data portions should select as percentages by considering activity occurrence sequence as the propagation of time.

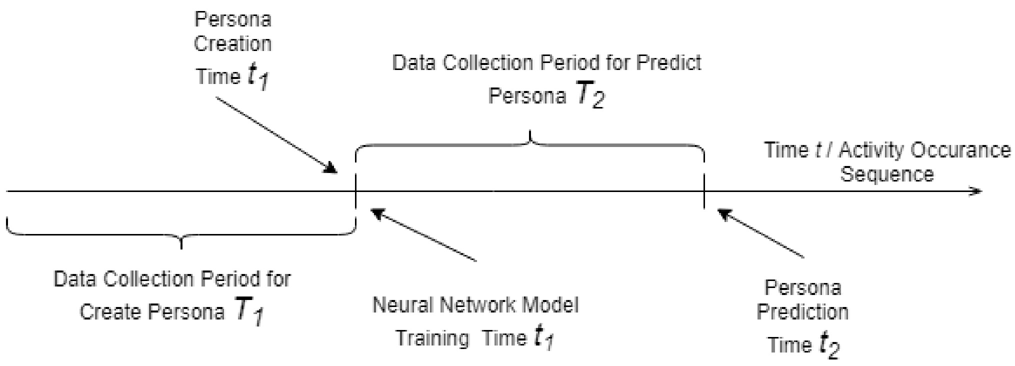

Figure 5.2 Data selection based on activity sequence.

40% of data from the beginning of the sequence, which belong to T_1 period according to figure 5.2, selected for neural network model training and rest of data use for predict user behaviours. Since the K-means algorithm user as persona creation mechanism, use the k-means algorithm and cluster data into three groups and found personas belongs to each user. K-means algorithm output for selected activities.

Table 5.9 Persona and user mapping by K-means algorithm.

Cluster	Number of Users Assigned
0	1558
1	1467
2	1653

Following images show the 3d space scatter plot for collected data before ration calculation and clustered data after ratio calculations between user's visits to information sections.

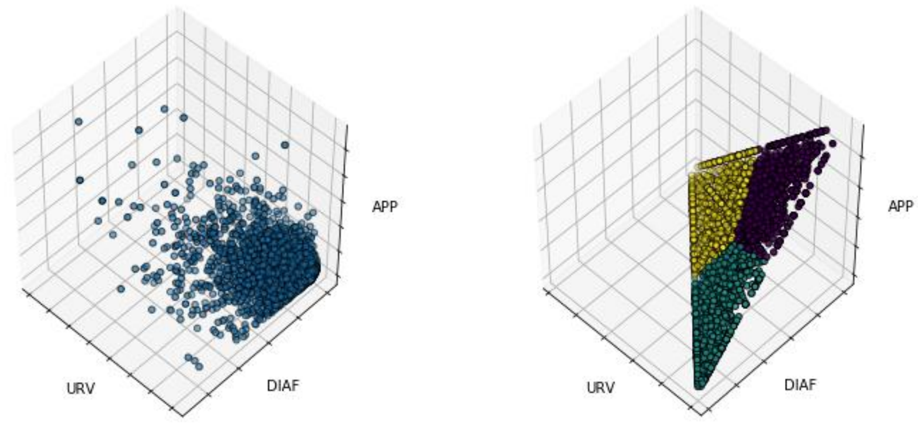

Figure 5.3 Scatter plots – Left: Before ratio calculation Right: After ratio calculation and clustering

5.7 Predicting Personas

Data related to T_2 period, as mentioned in Figure 5.2 need to prepare for predict personas. T_2 time can define as a window of time, and there are two different ways to define the length of a period for the window. One technique is growing window, and the other one is moving window.

The growing window can represent as following,

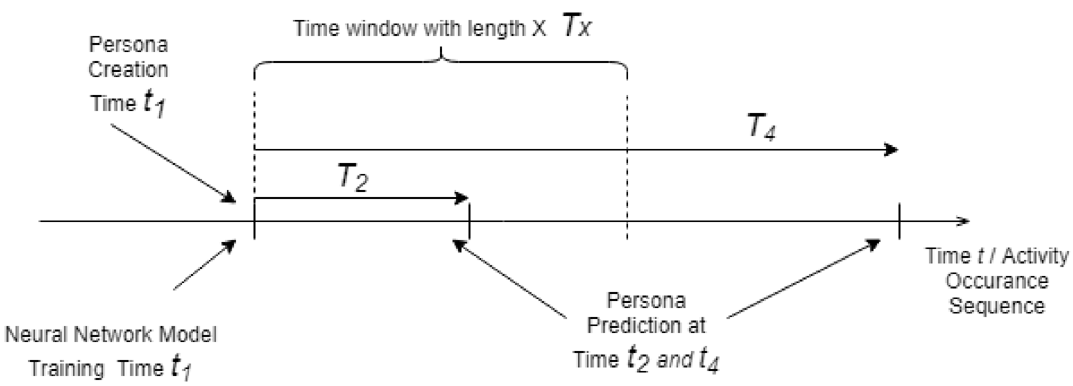

Figure 5.4 Growing time window

This window always starts at the same position, the persona creation time at T_1 and growing up to certain time x and allow collect data within that period. When x increases, the amount of data gathering for period *Tx* also increases.

Moving time window has fixed length, and the starting position moves forward with time when defining different periods for data collection. This method provides data sets that have the same amount of data. The main difference between growing and moving windows is, the growing window allows to remember the user's history of activity for a long time but moving window limits the user's memory for a short period. The moving window can represent as follows,

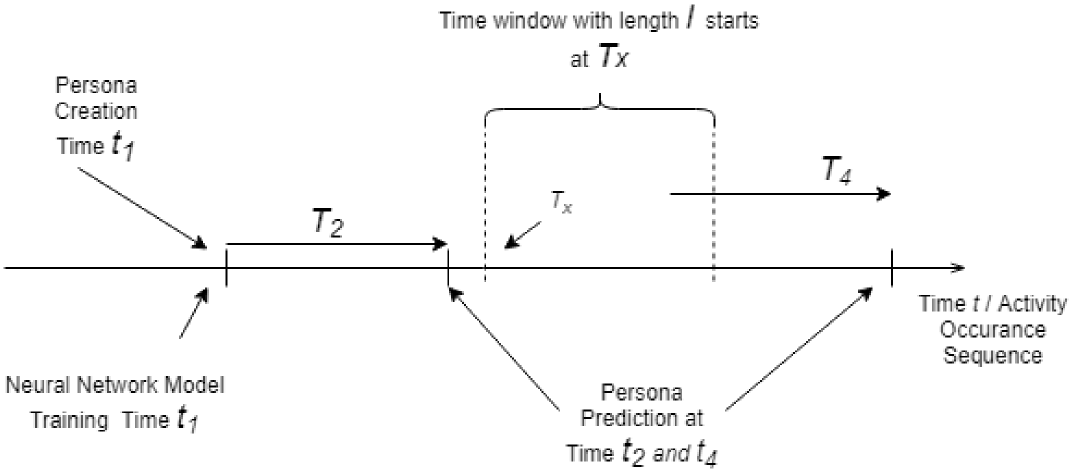

Figure 5.5 Moving time window

40% of data is already selected for persona creation and neural network. Only 60% of data remaining for user behaviour. Data sets predict personas create based on both techniques growing and moving tie window and following are the experimental data sets for persona prediction.

Table 5.10 Experiment set up to find user behaviour using persona prediction

Prediction Series	**Data Set**	**Selection Type**	**Length %**	**Start Position %**	**End Position %**	**Number of Activity Records**
1	1	GW	30	40	70	161273
	2	GW	30	40	100	322546
2	3	GW	20	40	60	107515
	4	GW	20	40	80	215030
	5	GW	20	40	100	322545
3	6	MW	30	40	70	161273
	7	MW	30	70	100	161273
4	8	MW	20	40	60	107515
	9	MW	20	60	80	107515
	10	MW	20	80	100	107515
	GW =Growing Window MW = Moving Window					

Predictions for chosen data sets done using the trained neural network model. Prediction for datasets within the same prediction series (e.g., data set 1 & 2) uses to find the behaviour within the entire period. Predicted results arranged to highlight on user's movements within personas by ordering results according to the percentage of users with persona changing pattern and results for prediction series one represent as following,

Table 5.11 Persona prediction results for series 1

Initial Persona	**Persona at 70%**	**Persona at**	**Percentage of users with persona changing pattern**

		100%	(%)
1	1	1	22.25
2	2	2	20.88
0	0	0	19.78
2	0	0	6.25
0	2	2	6.04
2	1	1	3.75
1	2	2	3.03
0	1	1	2.54
1	0	0	2.47
0	2	0	1.73
2	2	0	1.42
0	0	2	1.12
2	0	2	1.12
1	0	1	0.92
2	1	2	0.85
1	2	1	0.83
0	1	0	0.79
1	1	2	0.79
2	2	1	0.58
0	0	1	0.56
1	1	0	0.45
1	2	0	0.43
0	1	2	0.34
2	0	1	0.31
2	1	0	0.27
0	2	1	0.25
1	0	2	0.25

According to the above results of predictions, 62.91% of user never changed their persona because they have not changed their behaviour. Further, some users show consistency after changing persona for the first time and as a percentage is 24.08%. As a total of 86.99% user stick to their personas without changing their behaviour, and others do not show stable consistency with a persona. Rest of the full results of predictions are included in Appendix II.

Summary of results,

Table 5.12 Persona prediction result summaries

Prediction Series	Window Type	Window Length	Never Change Persona (%)	Consistent After First Change (%)	Other Changing Patterns (%)	Total Stable Consistency (%)
1	GW	30	62.91	24.08	13.01	86.99
2	GW	20	57.04	19.17	23.74	76.21
3	MW	30	53.12	13.32	33.56	66.44
4	MW	20	43.11	6.45	51.71	49.56
	GW =Growing Window MW = Moving Window					

5.7 Prediction Result Analysis.

According to the above summary, all experiments provide at total stable consistency very close or above 50%. When considering configurations of an experiment parameter such as data collection window type and window size, results show different percentage values. With the increase of window, length users show more consistent behaviours with persona and smaller the length lower consistency with persona. Base on the window type also shows the same type of behaviour that growing window type shows more consistent results than moving window data set results. Moving time windows with long length and growing time windows with smaller lengths are both contain a higher amount of data and provides a higher influence from older behaviour on predicting personas for new behaviours. When considering the results from prediction series 2 and 3 have a certain level of similar behaviours when compared to the results of other configurations and shows those configurations are capable of minimising the older behaviour influence for new predictions.

Identifying these configuration values (e.g. series 2 and 3) for window length and window type, allow to choose the best prediction series to decide persona update point. Considering both series 2 and 3, as an average of 71.32% of persona consistency shows with the system. According to that above result, personas still have better consistency with the majority of users

Chapter 6

CONCLUSION

Persona is a concept used within UCD to providing better user interfaces based on user behaviour and user goals. Since personas completely based on users, they can be outdated or not productive with time passes. The research tries to provide a mechanism to get statistics to decide the need for a persona revision.

6.1 Summary

With the focus of achieving the above goal, proposed a personal prediction and analysis technique using a neural network. In this method, a neural network model train to identifying user persona relationship and used it to find current user behaviour is matched with persona it is assigned. A product that promotes and sells other applications such as mobile apps and widgets, to apply the proposed method to implement only necessary components to prove the method proposed. Tested this method using a data collection which closely matching data set. Results show that the method can provide information about the consistency of user persona relationship and allowing to define a persona updating point.

Considering results received it is 71.32% final results that it system still serving well for the majority of its users and updates on a persona is not essential. If owners of the product that have a requirement that system should keep the persona user consistency in very high levels (e.g., above 90%), this is a persona revising point for the system.

6.2 Limitations

The proposed method is capable of providing user and persona relationship consistency. That information helps to take a decision, and it did not say that the point needs to do a persona update. Because that depends on the product owner's decision about what is the consistency percentage, they need to keep with the system.

Within the research, the application of the proposed method discussed using a certain type of product, product’s usage and data used for testing the method related to the considered product. Since a user behaviour is an activity that has high variation, usage of the method can be limited with different user behaviour patterns.

Since still need to conduct experiments that need to come up with optimal window length that minimise the influence of user behaviour history for the prediction. Due to final consistency results depends on these experimental values, it could be an effect on the accuracy of final consistency results.

Persona creation method K-means algorithm equipped within the research is not the actual user research for persona creation. K-means algorithm is capable of clustering things which have similar qualities since it is not representing actual persona creation method and add limitations to this method.

6.3 Future Work

The prediction component has to run multiple times for finding the data selection window type and window length to minimising user behavioural history to current persona state prediction. Since these parameters still based on experiments, it is drawn back to the proposed method. Improving the window length and window type selecting mechanism to give a vast advantage for this method to use accurately. This section should be taken in to research work.

REFERENCES

[1] A. Cooper, *The Inmates Are Running the Asylum*. United States of America: Sams Publishing, 2004

[2] J. Pruitt and T. Adlin, *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*, San Francisco, CA: Morgan Kaufmann Publishers Inc., 2005

[3] T. Adlin and J. Pruitt, *The Essential Persona Lifecycle: Your Guide to Building and Using Personas*, San Francisco, CA: Morgan Kaufmann Publishers Inc., 2010

[4] J. Grudin and J. Pruitt, "Personas, Participatory Design and Product Development: An Infrastructure for Engagement," Proc. PDC 2002, pp.144-161.

[5] T. Miaskiewicz and K.A. Kozar, "Personas and user-centred design: How can personas benefit product design processes?," *Design Studies*, vol.32, pp.417-430,2011.

[6] K. Flaherty,"How Much Time Does It Take to Create Personas?," www.nngroup.com , para 2, Oct. 25, 2015.[Online].Available: https://www.nngroup.com/articles/persona-budgets/**.** [Accessed Feb. 28, 2019].

[7] K. Flaherty,"Are Your Personas Outdated? Know When It's Right To Revise," www.nngroup.com , para 2, Feb. 14, 2016.[Online].Available: https://www.nngroup.com/articles/revising-personas/. [Accessed Feb. 28, 2019].

[8] T. Issa and P. Isaias, *Sustainable design: HCI, usability and environmental concerns*. United Kingdom: Springer London, 2015

[9] J. Preece, D. Benyon, T. Carey, S. Holl and Y. Rogers. *Human-computer interaction: Concepts and design*, Reading, MA: Addison-Wesley Educational Publishers, 1994

[10] A. Dix, J.E. Finlay and G.D. Abowd, *Human-computer interaction.* 3rd edn, Harlow: Pearson Prentice-Hall,2006

[11] X. Wang, "Personas in the User Interface Design," Alberta, University of Calgary, 2017, [Online]. Available: https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/UCALG_CA/U071112W.pdf. [Accessed Mar. 12, 2019]

[12] A. Cooper, “Perfecting Your Personas.” www.cooper.com, para 4, May. 15, 2008.[Online].Available: https://www.cooper.com/journal/2001/08/perfecting_your_personas. [Accessed Feb. 28, 2019].

[13] A. Cooper. R. Reimann, D. Cronin. *About Face 3 The Essentials of Interaction Design.* Indianapolis, IN: Wiley Publishing, Inc. 2007

[14] K.L. Priddy and P.E. Keller. *Artificial neural networks: an introduction*. Bellingham, WA: SPIE, 2005

[15] H.B. Burke, P. E. Keller, S. Hashem, L. J. Kangas and R. T. Kouzes “The Importance of Artificial Neural Networks in Biomedicine,” Applications of Neural Networks in Environment, Energy, and Health. Singapore: World Scientific Publishing: 1996, pp.145–153.

[16] M. H. Hassoun, Fundamentals of Artificial Neural Networks. London: The MIT Press, 1995

[17] H. Shah and T. R. Soomro, “Node.js Challenges in Implementation,” Global Journal of Computer Science and Technology: E Network, Web & Security,17, Jun. 2017

[18] A. Martín. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng. “TensorFlow: A System for Large-Scale Machine Learning,” In Proc. 12th USENIX Symposium on Operating Systems Design and Implementation ’11, 2016

[19] T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu,”An Efficient k-Means Clustering Algorithm: Analysis and Implementation”, IEEE Transactions on Pattern Analysis & Machine Intelligence, vol.24, pp. 881-892,2002.

[20] R.O. Duda and P.E. Hart, Pattern Classification and Scene Analysis. New York: John Wiley & Sons, 1973.

[21] M. Dagdoug, “Black Friday - A study of sales through consumer behaviours”. Kaggle.com, 2018. Available : https://www.kaggle.com/mehdidag/black-friday. [Accessed Nov. 20, 2018].

Appendix I

Number of Hidden Layers	1							
Number of Neuron Per Hidden Layer	10		20		30		40	
Number of Research	1	2	1	2	1	2	1	2
Training Result	F	F	F	F	F	F	F	F
Average Number of epochs	Stop at 10000 epochs		Stop at 10000 epochs		Stop at 10000 epochs		Stop at 10000 epochs	

Number of Hidden Layers	2							
Number of Neuron Per Hidden Layer	10/10		10/20		10/30		10/40	
Number of Research	1	2	1	2	1	2	1	2
Training Result	F	F	S	S	S	S	F	F
Average Number of epochs	Stop at 10000 epochs		2244		4026		Stop at 10000 epochs	

Number of Hidden Layers	2							
Number of Neuron Per Hidden Layer	20/10		20/20		20/30		20/40	
Number of Research	1	2	1	2	1	2	1	2
Training Result	F	F	F	F	F	F	F	F
Average Number of epochs	Stop at 10000 epochs		Stop at 10000 epochs		Stop at 10000 epochs		Stop at 10000 epochs	

Number of Hidden Layers	2							
Number of Neuron Per Hidden Layer	40/10		40/20		40/30		40/40	
Number of Research	1	2	1	2	1	2	1	2
Training Result	F	F	F	F	F	F	F	F
Average Number of epochs	Stop at 10000 epochs		Stop at 10000 epochs		Stop at 10000 epochs		Stop at 10000 epochs	

Appendix II

Prediction results for series 2

Initial Persona	Persona at 60 %	Persona at 80 %	Persona at 100 %	Percentage of users with persona changing pattern (%)
1	1	1	1	20.86
2	2	2	2	19.07
0	0	0	0	17.11
0	2	2	2	4.61
2	0	0	0	4.53
2	1	1	1	2.52
1	2	2	2	2.22
1	0	0	0	1.79
0	2	0	0	1.76
0	1	1	1	1.74
1	2	1	1	1.36
2	0	2	2	1.3
2	1	2	2	1.22
2	2	0	0	1.19
1	0	1	1	1.08
0	0	2	2	1.08
0	2	2	0	0.98
0	1	0	0	0.84
2	0	0	2	0.81
2	2	1	1	0.76
2	2	2	0	0.73
0	0	2	0	0.6
0	0	0	2	0.6
1	0	0	1	0.54
1	2	2	1	0.54

1	1	2	2	0.54
1	1	1	2	0.52
2	1	0	0	0.49
2	0	2	0	0.46
0	1	1	0	0.41
2	1	1	2	0.41
0	2	1	1	0.38
0	2	0	2	0.38
0	0	0	1	0.38
0	1	2	2	0.35
2	2	1	2	0.33
2	2	0	2	0.33
0	0	1	1	0.3
1	1	1	0	0.27
0	0	1	0	0.27
1	1	0	1	0.27
1	2	2	0	0.27
2	2	2	1	0.24
2	1	2	1	0.24
1	1	2	1	0.24
1	1	0	0	0.24
2	0	1	1	0.22
1	0	2	2	0.22
0	2	2	1	0.19
1	2	1	2	0.19
1	0	1	0	0.16
1	2	0	0	0.16
0	1	1	2	0.16
1	0	0	2	0.14
2	1	1	0	0.14
1	0	2	0	0.11
2	0	0	1	0.11
0	1	0	1	0.11

2	1	2	0	0.08
1	2	0	1	0.08
0	2	1	2	0.08
1	2	0	2	0.08
2	1	0	2	0.05
2	2	1	0	0.05
1	0	2	1	0.05
2	1	0	1	0.05
2	0	1	2	0.05
1	0	1	2	0.05
2	0	1	0	0.05
0	1	2	0	0.05
0	0	1	2	0.05
0	1	2	1	0.05
0	2	0	1	0.03
1	1	0	2	0.03

Prediction results for series 3

Initial Persona	**Persona at 70%**	**Persona at 100%**	**Percentage of users with persona changing pattern (%)**
1	1	1	20.23
2	2	2	17.38
0	0	0	15.51
0	0	2	4.2
2	2	0	3.95
2	0	0	3.62
0	2	2	3.6
0	2	0	3.51
2	0	2	3.13
1	1	2	2.44

2	2	1	2.11
2	1	2	2.06
2	1	1	1.87
1	1	0	1.87
1	2	2	1.73
1	2	1	1.7
0	0	1	1.59
1	0	1	1.4
0	1	0	1.37
0	1	1	1.26
1	0	0	1.24
2	0	1	0.82
0	2	1	0.8
2	1	0	0.74
0	1	2	0.66
1	2	0	0.66
1	0	2	0.55

Prediction results for series 4

Initial Persona	Persona at 60 %	Persona at 80 %	Persona at 100 %	Percentage of users with persona changing pattern (%)
1	1	1	1	17.13
2	2	2	2	14.16
0	0	0	0	11.82
0	0	2	0	2.4
2	2	0	2	2.11
0	0	0	2	2.11
2	2	2	0	2.08
0	2	0	0	1.91

2	0	0	0	1.84
2	0	2	2	1.71
0	0	2	2	1.58
1	1	2	1	1.52
0	2	2	0	1.45
0	2	2	2	1.45
2	2	1	2	1.45
1	2	1	1	1.45
2	1	2	2	1.42
2	0	0	2	1.42
2	2	0	0	1.25
1	1	0	1	1.25
2	0	2	0	1.19
0	2	0	2	1.19
1	1	1	2	1.19
2	2	2	1	1.19
1	0	1	1	1.19
1	2	2	2	1.02
0	1	0	0	0.99
0	0	1	0	0.89
2	1	1	1	0.89
1	1	1	0	0.86
1	1	2	2	0.82
2	1	2	1	0.79
1	2	1	2	0.69
2	2	1	1	0.69
0	0	0	1	0.66
2	1	1	2	0.66
1	0	0	0	0.66
0	1	1	1	0.59
1	0	0	1	0.53
1	2	2	1	0.53
1	1	0	0	0.49

0	2	0	1	0.46
0	1	0	1	0.43
0	0	2	1	0.43
2	1	0	0	0.43
2	0	0	1	0.4
1	0	1	0	0.4
0	1	2	1	0.36
2	1	0	2	0.36
0	0	1	1	0.33
0	1	0	2	0.33
0	2	1	2	0.33
0	0	1	2	0.33
1	0	1	2	0.33
0	2	1	0	0.3
1	2	2	0	0.3
2	2	1	0	0.3
1	1	0	2	0.3
2	0	2	1	0.26
1	2	1	0	0.26
0	1	1	0	0.26
0	1	1	2	0.26
1	2	0	1	0.26
2	0	1	0	0.23
0	1	2	0	0.23
2	1	2	0	0.23
2	0	1	1	0.23
0	2	2	1	0.23
2	1	0	1	0.23
1	0	2	0	0.23
1	0	2	1	0.2
0	1	2	2	0.2
0	2	1	1	0.2
2	0	1	2	0.2

1	1	2	0	0.16
2	1	1	0	0.16
1	2	0	2	0.13
2	2	0	1	0.13
1	2	0	0	0.13
1	0	0	2	0.13
1	0	2	2	0.13