JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 34, 1141-1173 (2018)
DOI: 10.6688/JISE.201809_34(5).0003

ID3P: Iterative Data-Driven Development of Personas to Improve Business Goals, Strategies, and Measurements

YASUHIRO WATANABE, HIRONORI WASHIZAKI,
KIYOSHI HONDA, YUKI NOYORI, YOSHIAKI FUKAZAWA,
AOI MORIZUKI, HIROYUKI SHIBATA, KENTARO OGAWA, MIKAKO ISHIGAKI,
SACHIYO SHIIZAKI, TEPPEI YAMAGUCHI AND TOMOAKI YAGI
School of Computer Science and Engineering
Waseda University
Tokyo, 169-8050 Japan
Yahoo Japan Corporation
Tokyo, 102-8282 Japan
E-mail: {jellyfish44-time; akskw-luck}@akane.waseda.jp;
{washizaki; Fukazawa}@waseda.jp; khonda@ruri.waseda.jp;
{aomorizu; hirshiba; keogawa; miishiga; sashiiza; tepyamag; tyagi}@yahoo-corp.jp

Personas are fictional characters used to understand users' requirements. Many researchers have proposed persona development methods from quantitative data (data-driven personas development). These works do not assume that personas are used continuously or that they can reflect on changes in users, making it difficult to plan reliable strategies in a web service due to dynamic changes in users' preference. Generally, measuring the effect of strategies is challenging. Personas, which do not reflect on actual current users, prevent effective measurements of business strategies and suitable decision-making. To develop more suitable personas for decision-making in a web service, we previously proposed Iterative Data-Driven Development of Personas (ID3P). To detect changes in users' characteristics, our proposal includes an iterative process where personas are quantitatively evaluated and revised. Moreover, it provides a quantitative evaluation of business strategies based on GQM+Strategies and personas to improve business strategies and goals. This paper is an extension of our previous work. ID3P can verify personas and strategies even when changes in personas are not drastic. We employed additional case study involving Yahoo! JAPAN's web service called Netallica to verify it.

Keywords: requirements engineering, market-driven requirements engineering, data analysis, personas, data-driven personas, GQM+strategies

1. INTRODUCTION

A persona, which is a fictional character designed to understand users' requirements, is a representative human-centered design method. Personas, which were created qualitatively, were proposed originally by A. Cooper. However, previous studies have noted some issues:

- Personas differ from actual users when they are not based on users' data.
- Personas are not used for decision-making in design, *etc.* [2, 3].
- Meeting the persona's requirement does not ensure that a business goal is achieved [4].

Received September 6, 2017; revised October 17, 2017; accepted December 11, 2017.
Communicated by Yu-Chin Cheng.

In connection with the second issue, some researchers have proposed methods to analyze the requirements based on personas [5-7, 34]. Additionally, some previous methods have proposed applying personas to Agile or Scrum [8, 9]. These methods can be considered as applications of personas to decision-making in the software design process. However, we conjecture that previous methods do not adequately address decision-making in web services because they do not include the continuous development of personas to reflect changes in users explicitly. Due to a rapidly changing environment and users switching between services, decision-making in a web service is challenging. Consequently, the following points are considered to be decision-making obstacles in a web service:

- According to other requirements engineering work [10, 11], understanding users' preferences and requirements is difficult because they dynamically change.
- The Hawthorne effect makes it challenging to measure the impact of strategies' on users [12]. Consequently, defining effective target users to achieve a goal is problematic.

To solve the above issues, we proposed Iterative Data-Driven Development of Personas (ID3P) for practical applications of data-driven persona development on real services. ID3P constructs personas based on actual users' data, and should be a solution to the first issues with personas. Many previous methods have proposed data-driven persona development methods using various types of data [13-19]. In 2016, data-driven development of personas from users' clickstreams on a website, which is a type of big data, was proposed [20].

In particular, we integrated data-driven persona development and evaluation into GQM+Strategies (GQM+S), which is a goal-oriented model to measure business goals. To address decision-making in a web service, ID3P includes:

- It assists in understanding users in a service via iterative evaluation and revision of personas.
- It quantitatively analyzes persona characteristics to easily derive strategies.
- It quantitatively evaluates goals and strategies based on personas to enhance a business goal, strategy, and measurement.

This paper is an extension of our previous work [1]. Our contribution in this paper is to verify the continuous application of ID3P. ID3P can verify personas and strategies even when changes in personas are not drastic. We employed an additional case study involving Yahoo! JAPAN's web service called Netallica to verify it. Through the additional case study, we verified that ID3P helps managers to plan strategies based on precise assumptions and suitable personas. We investigate the following research questions:

RQ1 Can personas' characteristics in ID3P derive business goals and strategies?
RQ2 Can ID3P verify personas constructed via a data-driven approach?
RQ3 Can the assumptions about personas be verified quantitatively via ID3P?
RQ4 Can strategies and goals be evaluated based on personas quantitatively?
RQ5 Does revising personas aid in understanding users and planning strategies?

To answer these research questions, we employ two case studies with Netallica, which is a web curation service of Yahoo! JAPAN.

Section 2 reviews the basic concept behind ID3P and describes the motivating examples. Section 3 explains ID3P. Section 4 presents our two case studies, while section 5 analyzes our results. Sections 6 and 7 discuss threats to validity and related work, respectively. Section 8 summarizes our conclusions and contributions.

2. BACKGROUND

2.1 Data-Driven Persona Development

A persona is a fictional character developed to understand users' requirements. Similar to a real person, a persona has attributes (*e.g.*, name, gender, job, characteristics, goal for using its service, *etc.*). Initially, personas were created qualitatively. Previous studies have reported the issue that personas differ from actual users when they are not based on users' data.

To solve this issue, previous work proposed data-driven construction of personas (Data-Driven personas) [13, 14]. Data-driven construction involves employing actual users' data to relate personas to actual users. Various techniques are employed to construct personas (*e.g.*, latent semantics analysis [16], prediction model [19], topic model [15], association rule mining [18], hierarchical clustering [20], simulation [17], *etc.*) In 2016, data-driven development from a certain type of big data was proposed [20], but it does not consider issues with running a long-term service (*e.g.*, changes in users' requirement).

2.2 GQM+Strategies

GQM+Strategies (GQM+S) is a measurement approach for business goals based on Goal-Question-Metric (GQM). GQM+S includes a hierarchical model of the goals and strategies where each strategy is derived from a goal. In GQM+S, the reasons why each strategy are derived are related to link between Goal and each Strategy as Assumption or Context. Context is a reason based on highly accurate information, whereas an assumption is any other reason. In GQM+S, every business goal is measured by several metrics, which are derived by the GQM approach, to determine whether a goal is achieved [21].

GQM+S involves the following activities to measure and achieve business goals [21]:

1. Characterize the environment
2. Define goals, strategies, and measurements
3. Plan model implementation
4. Execute plans
5. Analyze outcomes
6. Package improvement

In connection with practical cases, previous methods applied GQM+S to several types of real services and validated its effectiveness [22-24]. However, these methods do not improve the GQM+S model based on the strategy execution result. Other methods have demonstrated a method to improve the quality of the GQM+S model [25-27], but they focused on model construction. Iterative improvement of GQM+S is not well discussed. ID3P provides a method to improve the GQM+S model based on actual users' data.

2.3 Motivating Example

ID3P helps service managers make decisions by understanding changes in users' requirements. Here, we explain issues that ID3P can solve. We assumed a web service, which provides a web application to end-users. Such services make a profit on advertisements or licensing fees. The profit depends on the number and the persistence of users. Consequently, service managers must acquire more users and increase users' satisfaction. When service managers adopt such strategies, they often struggle with the following issues, which are related to decision-making:

- Diversification of users' requirements
- Measures for the effectiveness of business strategies related to users.

2.3.1 Diversification of users' requirements

Users' requirements or preferences fluctuate due to content diversity. Even in web articles, there are many types of articles (*e.g.*, about daily news, public entertainment, trivia news, *etc.*). In addition, each user has his or her own favorite patterns. Users are sensitive to changes in a service. If a service adopts a strategy, that unintentionally affects a users' login, users tend to terminate a service. Additionally, many other organizations provide similar services. Consequently, rapid responses to changes in users' requirement and attitudes are crucial.

2.3.2 Measurements of the effectiveness of business strategies

Merely monitoring the business KPI (Key Performance Indicators), which is a reflection of the achievement of business goal, is insufficient to evaluate the effectiveness of business strategies. A case study of an organization indicates that organizations struggle measuring the effectiveness of strategies. A goal can be achieved even if the strategies are ineffective because measuring performance often influence employees' behaviors. This phenomenon is known as the Hawthorne effect and is common in software development [12].

Effective target users cannot be defined because evaluating the effect of strategies on users and a goal is challenging. In practice, the cost effectiveness of strategies must also be considered. To achieve a goal under resource constraints, the target users should be defined and resources should be allocated appropriately. Unfortunately, defining effective target users is burdensome.

2.3.3 Ideas to resolve problems

With regards to the first issue, Data-Driven Personas Development helps a service manager understand actual users of a service. However, Data-Driven Personas Developments proposed in previous work do not evaluate and revise personas to detect the changes in actual users.

Herein we propose ID3P to detect the user's change rapidly and evaluate strategies. ID3P includes the following ideas:

- Data-driven personas development to create personas that reflect actual users in a service,
- Iterative evaluations and revisions of personas to determine whether they are suitable for current users,
- A GQM+Strategies process to evaluate the effectiveness of strategies and achievement of a goal.

3. PROPOSAL

3.1 Overview

It is assumed that ID3P is applied over multiple iterations. Our proposal includes: (1) a quantitative evaluation and revision of personas developed through a data-driven construction approach and (2) a quantitative evaluation of business strategies or assumptions via the analysis of personas (Fig. 1).

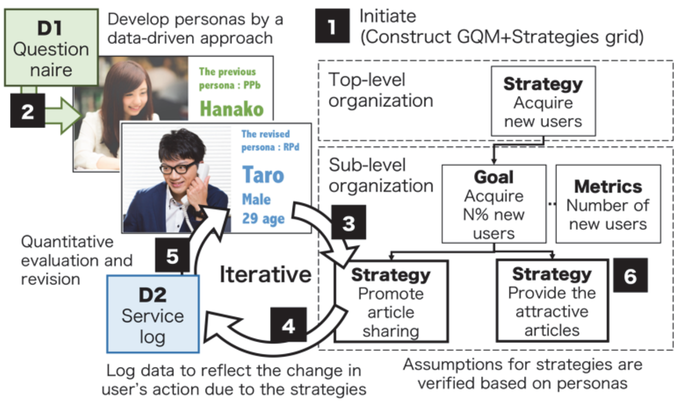


Fig. 1. Overview of an iteration in ID3P.

To cope with an unpredictable user changes in ID3P, personas are verified quantitatively in each iteration. Additionally, to evaluate the effect of business strategies on usability quantitatively, GQM+S is integrated into above data-driven persona iteration because GQM+S itself includes an iterative and data-driven improvement of strategies like Agile.

Thus, ID3P involves the following steps and 3 to 6 steps are applied in each iteration:

1. Initiate

2. Develop personas by a data-driven construction approach
3. Deduce the assumptions to plan strategies
4. Plan and execute strategies
5. Revise personas
6. Verify assumptions and evaluate strategies

Fig. 2 shows the procedures' correspondence between GQM+S and ID3P. ID3P extends or includes each step in GQM+S.

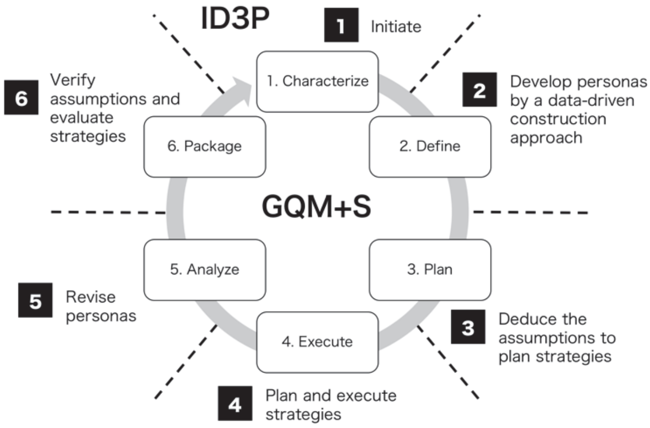


Fig. 2. Procedures' correspondence between GQM+S and ID3P.

3.2 Step 1: Initiate

In this step, a GQM+S model is constructed to quantitatively evaluate the strategy. Additionally, metrics are selected according to the definition in ID3P to develop personas. Fig. 3 depicts the relationship among the attributes in ID3P.

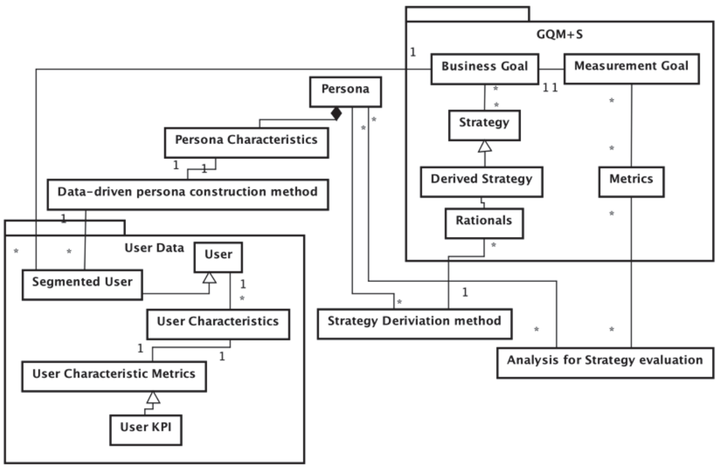


Fig. 3. UML class diagram-based relations of attributes in ID3P.

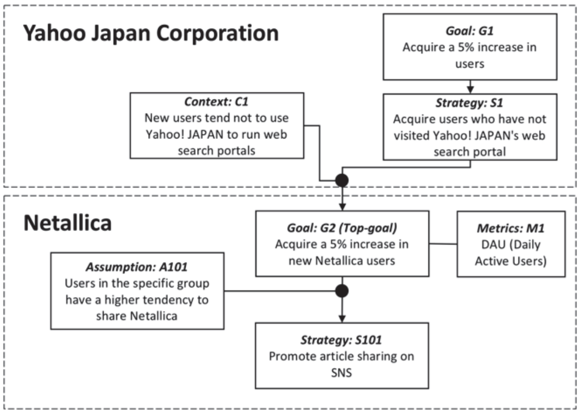


Fig. 4. GQM+S model of the Netallica case study.

3.2.1 GQM+Strategies

To quantitatively evaluate a goal, all goals should be measurable. Therefore, the relationship between goal strategies should be clarified as a GQM+S model in ID3P. A goal is often related to a higher-level organization's goal. For example, in Fig. 4 the top-level organization's goal G1 is "*Acquire a 5% increase in users*", while "*Acquire a 5% increase in new Netallica users*" is defined as the service goal G2.

3.2.2 User characteristic metrics

User characteristic metrics are used to develop a persona and are independent of GQM+S. These metrics correspond to each user and must be reflected in an action of a user on a web service. The service can track the logged-in user's click points on web page. In this case, a user's click log of an item or a coordinate in each web page can be defined as a user characteristic metric because each click log corresponds to the user's action on the web service.

3.2.3 User characteristic metrics

Metrics that reflect each user's satisfaction, effectiveness, or other usability aspects are defined as the user KPI. Analysis of user KPI evaluates strategies via the relationship between users' intention and GQM+S metrics. For example, because the number of logins reflects each user's intention to use a service, it can be categorized into user KPI.

3.3 Step 2: Develop Personas

In this step, personas are developed by a data-driven construction approach. In ID3P, we assume that the user characteristic metrics are relatively large or big data. There are several reasons why personas can be constructed from user characteristic metrics.

- User characteristics metrics reflect the users' behaviors.
- As previous work reported, metric patterns are derived from the user characteristic metrics by data mining techniques.

- Metric patterns are summaries of users' behaviors, so users' behaviors can be derived from such metric patterns.

Some common patterns are derived by clustering click logs, which is a user characteristic metric. When a pattern has a high frequency of clicks on the help page, this pattern can be defined as the action of watching the help page.

3.4 Step 3: Deduce Assumptions to Plan Strategies

Assumptions to plan strategies are derived based on the personas' characteristics and are represented as attributes of GQM+S. ID3P assumes the following relationships between the behavior of a persona, actual users, and other metrics are assumed:

- Each user corresponds to a user's behaviors derived from the user's characteristic metrics.
- Each user must also correspond to his or her own user KPI.

In ID3P, the difference in the user characteristic metrics is helpful to plan strategies. For example, the intention to use a service can be measured by the login count indirectly. When one persona has longer login count than the others, the reason for the difference can be assumptions for an effective strategy to promote user's login and can be assumed through quantitative analysis of user characteristic metrics.

While taking actions based on planned strategies, the assumptions are validated via the user characteristic metrics. For example, the number of clicks of the share button on a web service, which is a user characteristic metric, can validate the assumption that some personas tend to recommend the web service to others more than other personas.

3.5 Step 4: Plan and Execute Strategies

Strategies are planned based on derived assumptions and integrated into the GQM+S model. Although ID3P assumes that strategies are planned manually, ID3P can evaluate planned strategy based on persona changes and assist in planning precise strategies.

3.6 Step 5: Revise Personas

In this step, the personas in the previous step are evaluated and revised to understand the change in users. In practical situations, reconstruction of personas in each iteration is time-consuming. To restrain the time and cost to reconstruct personas, ID3P quantitatively evaluates the personas to determine whether revision is necessary. Moreover, this step is employed to help to plan business strategies based on precise personas. Though the evaluation of personas in ID3P is proceeded manually, our contribution is the framework to detect the changes in personas when planning business strategies. Therefore, this step must not be omitted. Fig.5 shows the persona revision procedure in ID3P. The evaluation and revision of personas involves the following steps:

1. Build a classifier from the users' data used to develop personas as a label. This persona is defined as the previous persona.

2. Predict a suitable previous persona for every user in this service iteration.
3. Discuss the classification results based on quantitative criteria.
4. Develop personas from users' data in a service iteration when the classification result is unsuitable. The developed personas are defined as the revised personas. If the personas are free of issues, define the previous personas as the revised persona.

If personas are unsuitable for actual users, they should be reconstructed from current users (Fig. 5, right). For example, when previous personas are developed from the click logs in the previous step, every user in a given iteration can be categorized into one of the personas determined by the classifier, which is built based on the previous persona (the training input is the click log and labels are the previous personas). After the classification, the criteria for unsupervised clustering is calculated. If the results become worse, it means that the labeling of the previous personas is not suited for the current users. Therefore, new personas should be developed from latest user data.

On the other hand, when the results of the criteria meet a minimum threshold, the persona can be used in the next iteration (Fig. 5, left). In this situation, the change in the user KPI of each persona can be helpful to understand the change of a persona's attitude towards a web service. Employing these steps allows the change in the user's behaviors and attitudes to be detected.

The suitable unsupervised clustering criteria depend on the types of user characteristic metrics. To avoid overfitting, the criteria should become worse as the number of personas is increased. If thresholds are difficult to determine definitely, they should be determined empirically through iterations.

Additionally, the clustering criteria also depends on the clustering algorithm. ID3P assumes that the user characteristic metrics or clustering algorithm is changed to construct more suitable personas for current actual users. In this case, persona validity should be discussed qualitatively. However, the continuous application of ID3P can validate personas quantitatively.

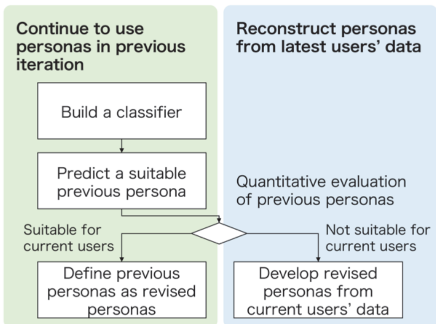


Fig. 5. Persona revision procedures in ID3P.

3.7 Step 6: Deduce Assumptions to Plan Strategies

The assumptions derived in the step 3 are verified by analyzing the revised personas. Additionally, a strategy is quantitatively evaluated based on the personas, and the GQM+S model is improved. ID3P extends strategies improvement in GQM+S with respect to analysis about personas' attitudes.

ID3P verifies assumption to plan and quantitatively evaluate strategies based on the changes in the personas. In GQM+S, the goals and strategies are analyzed and evaluated with respect to following aspects [21]:

- Feasibility and suitability of goals (*e.g.*, scope, magnitude, time frame, *etc.*)
- Validity of the relationships between goals and strategies
- Effectiveness and sufficiency of strategies
- Validity of assumptions

ID3P's improvement of a business goal is an extension of an evaluation from the following two perspective in GQM+S:

- **Validity of assumptions:** ID3P includes GQM+S practices. It is assumed that every strategy is derived from the upper-level goal based on assumptions or context. When an assumption is wrong, strategies based on it may also be incorrect. To avoid keeping inappropriate strategies, ID3P provides assumption verification based on personas.
- **Effectiveness and sufficiency of strategies:** Ineffective strategies are planned when their basis is unsuitable. When necessary assumptions are missed or the definition of a sub-goal is inappropriate, a strategy effectiveness is suspicious. Additionally, when metrics selection is incorrect, goal achievement cannot be determined, making it difficult to refine strategies. In ID3P, strategies are also evaluated via the correlation between GQM+S metrics and user characteristic metrics to assist in efficient achievement of the top-goal.

ID3P helps managers to improve the GQM+S model based on personas. After the strategies are evaluated, the GQM+S model is revised. The GQM+S practice includes the following four activities to improve the GQM+S:

- Replace goals related to the improvement with goals necessary to maintain the current level of performance
- Modify the attributes of existing goals or strategies
- Remove obsolete goals or strategies
- Add new goals or strategies

ID3P helps managers tackle the above activities through persona analysis, assuming that the top-goal is related to users (*e.g.*, user acquisition or improvement of users' satisfaction). Below are examples of GQM+S model improvement in ID3P:

- **Modify the goals or strategies:** Persona analysis can identify personas that influence a strategy to define more suitable target users.
- **Remove the goals or strategies:** Persona analysis can clarify inaccurate assumptions. Therefore, goals and strategies based on incorrect assumptions can be removed.
- **Add new goals** Persona analysis can identify preferences of personas to define new assumptions and strategies.

Non-verified assumptions should be held and reassessed in the next iteration. For example, when a user's intention to recommend a web service is aligned with the assumption in the previous step, but the difference is insignificant, the assumption should be also verified in the next iteration.

The effectiveness of a strategy is discussed based on the relationship between the change in the user characteristic metrics and the metrics for a business goal. In ID3P, the metrics in GQM+S can measure the achievement of every business goal. The effect of strategy is evaluated by the relationship between a user behavior and the metric of a business goal. For example, when the number of daily active users and a persona's login time are correlated, it can be hypothesized that strategies improving the login time are effective for user acquisition.

4. CASE STUDY

4.1 Overview of the two case studies in Netallica

ID3P is the first data-driven persona development method that includes quantitative evaluation and revision of personas to understand users' changes. To answer the following research questions, we employed two case studies involving Netallica, which is a service in Yahoo! JAPAN:

RQ1: Can personas' characteristics in ID3P derive business goals and strategies? To answer RQ1, we tried to derive business goals and strategies based on reconstructed personas in the first case study and reused personas in the second case study.
RQ2: Can ID3P verify personas constructed via a data-driven approach? To answer RQ2, personas were verified in two case studies. The first case study assumes that persona changes are drastic and the second case study assumes that they are minor.
RQ3: Can the assumptions about personas be verified quantitatively via ID3P? To answer RQ3, strategies were verified through identifying incorrect assumptions. Assumption verification is based on analyzing persona changes. Therefore, we employed two case studies to verify it in two types of persona changes.
RQ4: Can strategies and goals be evaluated based on personas quantitatively? To answer RQ4, strategies and goals were evaluated based on personas changes in two case studies. Additionally, we improved the GQM+S model based on evaluation in two case studies.
RQ5: Does revising personas aid in understanding users and planning strategies? To answer RQ5, two types of persona revision were verified through two case studies. Additionally, we discussed about the persona revision's contribution to improvement of the GQM+S model.

Netallica has tried to take an action to acquire new users based on web service metrics in the PDCA (Plan-Do-Check-Action) cycle. Although Netallica constructed personas from a questionnaire survey through a qualitative approach, it could not confirm that the personas represent current users on the web service because users can easily quit the service. Moreover, the Netallica wanted to identify the effective target users and to plan the effective strategies to achieve the top-goal efficiently.

To apply ID3P to Netallica, two types of data sets were used: ND1) a questionnaire survey implemented by a research company and ND2) the log data of users on the service. The questionnaire was completed by some Netallica users, including some user groups. Participants were randomly selected by a research company. Table 1 shows the details of the questionnaire. Users' log data is composed of actual log data of users who visited Netallica in Oct 2016 (Table 2). Netallica managed each user's history of articles through user accounts adequately. Article information contained an article ID, release date, and the name of the original website. Although the size of datasets cannot be described, the datasets were sufficient to apply machine-learning or statistical method in following sections. We used Python and R to analyze the data. We selected a machine-learning method and evaluation metrics according to the following criteria so that a service manager, who is unfamiliar with machine-learning, can apply ID3P to a service:

- It is implemented by a major library and anyone can apply it.
- It can be applied without complex performance tuning.

Table 1. Details of the questionnaire survey (ND1).

No.	Contents	Possible values
Q1	Interest in each article category	1 (favorite) to 5 (not a favorite)
Q2	Frequency of reading articles in each category	1 (usually) to 4 (never)
Q3	Intention to use Netallica continuously	1 (intend to) to 5 (never intend to)
Q4	Intention to recommend Netallica to others as a percentage	100 to 0 in 10% increments

Table 2. Details of the log data of users on a service (ND2).

Item	Contents	Possible values
User id	Number used to identify a user	Integer
Article category	Categories with articles that users read in October	List of categories (string)
Article	Articles users read in October and its information	List of article IDs with dates and names of the original websites (string)
Count of shares	Number of shares by users on Twitter and on Facebook	Count on Twitter and that on Facebook (two integers)

4.1.1 Overview of the first case study

ID3P assumes the two types of persona revisions: persona reuse and persona reconstruction (Fig. 5). The first case study verifies the effectiveness of the quantitative evaluation in ID3P to determine whether personas should be reconstructed. Because Netallica's personas are constructed from questionnaires and they are not based on actual users, the original personas may not represent current actual users.

The first case study, which answers RQ1, RQ2, RQ3, and RQ5, involves the following steps:

1. Construct a simple GQM+S model of Netallica to clarify the goal.
2. Construct personas from a questionnaire survey through a data-driven approach (previous personas).

3. Evaluate personas in step 2 by classifying users' log data into personas and revise personas.
4. Verify the assumption to plan a strategy based on changes in personas.

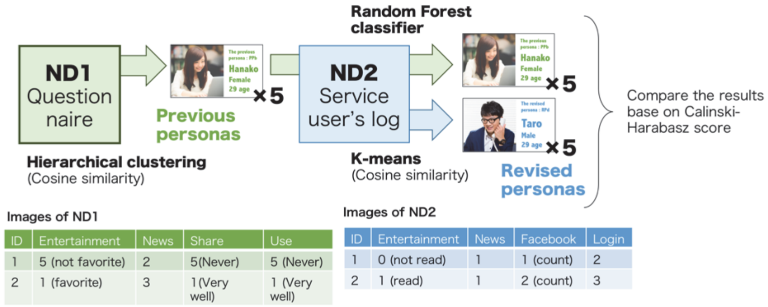


Fig. 6. Overview of persona revisions in the first case study.

Fig. 6 overviews the persona revision. Previous personas are constructed through hierarchical clustering. After quantitatively evaluating previous personas, we define personas, which are constructed from current users' data by K-means, as revised personas.

4.1.2 Overview of the second case study

The second case study validates the quantitative evaluation of business strategies based on persona reuse. ID3P continuously improves the business strategies based on the revision of personas. Although we tried to validate the processes to improve the GQM+S model based on changes in personas, we could not acquire the additional actual users' data. Hence, we divided users' log data into more short-term iterations based on the article release date. We defined each iteration as one week because we assumed that the improvement cycle of strategies depends on the working day in an organization. We define an Article as a user characteristic metric because the Article category, which was a user characteristic metric in the first case study, is not related to the release date (Table 2).

This case study answers RQ2, RQ4, and RQ5. Fig. 7 overviews the second case study, which involves the following steps:

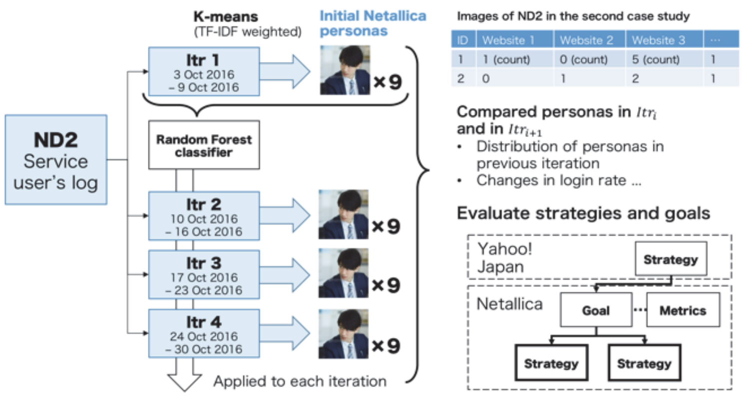


Fig. 7. Overview of persona construction in the second case study.

1. Construct personas from user's log in the first iteration.
2. Classify users into personas and determine that they are suitable quantitatively.
3. Evaluate the effectiveness of strategies quantitatively and improve GQM+S model based on the results.

4.2 Case Study to Evaluate a Persona Revision

4.2.1 Initiate

First, we constructed a simple GQM+S model of Netallica (Fig. 4). In this case study, the target organizations are the Netallica team and Yahoo Japan Corporation. The top organization is Yahoo Japan Corporation. The Netallica team's goal is to acquire new users, who do not use Yahoo! JAPAN to run web search portals. We defined the Yahoo Japan Corporation's strategy as S1, "*Acquire users who have not visited Yahoo! JAPAN's web search portal*" and Netallica team's goal as G2, "*Acquire a 5% increase in new Netallica users*".

4.2.2 Develop personas

Second, we developed personas from ND1 (Table 1). Netallica categorizes articles into 11 categories (public entertainment, news, trends, love, beauty, food, travel, movies & music, animation, humor, and trivia news). In this case study, user characteristic metrics are the responses to Q1 and Q2 (Table 1) because they are related to reading articles, which is an action in Netallica web service. Additionally, user KPI are the answers to Q3 and Q4 (Table 1) because they are related to users' attitudes toward Netallica.

To derive the persona's characteristics, we applied hierarchical clustering based on the cosine similarity to the Q1 answers. To simplify the categorization of personas, we identified five clusters. Each cluster was defined as a persona (Table 3 previous persona). Each attribute in Table 3 is assumed from the response distribution to related questions, which are answered by users in each cluster.

Table 3. Previous Netallica persona.

Persona	Goal	Use intention	Recommend intention
PPa	High *beauty* and low *animation*	Not particularly	Relatively high
PPb	Almost high, particularly *beauty* and *trip*	Relatively high	Relatively high
PPc	High *news*	Relatively low	Low
PPd	Almost high but low *love* and *beauty*	Relatively high	Relatively low
PPe	All categories	High	High

4.2.3 Derive assumptions and plan strategies

We tried to derive assumptions (step 3 in Fig. 1) and to plan strategies (step 4 in Fig. 1). We assessed the Q4 distribution of the users for each persona. Some personas show higher intentions than others. Additionally, the questionnaire asked the intentions of using SNS (Social Network Service). The responses of users in a specific group are higher than those of the other users (Assumption A101 in Fig. 4). The Mann-Whitney U test was used to determine if the difference is significant. Although confidentiality prevents sharing of group details, Yahoo Japan Corporation and Netallica have not increased users in this group. Consequently, we derived strategy S101 to "*promote the sharing articles on SNS*" to acquire such users.

4.2.4 Revise personas

We tried to detect change of the personas based on a quantitative evaluation. First, we classified ND2 into the previous personas using a RandomForest classifier. In this case study, the training data of the classifier was each user's answer to Q2 in ND1 (Table 1) and the input data was the categories with articles that users read in ND2 (Table 2). To match the scale of ND2, the answers of 1 or 2 to Q2 were transformed into 1, while responses of 3 or 4 were converted into 0. Because ND2 was missing an *animation* category, the answers for the other 10 categories to Q2 were used as training data. After the classification, we evaluated the results quantitatively. In this case study, the Calinski-Harabasz score was adopted as the evaluation criterion. Let

$$w_k = \sum_{q=1}^{k} \sum_{x \in C_q} (x - c_q)(x - c_q)^T \tag{1}$$

$$B_k = \sum_q n_q (c_q - c)(c_q - c)^T \tag{2}$$

The Calinski-Harabasz score was calculated by

$$s(k) = \frac{B_k}{W_k} \cdot \frac{N-k}{K-1} \tag{3}$$

where N is the number of data and k is the number of clusters. W_k denotes the variance of each cluster and B_k denotes the variance between clusters. When every cluster can be identified well, W_k becomes smaller, while B_k becomes larger. Additionally, the Calinski-Harabasz score is restrained by the number of clusters k to adjust overfitting. Therefore, the larger the Calinski-Harabasz score is, the better the clustering result. We calculated the Calinski-Harabasz score of ND2 based on the previous personas. Table 4 (C1) shows the results. This calculated score is relatively small and is the static part in formula. Additionally, we suspected that the previous personas may be insufficient for actual users because they were constructed from some users who were conscious of using Netallica.

This analysis suggests that the previous personas are unsuited as representatives of the users of the service. Therefore, we developed new personas from the Article category in ND2 (revised persona). We applied *k*-means based on the Jaccard distance to Article category in ND2 and identified five clusters. To compare the revised personas with the previous personas, the Calinski-Harabasz score of ND2 based on the revised personas was also calculated (Table 4, C2). The revised personas produced better results than the previous personas. Moreover, we calculated the Calinski-Harabasz score of ND1 based on the previous and revised personas (C3 and C4 in Table 4). To consider the impact of preprocessing, we also calculated the Calinski-Harabasz scores of users completing the questionnaire survey before preprocessing the responses to questionnaire based on the previous and revised personas (C5 and C6 in Table 4). These calculated scores show that the revised personas are suitable even for users completing the questionnaire survey.

Table 5 summarizes the revised personas. In this step, the goal of the personas (preference of categories) and login count as user intention to use Netallica are concluded.

Table 4. Calinski-Harabasz score in the first case study.

Case	Dataset type	Persona type	Number of personas	Score	Baseline and Increase
C1	Service(ND2)	Previous	5	24431	–
C2	Service(ND2)	Revised	5	105659	Increase from C1
C3	Survey(ND1)	Previous	5	7.0866	–
C4	Survey(ND1)	Revised	5	20.686	Increase from C3
C5	Survey(ND1) (before preprocessing)	Previous	5	12.580	–
C6	Survey(ND1) (before preprocessing)	Previous	5	15.489	Increase from C5

Table 5. Revised Netallica persona.

Persona	Goal	Use intention	Recommend intention
RPa	Relatively high *Humor*	Low	Relatively high
RPb	High *news*	Relatively low	Not particularly
RPc	High *public entertainment*	Low	Not particularly
RPd	High *trivia news*	Low	Relatively high
RPe	High *trivia news* and *news* Additionally, relatively high other categories	Relatively high	Not particularly

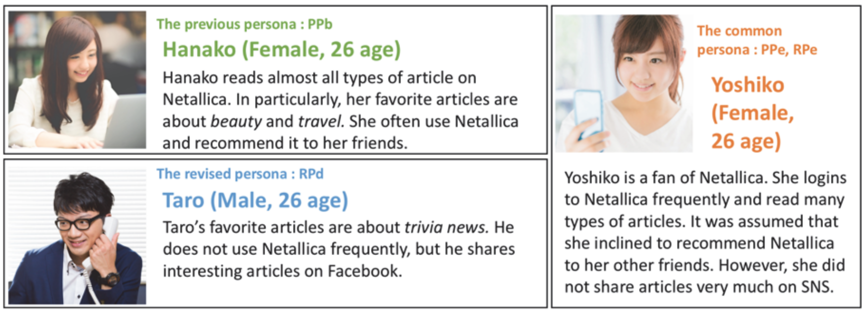
4.2.5 Verify assumptions

This case study was designed to verify that some personas are more willing to share articles on SNS than others (A101 in Fig. 4). Our assumption A101 is that some personas in a specific group are more willing to share of articles on SNS than personas in other groups. Due to the size restriction of the users' log data, personas in a specific group could not be compared to those in other groups. Instead, we determined the statistical difference between personas because the method to determine the difference is similar to the one used to verify the original assumption.

First, we showed each persona's distribution of the share count on Facebook (Table 2). We applied the Kruskal-Wallis test to the persona's share count and determined if a significant difference exists. To specify which personas produce a significant difference, we applied Dunn's test, which is a multiple comparison statistical test, to the share counts. Persona CPA and RPd differ significantly from the others with regard to shares on Facebook and have a slightly higher means than RPb. Therefore, we hypothesize that RPa and RPd are slightly more inclined to share articles on Facebook than the other personas (Table 5 Recommend intention).

To identify the difference between the previous and the revised personas, we also compared Tables 3 and 5 qualitatively. Persona PPc and RPb have the same goal (*News* category). Moreover, both personas PPe and RPe read many types of articles, but unlike

PPe's, RPe's recommendation intention is not particularly high. Fig. 8 depicts some persona descriptions to highlight the difference between the previous personas and the revised ones. Yoshiko depicts an example of PPe and RPe. In contrast, the previous personas do not include the same personas as RPa, RPc, and RPd with regard to the goal of personas. Additionally, many revised personas do not intend to login into Netallica as frequently as the previous ones. Consequently, it is assumed that many personas are not in the habit of using Netallica. However, this assumption is not well considered in the planned strategies to improve existing user's satisfaction because the previous persona's intention to use Netallica is so high. Therefore, in the next iteration, we should try not only to promote the sharing of user-preferred articles, but also to provide attractive articles to promote the login of existing users.



Above three pictures are downloaded from Pakutaso(www.pakutaso.com). Please accept the Pakutaso's user policy and download by yourself from Pakutaso when you use above pictures on your material.

Fig. 8. Three exemplary personas in the first case study.

4.3 Case Study to Assess the Continuous Application of ID3P

4.3.1 Develop personas

First, we constructed personas from the users' history of articles from 3 Oct 2016 to 9 Oct 2016. Netallica provides articles from other websites. The article content depends on the original website. Each article is connected with the released date and the original website (Table 2). In this case study, we defined users in each iteration as users who read articles which is released in that iteration. We also defined the login count as the count of the released date of an article which users read in each iteration.

To construct personas, we defined the frequency of the original website of an article as a user characteristic metric. We also defined login count as user KPI because it can reflect on user's satisfaction. As a preprocessing measure, we applied TF-IDF to each user's count of the website. Generally, TF-IDF is used in natural language processing tasks to distinguish less frequent terms in a document. In this case study, the number of articles differs among the websites. To distinguish less frequent websites, we considered each user as a document and the website name as a term in a document.

After preprocessing, we applied the k-means clustering algorithm to TF-IDF weighted count of websites. We defined nine initial Netallica personas based on the clustering results (Table 6). Each persona's goal in Netallica is its favorite type of article, while the intention to use is a persona's willingness to use Netallica. The goal is based on frequent websites for each persona. Intention to use in Itr_1 is also assumed based on the login count distribution for each persona. For simplicity, we divided personas' login counts into four categories (High, Relatively high, Relatively low and Low).

Table 6. Initial Netallica personas.

Persona	Goal	Use intention	Recommend intention
NPa	Various type of categories, especially news about Japanese pop stars in *public entertainment*	High	High
NPb	*Public entertainment*, especially Japanese TV personality	Relatively low	High
NPc	News about animals in *trivia news*	Relatively high	Relatively high
NPd	*Public entertainment* and *news*	Relatively low	Relatively high
NPe	*Public entertainment* and *news* in foreign countries	Relatively low	Relatively low
NPf	Legal topics in *trivia news*	Low	Low
NPg	Japanese TV personality (especially actresses) in *public entertainment* and others in *news*	High	High
NPh	Lifehack in *trivia news*	Low	Low
NPi	*Trivia news* and *news*	Relatively high	Relatively high

4.3.2 Plan and execute strategies

We tried to plan a strategy based on the initial Netallica personas and integrate the strategy into the GQM+S model (Fig. 9). We assumed that users with a higher satisfaction will visit Netallica again (A201 in Fig. 9). The top goal is the same as Fig. 4, but goal G201, "*Improve the satisfaction of users based on personas*" is defined as the subgoal of G2.

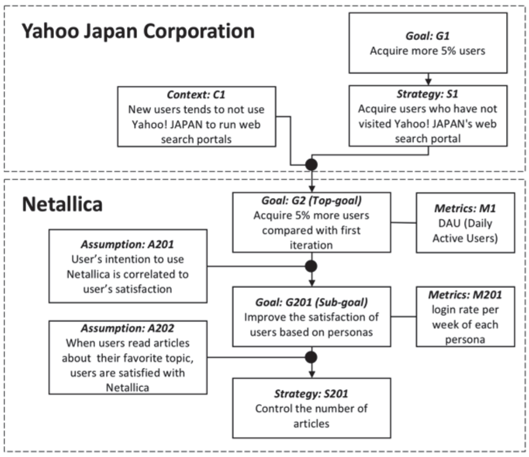


Fig. 9. Initial GQM+S model of the Netallica case study.

While monitoring users' action in the case study, Netallica did not convert a specific strategy into an action. In this case, we assumed that the number of articles read influences user's satisfaction (A202 in Fig. 9) and the dynamic change in the number of articles is a certain strategy (S201 in Fig. 9).

4.3.3 Revise personas

To evaluate the initial Netallica personas, we classified users in other iterations into the initial Netallica personas and calculated the criteria score. First, we built a Random-Forest classifier using each user's count of the website as training input and the Netallica personas as labels. We used the rate of each original website in the articles that each user read as the input of the classifier. After building a classifier, we classified users in each iteration into Netallica personas and calculated the Calinski-Harabasz score (Table 7). Then the Calinski-Harabasz scores were calculated by the TF-IDF weighted value.

Table 7 shows that the Calinski-Harabasz scores gradually increase. We assumed that the personas in Itr_1 are also suitable for representations of users in other iterations. Moreover, Netallica did not convert specific strategies into action in Oct 2016. Therefore, large changes in the users' requirements might not happen. We assumed that we could apply the same personas to users in other iterations.

Table 7. Details of iterations and the Calinski-Harabasz score.

Iteration	Duration	Number of personas	Score	Baseline and Increase
Itr_1	2016/10/3 ~ 2016/10/9	9	42336.018	–
Itr_2	2016/10/10 ~ 2016/10/16	9	50817.354	Increase from Itr_1
Itr_3	2016/10/17 ~ 2016/10/23	9	61080.096	Increase from Itr_2
Itr_4	2016/10/24 ~ 2016/10/30	9	67803.591	Increase from Itr_3

4.3.4 Evaluation of a goal and strategy

We verified the evaluation of a strategy through ID3P in the following steps:

1. Evaluate the achievement of top-goal (G2) and sub-goal (G201).
2. Specify the effect of the strategy (S201 in Fig. 9) by comparing the persona characteristics in the previous iterations to the current iteration.
3. Improve the GQM+S model based on the evaluation.

ID3P includes an evaluation of business goals and strategies based on GQM+S. Business goals are evaluated from the top-level goals. After determining goal achievement, the effect of each strategy to the related goals is discussed. First, we tried to determine whether Netallica's top-level goal G2, "*Acquire 5% more users compared with first iteration*" was achieved. To evaluate the achievement of G2, we checked the average number of user logins by iteration as DAU (Daily Active User). Although the actual DAU cannot be described, but Itr_4 has the highest average number of user logins among the four iterations. Because Itr_4 is 5% higher than the average in Itr_1, G2 was achieved. After determining the achievement of G2, we evaluated the achievement of a sub-goal G201, "*Improve the satisfaction of users based on personas*" by analyzing the changes in each persona's login count distribution (Table 9). Persona NPd's average login count gradually decreases and NPd's login count is relatively low among the initial Netallica personas. Additionally, we compared the average number of read articles for NPd in Itr_3 with NPd in Itr_4. These differences suggest that NPd's intention to use Netallica declines. Therefore, we suspect that increasing or decreasing the amount of articles affects NPd's intention to use Netallica.

To identify the effect of articles on users, we assessed the effectiveness of a strategy S201, "*Control the number of articles*" quantitatively. We applied a binomial regression model to the user's login count in Itr_4 assuming that each user's login count follows a binomial distribution. We also assumed that the login possibility, which is a parameter of the binomial distribution, could be estimated by each user's count of articles released from the top ten most frequent websites. The user's count was estimated using Rstan, which is the R distribution of a statistical modeling tool with MCMC sampling (Table 10). We verified the convergence of MCMC sampling by checking $Rhat < 1.1$, which is a criterion of convergence. The regression results suggest that $Site_2$, $Site_3$, $Site_5$ and $Site_8$ influence users' login because the means in Table 10 are larger than the other convergence of websites. Additionally, we checked the average number of articles that each website provided to Netallica by iteration. More articles are provided in Itr_4 compared to Itr_3. Additionally, $Site_2$, $Site_3$, $Site_5$ and $Site_8$ also provide more articles. We cannot immediately determine if the effect of articles results in users' visits to Netallica. However, we suspect that merely increasing the number of articles is not a suitable strategy for NPd. Consequently, we assume that the sub-goal G201 is not achieved and Strategy S201 did not contribute to achievement of G201 in Itr_4.

The top-goal G2 is achieved, but not the sub-goal G201. We suspect that some goals and strategies in Fig. 4 are inadequate due to the missing assumptions or an incorrect definition of the sub-goal.

Table 8. Ratio of the number of a persona to the previous iteration.

Iteration	NPa	NPb	NPc	NPd	NPe	NPf	NPg	NPj	NPi
Itr_1	–	–	–	–	–	–	–	–	–
Itr_2	1.053	1.125	0.791	0.817	1.490	1.451	0.730	2.066	1.294
Itr_3	1.055	2.181	0.891	1.034	0.662	1.993	1.070	1.500	1.423
Itr_4	1.161	1.495	1.565	0.578	1.152	1.070	0.898	0.947	1.056

Table 9. Each persona's average login count by iteration.

Iteration	NPa	NPb	NPc	NPd	NPe	NPf	NPg	NPj	NPi
Itr_1	2.783	1.687	1.967	1.776	1.768	1.402	2.183	1.422	1.930
Itr_2	2.647	1.770	1.942	1.670	1.810	1.462	2.001	1.519	1.819
Itr_3	2.684	1.865	1.797	1.753	1.627	1.332	1.943	1.428	1.932
Itr_4	2.841	2.061	1.877	1.616	1.717	1.371	2.065	1.462	1.831

Table 10. Result of the top 10 websites' coefficient distribution in MCMC sampling.

Website	Mean	2.5%	97.5%	Rhat
$Site_1$	0.33	0.28	0.37	1
$Site_2$	0.53	0.49	0.58	1
$Site_3$	0.50	0.45	0.54	1
$Site_4$	0.47	0.42	0.52	1
$Site_5$	0.50	0.43	0.56	1
$Site_6$	0.33	0.26	0.41	1
$Site_7$	0.37	0.30	0.45	1
$Site_8$	0.50	0.42	0.58	1
$Site_9$	0.41	0.34	0.49	1
$Site_{10}$	0.40	0.30	0.50	1

4.3.5 Improvement of the GQM+S model

To specify the issues of GQM+S in Fig. 9, we identified the relationships between personas in the iterations. Table 11 shows the rate for select personas in Itr_4 based on each persona in Itr_3. For example, almost 40% of users classified into NPd in Itr_3 are classified into NPa in Itr_4. NPa is defined as the persona with a high intention to use Netallica in Table 6. Table 9 shows that NPa is satisfied better in Itr_4 than in Itr_3. Moreover, Table 8 also shows that the number of NPa's is larger in Itr_4 than Itr_3. The results suggest that some users, who are classified into NPd in Itr_3, are satisfied with Netallica. Consequently, NPd users in Itr_3 are classified into NPa in Itr_4 because they begin to read more articles. We assume that this relationship is suitable because NPa and NPd prefer the same category of articles (*public entertainment*). We also assume that the results in Table 11 are suitable because NPf and NPh, which have low intentions to use Netallica in Itr_3, do not visit Netallica much in Itr_4. Due to confidentiality, we cannot show how many users classified into a persona in an iteration did not visit to Netallica in the next iteration. However, users classified into NPf or NPh visited to Netallica less frequently than other personas. This result is consistent with NPf and NPh's low login count in Table 9.

To reflect these results on personas, we revised each persona's intention to use Netallica (Table 6) and defined the description of some personas (Fig. 10). We described examples of NPa (Tetsuya), NPd (Kenji), and NPh (Yoko). NPa (Tetsuya) and NPh (Yoko) do not differ from those in Itr_1, while NPd (Kenji) differs from that in Itr_1. Additionally, Tetsuya and Kenji are defined as examples of effective targets, which have high intentions to use Netallica, while Yoko is defined as an example of a persona without a high intention.

Through this revision of personas, we assumed that there are two issues in Netallica's GQM+S model. The first issue is the missing of assumptions about the relationship in personas and vagueness of the target user in G2 and G201. Ineffectiveness of G201 indicates that increasing the number of all personas will not necessarily contribute to the achievement of upper-level goal G2. Moreover, in practical decision-making, the balance between cost and benefit should be considered. Therefore, G201 should be improved and the target persona should be specified to concentrate the resource to achieve goal G2.

The second issue is inappropriate selection of metrics. Table 6 and Fig. 10 show that some personas' satisfactions are misunderstood by judging based on the login count in

Itr_1. For example, NPd is defined as a persona with a relatively low intention to use Netallica in Itr_1 based on the login rate (Table 9). Due to confidentiality, we cannot show the persistent rate of each personas. However, actually, NPd is a persona with a relatively high persistency rate from the previous week. Additionally, the relationship between NPa and NPd suggests that the persistency rate is a more suitable metrics to understand users' intention to use Netallica.

Above three pictures are downloaded from Pakutaso(www.pakutaso.com). Please accept the Pakutaso's user policy and download by yourself from Pakutaso when you use above pictures on your material.

Fig. 10. Three exemplary personas in the second case study.

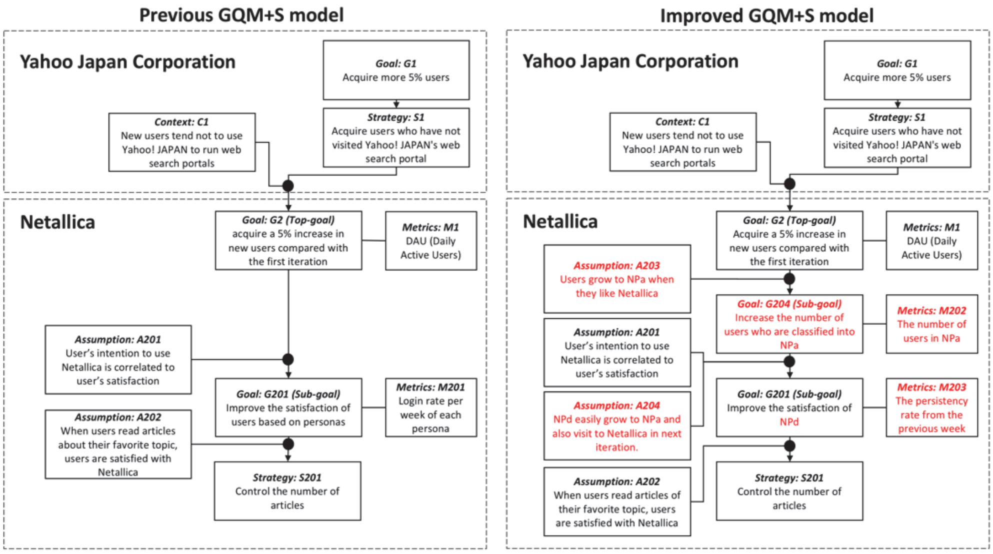


Fig. 11. GQM+S model of the second Netallica case study.

Consequently, we improved the GQM+S model based on above analysis. Fig. 11 shows the improved model and the differences from Fig. 9. The improvements are:

- Based on results in Table 11, we implemented new Assumptions A203: "*Users grow to NPa when they like Netallica*" and A204: "*NPd easily grow to NPa and also visit to*

Netallica in next iteration".

- Based on A203 and the evaluation of G201, we defined a new goal G204: "*Increase the number of users who are classified into NPa*" because G204 will contribute the achievement of G2 more precisely.
- Based on A204, the target personas are specified in G201: "*Increase the number of users, who are classified into NPa*".
- Based on evaluation of G201 and results in Table 11, we defined metrics M203: "*The persistency rate from previous week*" as metrics for G201 because M203 will measure user's satisfaction more precisely.

Table 11. Rate of personas in Itr_4 based on each persona in Itr_3.

Label in Itr_3	NPa	NPd	NPf	NPh	Other
NPa	0.494	0.021	0.006	0.004	0.475
NPd	0.395	0.039	0.003	0.004	0.559
NPf	0.151	0.004	0.075	0.018	0.752
NPh	0.086	0.006	0.050	0.095	0.763

5. DISCUSSION

5.1 RQ1: Can personas' characteristics in ID3P derive business goals and strategies?

In the first case study, strategies can be derived based on changes in previous and revised personas. Additionally, reusing personas in several iterations can derive new goals and improve GQM+S.

In the first case study, we derived a strategy S101, "*Promote article sharing on SNS*" based on the assumption A101, "*Users in the specific group show higher intention to share Netallica*" in Fig. 4. We also planned a new strategy to improve user satisfaction through a qualitative comparison between the previous and the revised personas.

In the second case study, we derived a sub-goal G204, "*Increase the number of users who are classified into NPa*" and improved goal G201, "*Improve the satisfaction of NPd*" in Fig. 11. Both goals are based on A201 and A204, which are deduced from changes in personas.

These results suggest that both ad-hoc strategies to satisfy current users and strategies to contribute to the upper-level goal through ID3P can be derived. Therefore, the change or difference of personas is helpful in planning strategies.

5.2 RQ2: Can ID3P verify personas constructed via a data-driven approach?

Both case studies verified personas based on the Calinski-Harabasz score, which is a quantitative criterion.

In the first case study, we determined that the previous personas are unsuitable for current users. This result is reasonable because the questionnaire responders likely differ from most users in Netallica. Users, who answered the questionnaire, were gathered by an external research organization. The questionnaire respondents are conscious of using

Netallica. It is suspected that such users are heavy users of Netallica, and they are different from most of users. In this case, quantitative criteria indicate that revised personas are more suitable for current users than previous personas (Table 4).

On the other hand, in the second case study, we determined that personas in Itr_1 are suitable for users in other iterations. This result is appropriate because Netallica did not adopt any strategies that would lead to changes in the users' action. An example of such strategies would be to implement a new category or place a new advertisement. In this case, quantitative criteria also indicated that personas do not have to be reconstructed (Table 7).

Based on above two results, we believe that unsuitable personas can be verified through a quantitative evaluation in ID3P, leading to an appropriate revision of personas.

5.3 RQ3: Can the assumptions about personas be verified quantitatively via ID3P?

We can verify the assumptions about users' attitude based on changes in personas' characteristics and by deriving new assumptions even when personas are reused.

In the first case study, we were unable to confirm the assumption that personas in the specific user group have a relatively higher intention of offering recommendations on SNS than others due to restrictions of the data. However, we were able to verify the assumption that some personas are more willing to share articles on SNS than others. Both assumptions are similar because it is necessary to compare a persona with others. We show that some personas have a higher share count than others, and the difference in the distribution is a statistically significant. Therefore, the original assumption also can be derived via this method because a similar assumption can be verified.

In the second case study, we confirmed the assumptions by deriving new assumptions even if changes in personas' characteristics are not drastic. We derived missing assumption A204 and defined the target user to achieve G2 more successfully. In the previous GQM+S model (Fig. 11, left), we implicitly assumed each user requires the same amount of effort to acquire. This case study adopted the same personas in every iteration. A204 is derived from changes in the number of users, which are classified into each persona. A204 indicates that NPa can be acquired more easily than other personas. Therefore, we can verify assumptions by deriving supplementary assumptions even if the same personas are adopted.

5.4 RQ4: Can strategies and goals be evaluated based on personas quantitatively?

In the second case study, we evaluated the validity of S201 and G201. Moreover, we derived suitable strategies and goals to achieve G2 based on personas.

In the evaluation of business goals and strategies, we identified the effect of the number of articles on users' login rate through a regression (Table 10) and determined that S201 is ineffective to achieve G201. Although we did not immediately determine the effect of articles on users' visits to Netallica because the regression results do not indicate a causal relationship, we identified which strategies are implemented and which metrics evaluate the effect of strategies. For example, when an iteration increases the amount of articles from a specific website to satisfy a persona who always read articles from a specific website, the effect of this strategy can be evaluated by comparing the target persona's satisfaction with the one in the previous iteration. Therefore, we assume that the effectiveness of strategies can be evaluated by a more simple and explicit method using ID3P.

We found that sub-goal G201 and upper-level goal G2 are achieved. Based on changes in personas' satisfaction, we determined that G201 did not contribute to G2, allowing a more precise strategy to be implemented to clarify the business goal in Netallica. It means that ID3P provides an evaluation of the sub-level goal, implying that ID3P can assess G1, which is a goal in the upper-level organization.

Consequently, strategies and goals can be evaluated through ID3P.

5.5 RQ5: Does revising of personas aid in understanding users and planning strategies?

To understand users, we assumed that ID3P can provide reliable personas, which are reflected in current actual users through quantitative evaluations and revisions. Adding the new goals and strategies, replacing the metrics, and verifying the assumption can improve the GQM+S model. Thus, ID3P helps service managers to understand users and plan more precise strategies.

RQ2 shows that misunderstandings about the most current users based on biased personas can be avoided through a quantitative evaluation and revision of personas in ID3P. Previous personas in the first case study are biased toward heavy users of Netallica, leading to a misunderstanding about typical current users. This misunderstanding prevents planning precise strategies to acquire new users. In our first case study, we confirmed that biased personas are unsuitable for current users based on the evaluation in ID3P. Additionally, in second case study, we determined that personas are suitable if changes in users did not occur in the iterations. Therefore, we can construct reliable personas, which are helpful for understanding users through ID3P.

ID3P constructs reusable personas, avoiding unnecessary documentation. We believe that it will be helpful to maintain personas and rapid decision-making for not only businesses but also other purposes (software design *etc.*) based on the latest personas. In the first case study, we identified common personas in previous and revised personas (Fig. 8) and confirmed that the revision in a common persona is reduced to a change in satisfaction. The second case study demonstrated that personas can reflect actual users by revising persona's satisfaction (Fig. 10). Hence, personas do not have to be identified in every iteration, reducing the burden to manage persona documents. For the common interpretation of personas, it is necessary for service team members to make a persona document. However, we believe that documentation of personas is time-consuming because many attributes and experts' opinions are often needed. However, especially in agile software development, the priority of documentation is often lower than other practices. Therefore, reconstruction in each iteration is not suitable for practical situations. ID3P will be helpful in maintaining the latest persona document because a service team can determine whether personas' revision is minor or major through a quantitative evaluation.

With regard to planning strategies, we planned suitable strategies for the upper goal. In the first case study, strategy S101 is withdrawn by verifying the assumption to derive this strategy based on revised personas. Additionally, a new strategy is derived by the revised personas, which had a low priority when using the previous personas. In the second case study, we improved the metrics for G201, specified the target users of G201, and implemented more helpful goal G203 to achieve G2. In this case, Assumptions A203 and A204 are based on the persona attributes (Table 6) and the change of personas, which users are classified (Table 11). We assumed that ID3P contributes to specifying the appropriate users to achieve the goal by analyzing the changes in personas.

Consequently, ID3P, which includes quantitative evaluation and revision of personas, is helpful for understanding users even in practical situations, and strategies and goals can be planned by specifying the effect on users.

6. THREATS TO VALIDITY

6.1 Internal Validity

6.1.1 Clustering algorithm

In ID3P, the clustering method impacts the verification of personas based on the clustering criteria. In the case study, we changed the clustering algorithm from hierarchical clustering in step 2 to *k*-means in step 5 due to the limitations in the computation resources. In particular, the Calinski-Harabasz score has a relatively large value when *k*-means is applied because the evaluation function of *k*-means is the same as W_k in the Calinski-Harabasz score formula. A future case study will address this issue. Regardless, Tables 3 and 5 and Fig. 8 show that revised personas, which reflect the users who did not complete the questionnaire survey, seems to be reasonable.

Additionally, the clustering algorithm in ID3P impacts the analysis process because the clustering results affect the metrics distribution of each persona. Because the results of the *k*-means algorithm depend on the initial values, the results in the evaluation step are not always the same even when using the same data set. The unstable result in step 6 is verified by applying ID3P iteratively. When the result in step 6 is suspicious, the result is used as an assumption in the next iterations. The process allows suspicious assumptions to be verified because the impact of the change in the personas is small when change in the service environment is small over the short-term.

6.1.2 Training metrics

It is possible that the results are affected by the handling of the training data. In this case study, the training input of the classifier in the persona revision step was Q2: Frequency of reading articles in each category in Table 1. However, the metric used in the persona development step is Q1: Interest in each article category in Table 1. Although it can be assumed that there is correlation between Q1 and Q2, the impact on the results might not be ignored.

Tables 1 and 2 (col Possible values) shows the measurable degree of the difference for the users' action in ND1 and ND2. However, we believe that reconstruction of the previous personas is unsuited for practical situations, but the above issues can be solved in the next iteration by feedback about the format of the metrics.

6.1.3 Evaluation of strategy

ID3P includes an evaluation of the business strategy and improvement of GQM+S, which depend on knowledge of the analysis method and service situations. In the evaluation method, we verified the effectiveness of the analysis based on changes in the users, which is proposed in ID3P. The analysis method also affects the improvement result of GQM+S. However, the analysis may be too flexible to ensure that the same analysis in ID3P is effective in the other cases. In the future, we will conduct a case study in other domains to verify effectiveness of focusing on changes in users.

6.2 External Validity

In the Netallica's case study, the users' action is defined as the article preference, which dynamically changes. However, other web services often provide more specific services (*e.g.*, online shopping, communication tool, *etc.*) and users' scenario do not change so much. ID3P contains an iterative improvement of personas based on verification of assumptions about personas' action through quantitative analysis. We assumed that this concept is independent of the service situation.

Service characteristics affect the improvement of GQM+S. In the second case study, every goal is related to the acquisition of users though business goals and is often related to other values (Benefit and Brand images, *etc.*). It means that the metrics of the upper-level goal are correlated with those of the sub-goals. This affects the evaluation of the upper-level goal because it includes achievement of sub-goals based on metrics.

In the future, to verify ID3P and avoid the above issues, we will evaluate the effectiveness of ID3P in other services based on GQM+S with multiple levels.

7. RELATED WORKS

ID3P proposes an iterative revision of personas to understand actual users' requirement and to help plan business strategies. Personas in ID3P are constructed via a data-driven approach, which is a practice of data-driven requirements engineering [28]. ID3P can be considered as a data-driven requirements engineering framework that includes verification of business strategies.

In Agile, the relationship of stakeholders is an important concept. Some previous methods have proposed integrating personas as models of end-users into Agile [5, 8, 9, 29]. Additionally, other methods have proposed integrating personas into a goal-oriented requirements model [6, 7]. However, we believe that these methods do not adequately address decision-making in a web service because they do not include the continuous development of personas to reflect changes in users explicitly. We assume that dynamic changes in users and measures to determine the effectiveness of a strategy should be considered to address decision-making in web service precisely. Consequently, ID3P extends previous methods in the terms of continuous revision of personas and integration of GQM+S.

With regards to an Agile practice, we believe that an iterative evaluation of personas will be helpful for reuse of persona documents. In the terms of avoiding documentations or other representations of personas for development, O'Leary *et al.* proposed reusable personas [30] and Vandenberghe proposed bot personas [31]. We assumed that ID3P would be applicable in different situations. O'Leary *et al.*'s personas are reused over projects while personas in ID3P are reused in the iterations in a project. Additionally, Vandenberghe's personas are used to acquire interactive feedback while personas in ID3P are used to evaluate a business strategy based on changes in users.

To solve issues about measuring the effectiveness of strategy, ID3P includes GQM+S as a Kaizen practice. A Kaizen practice is a key activity in continuous software engineering [32]. GQM+S itself can be considered as a Kaizen practice because it contains the PDCA cycle to improve business strategies based on metrics. Some previous methods reported that extensions of GQM+S and ID3P is an extension of Uchida *et al.*'s work in terms of data-driven personas development. Uchida *et al.* described that the relationship between usability metrics and business KPI is useful to improve the usability and planning strategies [33].

8. CONCLUSION AND FUTURE WORK

A persona is a fictional character developed to understand users' requirement. Many previous methods have proposed data-driven persona development approaches to reflect actual users. However, an actual web service has two issues. One is understanding users' preference and requirements because they dynamically change. The other is measuring strategies' impact on users due to the Hawthorne effect. (The latter makes it difficult to define effective target users to achieve a business goal.)

To solve these issues, we propose ID3P based on a quantitative evaluation and the revision of personas. Our contributions through ID3P include:

- Quantitative evaluations and revisions of personas to better understand users on a service
- Quantitative analysis of personas to derive business strategies
- Quantitative evaluation of strategies based on personas to improve business goals, strategies, and measurements

To verify ID3P, we implemented case study in Netallica, which is a web curation service of Yahoo! JAPAN. The result revealed:

- ID3P can provide personas that reflect the changes in actual current users. The evaluation results of persona revisions in two situations show that ID3P is suitable regardless of persona reconstruction. Therefore, we assume that ID3P can provide suitable personas.
- ID3P is helpful to improve the measurement of a business goal and to specify the target users. The improved GQM+S model in the second case study involves identifying target users and evaluating goals, strategies, and measurements. We believe that ID3P can improve that business strategies and goals through measurements of goal and analysis based on suitable personas.

In the future, we plan to conduct additional case studies using long-term data to verify the effectiveness of ID3P in other business domains. Moreover, we will propose an evaluation method for requirements using the relationships between business goal and users' personas.

REFERENCES

1. Y. Watanabe, H. Washizaki, K. Honda, Y. Noyori, Y. Fukazawa, A. Morizuki, H. Shibata, K. Ogawa, M. Ishigaki, S. Shiizaki, T. Yamaguchi, and T. Yagi, "ID3P: Iterative data-driven development of persona based on quantitative evaluation and revision," in *Proceedings of IEEE/ACM 10th International Workshop on Cooperative and Human Aspects of Software Engineering*, 2017, pp. 49-55.
2. E. Friess, "Personas and decision making in the design process: An ethnographic case study," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 1209-1218.
3. T. Matthews, T. Judge, and S. Whittaker, "How do designers and user experience professionals actually perceive and use personas?" in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 1219-1228.
4. K. Rönkkö, M. Hellman, B. Kilander, and Y. Dittrich, "Personas is not applicable: Local remedies interpreted in a wider context," in *Proceedings of the 8th Conference on Participatory Design: Artful Integration: Interweaving Media*, *Materials and Practices*, Vol. 1, 2004, p. 112.
5. J. Cleland-Huang, A. Czauderna, and E. Keenan, "A persona-based approach for exploring architecturally significant requirements in agile projects," in *Proceedings of the 19th International Conference on Requirements Engineering: Foundation for Software Quality*, 2013, pp. 18-33
6. M. Aoyama, "Persona-scenario-goal methodology for user-centered requirements engineering," in *Proceedings of the 15th IEEE International Requirements Engineering Conference*, 2007, pp. 185-194.
7. S. Faily and I. Fl'echais, "Eliciting and visualising trust expectations using persona trust characteristics and goal models," in *Proceedings of the 6th International Workshop on Social Software Engineering*, 2014, pp. 17-24.
8. M. Singh, "U-SCRUM: An agile methodology for promoting usability," in *Proceedings of Agile Conference*, 2008, pp. 555-560.
9. P. Wolkerstorfer *et al.*, "Probing an agile usability process," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008, p. 2151.
10. L. Karlsson *et al.*, "Requirements engineering challenges in market-driven software development – an interview study with practitioners," *Information and Software Technology*, Vol. 49, 2007, pp. 588-604.
11. L. Lehtola *et al.*, "Linking business and requirements engineering: Is solution planning a missing activity in software product companies?" *Requirements Engineering*, Vol. 14, 2009, pp. 113-128.
12. H. Washizaki, "Pitfalls and countermeasures in software quality measurements and evaluations," *Advances in Computers*, Vol. 107, 2017, pp. 1-22.
13. J. Pruitt and J. Grudin, "Personas: Practice and theory," in *Proceedings of Conference on Designing for User Experiences*, 2003, pp. 1-15.
14. J. J. McGinn and N. Kotamraju, "Data-driven persona development," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008, pp. 1521-1524.
15. D. M. Yang *et al.*, "Persona analysis with text topic modelling," in *Proceedings of International Conference on Automatic Control and Artificial Intelligence*, 2012, pp.

1559-1563.
16. T. Miaskiewicz *et al.*, "A latent semantic analysis methodology for the identification and creation of personas," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008, pp. 1501-1510.
17. R. Marshall *et al.*, "Design and evaluation: End users, user datasets and personas," *Applied Ergonomics*, Vol. 46, Part B, 2015, pp. 311-317.
18. M. Rahimi and J. Cleland-Huang, "Personas in the middle: Automated support for creating personas as focal points in feature gathering forums," in *Proceedings of the 29th ACM/IEEE International Conference on Automated Software Engineering*, 2014, pp. 479-484.
19. C. Brooks and J. Greer, "Explaining predictive models to learning specialists using personas," in *Proceedings of the 4th International Conference on Learning Analytics And Knowledge*, 2014, pp. 26-30.
20. X. Zhang *et al.*, "Data-driven personas: Constructing archetypal users with clickstreams and user telemetry," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2016, pp. 5350-5359.
21. V. Basili *et al.*, *Aligning Organizations Through Measurement: The GQM+Strategies Approach*, Springer, Berlin, 2014.
22. T. Kaneko *et al.*, "Application of GQM+Strategies in the Japanese space industry," in *Proceedings of Joint Conference of the 21st International Workshop on Software Measurement and the 6th International Conference on Software Process and Product Measurement*, 2011, pp. 221-226.
23. S. A. Sarcia', "Is GQM+Strategies really applicable as is to non-software development domains?" in *Proceedings of ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, 2010, pp. 45:1-45:4.
24. J. Mnch *et al.*, "Experiences and insights from applying GQM+Strategies in a systems product development organization," in *Proceedings of the 39th Euromicro Conference on Software Engineering and Advanced Applications*, 2013, pp. 70-77.
25. Y. Aoki *et al.*, "Identifying misalignment of goal and strategies across organizational units by interpretive structural modeling," in *Proceedings of the 49th Hawaii International Conference on System Sciences*, 2016, pp. 4576-4585.
26. T. Kobori *et al.*, "Exhaustive and efficient identification of rationales using GQM+Strategies with stakeholder relationship analysis," *IEICE Transactions on Information and Systems*, Vol. E99.D, 2016, pp. 2219-2228.
27. C. Shimura *et al.*, "Identifying potential problems and risks in GQM+strategies Models Using Metamodel and Design Principles," in *Proceedings of the 50th Annual Hawaii International Conference on System Sciences*, 2017, pp. 4857-4866.
28. W. Maalej *et al.*, "Toward data-driven requirements engineering," *IEEE Software*, Vol. 33, 2016, pp. 48-54.
29. L. Caballero *et al.*, "Persona as a tool to involving human in agile methods: Contributions from HCI and marketing," in *Proceedings of the 5th IFIP WG 13.2 International Conference on Human-Centered Software Engineering*, Vol. 8742, 2014, pp. 283-290.
30. C. O'Leary *et al.*, "Towards reusable personas for everyday design," in *Proceedings of ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 2016, pp. 2915-2922.

31. B. Vandenberghe, "Bot personas as off-the-shelf users," in *Proceedings of ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 2017, pp. 782-789.
32. B. Fitzgerald and K. J. Stol, "Continuous software engineering and beyond: Trends and challenges," in *Proceedings of the 1st ACM International Workshop on Rapid Continuous Software Engineering*, 2014, pp. 1-9.
33. C. Uchida *et al.*, "GO-MUC: A strategy design method considering requirements of user and business by goal-oriented measurement," in *Proceedings of IEEE/ACM Co-operative and Human Aspects of Software Engineering*, 2016, pp. 93-96.
34. Y. Watanabe *et al.*, "Retrospective based on data-driven persona significance in B-to-B software development," in *Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results*, 2018, pp. 89-92.

Yasuhiro Watanabe received the B.E. degree in Computer Science and Engineering from Waseda University, Tokyo, Japan in 2017. His research interests include software engineering especially requirement engineering with employing data analysis.

Hironori Washizaki is Director and Professor at the Global Software Engineering Laboratory, Waseda University, Japan. He also works at the National Institute of Informatics as a Visiting Professor, at System Information Co., Ltd. as an Outside Director, and at eXmotion Co., Ltd. as an Outside Director. He is the Editor-in-Chief of International Journal of Agile and Extreme Software Development. He has served as the editor of journals including International Journal Software Engineering and Knowledge Engineering, IEICE Transactions on Information Systems, Computer Software, Heliyon, and Applied Computing and Informatics. He has contributed to societies such as IEEE Computer Society Membership at Large for the Professional and Educational Activities Board, IEEE Computer Society Japan Chapter Chair, SEMAT Japan Chapter Chair, IPSJ SamurAI Coding Director, ISO/IEC/JTC1 SC7/WG20 Convenor, ICST 2017 PC Co-Chair, CSEE&T 2017 PC Co-Chair and APSEC 2018 PC Co-Chair.

Kiyoshi Honda received the B.E, M.S. and D.E. degrees from Waseda University, Tokyo, Japan in 2009, 2011 and 2017, respectively. He is an Assistant Professor at the Global Education Center, Waseda University, Japan. His research interests include empirical software engineering and software reliability.

Yuki Noyori received the B.E. degree in Computer Science and Engineering from Waseda University, Tokyo, Japan in 2018. His research interests include software engineering with machine-learning and natural language processing.

Yoshiaki Fukazawa received the B.E, M.E. and D.E. degrees in Electrical Engineering from Waseda University, Tokyo, Japan in 1976, 1978 and 1986, respectively. He is now a Professor of Department of Information and Computer Science, Waseda University. Also he is Director, Institute of Open Source Software, Waseda University. His research interests include software engineering especially reuse of object-oriented software and agent-based software.

Aoi Morizuki was a Senior Manager of Local Service Department, Yahoo Japan Corporation. Her work was distributing Yahoo Map and Yahoo Car Navigation.

Hiroyuki Shibata is a Planner of Membership Business Division and Business Strategy Division, Yahoo Japan Corporation. His work is planning a business for premium membership benefits.

Kentaro Ogawa is a Manager of Technology and Solution Division, Yahoo Japan Corporation. His work is supporting service development.

Mikako Ishigaki is an Engineer of Technology and Solution Division, Yahoo Japan Corporation. Her work is supporting service development.

Sachiyo Shiizaki is a Planner of Media Business Department, Yahoo Japan Corporation. Her work focuses on planning to increase numbers of Netallica user.

Teppei Yamaguchi is a Manager of Technology and Solution Division, Yahoo Japan Corporation. His research interests include agile development, quality assurance, and model based development. His current work focuses on spreading the company testing and test automation.

Tomoaki Yagi is a Manager of Technology and Solution Division, Yahoo Japan Corporation. His work is improving of service quality in the company.