The Quality Guardian: Improving Activity Label Quality in Event Logs Through Gamification (Extended Abstract)

Sareh Sadeghianasl

Queensland University of Technology, Brisbane, Australia

Abstract

Data cleaning, the most tedious task of data analysis, can turn into a fun experience when performed through a game. This thesis shows that the use of gamification and crowdsourcing techniques can mitigate the problem of poor quality of process data. The Quality Guardian, a family of gamified systems, is proposed, which exploits the motivational drives of domain experts to engage with the detection and repair of imperfect activity labels in process data. Evaluation of the developed games using real-life data sets and domain experts shows quality improvement as well as a positive user experience.

Keywords

Process mining, Data quality, Activity labels, Gamification, Crowdsourcing

1. Introduction

Data quality is critical for efficient and low-risk data-driven decision making in organizations. Process mining concerns the analysis of event logs to provide a better understanding of the *real* processes executed within an organization to support decision making. Low-quality event logs negatively affect the reliability of process mining results—*garbage in, garbage out.* Activity labels, the recorded names of the tasks performed in a process, are key elements of event logs. However, their quality can be compromised. Multiple activity labels with different syntax may refer to identical tasks. Detecting and repairing imperfect activity labels requires a *deep insight* into the domain involved to understand the meaning of labels [1, 2, 3]. Automatic approaches to detect and repair imperfect activity labels often underestimate the complexity of this task and suffer from low effectiveness in real-life data sets [1, 4]. Domain experts are eminently suited to fix imperfect labels, but it is hard to engage them, as repair can be time-consuming and tedious [5].

Gamification incorporates game elements in system design to improve user engagement with non-game tasks [6]. It has the potential to offer a promising solution to the challenge of domain expert engagement in the task of detecting and repairing imperfect activity labels. The main research question that this study aims to answer is as follows:

BPM 2022 Best Dissertation Award, Doctoral Consortium, and Demonstration & Resources Track, 13–15 September, Münster, Germany

s.sadeghianasl@qut.edu.au (S. Sadeghianasl)

0000-0002-0338-958X (S. Sadeghianasl)

Research Question (RQ): *To what extent is gamification a promising approach for detecting and repairing activity labels with different syntax but the same semantics?*

To be more precise, this research question can be broken down into two sub-questions. First, assuming that domain experts are engaged with the gamified system, it is essential to determine if the approach works technically, i.e., if it improves the quality of event logs. Hence, the first sub-question is:

RQ1: *To what extent is gamification a promising approach for improving the quality of event logs by detecting and repairing activity labels with different syntax but the same semantics?*

Knowing that the approach technically works, its practical success now hinges on whether enough of the right type of people will become involved in the gamified system. Hence, the second research sub-question is:

RQ2: *To what extent is gamification a promising approach for supporting user engagement in the task of detecting and repairing activity labels with different syntax but the same semantics?*

2. Approach

This thesis addresses the research questions through developing gamification solutions to detect and repair imperfect activity labels and testing their effectiveness. Due to the large number of distinct activity labels in real-life event logs (e.g., 624 unique labels in the Hospital log used for the BPI Challenge 2011 [7], and 304 labels in the second Municipality log used for the BPI Challenge 2015 [8]), it might not be feasible for domain experts to check all activity labels. The first approach (Chapter 3) presents a new technique to automatically identify imperfect label candidates in an event log to be investigated further by experts. This approach looks at activity context, i.e. control flow, resource, data, and time. If two activity labels are close in terms of their context, they are identified as candidate imperfect labels.

In order to support domain expert engagement with activity label quality improvement, *The Quality Guardian*, a family of gamified systems, is proposed. This family consists of three gamified systems: (i) *The Quality Guardian* (Chapter 4) (ii) *The Quality Guardian Redux* (Chapter 5), and (iii) *The Quality Guardian Rosebud* (Chapter 6). These games aim to harvest knowledge of multiple domain experts (a form of crowdsourcing [9]) to improve the quality of activity labels in a collaborative and interactive manner.

The Quality Guardian gamified system is an initial attempt to gamify activity label quality improvement using common game elements such as points, badges, and progress bars. The Quality Guardian Redux game is designed with the main focus on motivating domain experts to engage with activity label repair. More specifically, the motivations of altruism, self-development, and reputation building (identified from the literature as the top knowledge sharing motivations [10, 11, 12]) were incorporated in the design of this game. These drives are further framed

using the Octalysis framework for gamification design [13] and Self-Determination Theory (SDT) [14].

The first two games improve the quality of activity labels in event logs; however, they are dependent on the participation of domain experts. The domain knowledge required for activity label repair can be provided from other sources, such as a domain ontology. However, such an ontology may not always be available. The Quality Guardian Rosebud gamified system aims to create an ontology of activity labels from an event log, which can be used for repair. Four types of semantic relations between activity labels are covered in the ontology: *synonymy*, *hypernymy* (i.e., the super-class sub-class relation), *holonymy* (i.e., the whole-part relation), and *antonymy*.

3. Evaluation and Results

The contextual approach proposed in Chapter 3 was evaluated using real-life logs from a hospital and an insurance company. The results show efficient detection of frequent imperfect labels, which pose more serious problems than infrequent ones. The results also suggest that the control flow and resource dimensions are more informative for detecting frequent imperfect activity labels, while for detecting infrequent ones, the resource and data dimensions are more helpful. Furthermore, the temporal dimension seemed to be the least informative dimension for detecting frequent and infrequent imperfect activity labels.

The Quality Guardian gamified system was evaluated by 21 participants who repaired the real-life BPIC 2019 event log [15] with injected imperfect labels. The results show that 21 out of 25 imperfect labels were detected and repaired (i.e., 84% success rate; RQ1). A custom-developed survey based on the GameFlow framework [16] was used to measure participants' engagement. The survey responses suggest that most participants found the system very easy to use and engaging overall (RQ2).

A real-life event log from an Australian insurance company was used for the evaluation of the Quality Guardian Redux gamified system. In order to measure the engagement of domain experts, interviews were conducted with 14 experts from the same company after they played the game. The results show that 48 out of 54 imperfect labels in the game are detected and repaired (i.e., 88.89% success rate; RQ1). An extensive qualitative analysis of the interviews showed that almost all the experts were inspired by some aspects of the three aforementioned motivations. Overall, the game was perceived as a useful way to improve user experience in the task of detecting and repairing imperfect activity labels (RQ2).

The Quality Guardian Rosebud gamified system was evaluated with the BPIC15_2 event log [8] containing real imperfect labels and 35 participants from the public. Participants' engagement was measured using a custom-developed survey based on the Octalysis [13] and GameFlow [16] frameworks and SDT [14]. The results confirmed the high quality of the created process activity ontology and a positive user experience (RQ2). After creation, a process activity ontology can be used for activity label quality improvement (e.g., to repair synonymous labels) and choosing activity labels at different levels of granularity in event logs (RQ1).

4. Contributions

This thesis contributes to the fields of process mining (contributions 1, 2, and 3), human-computer interaction (contribution 4), and semantic web (contribution 5), as follows:

1. An automatic approach to detect candidates of activity labels with the same meaning in event logs (Chapter 3). This approach conceptualizes activity context in event logs and defines context distance measures between activities.
2. Approaches to data cleaning through playing games to turn the most tedious task of data science into an engaging experience (Chapters 4 and 5). The Quality Guardian and The Quality Guardian Redux gamified systems have achieved promising results both in terms of quality improvement of activity labels and user engagement.
3. Creating activity ontologies from event logs to formalize domain knowledge (Chapter 6). The Quality Guardian Rosebud gamified system generates an ontology of activity labels and their semantic relations, which can be used for detecting and repairing imperfect activity labels.
4. Identifying the most promising motivational drives of domain experts to repair activity labels in event logs and their associated game elements (Chapters 4, 5, and 6). The surveys and the interviews examine the effect of a range of motivations and game techniques in the engagement of participants with activity label repair.
5. The first-ever approach to ontology learning through playing a game (Chapter 6).

5. Conclusion and Future Work

This thesis proposed the Quality Guardian, a family of gamified systems to improve the quality of activity labels in process event logs. There are several pathways for future research. This study focused on activity labels, which are one of the three necessary elements of event logs. However, the quality of the other two elements (i.e., timestamp and case) can also be compromised. The detection and repair approaches for other types of quality issues, possibly through gamification, can be explored in the future. Preventing the occurrence of data quality issues, incorporating other types of motivational drives in the game design, and developing methodologies for systematically designing games for data cleaning are other possible topics of future research. This thesis paves the way to turn data cleaning from the most tedious task of data science projects into an entertaining experience for domain experts in the future.

References

[1] C. Klinkmüller, H. Leopold, I. Weber, J. Mendling, A. Ludwig, Listen to me: Improving process model matching through user feedback, in: Business Process Management (BPM) Conference, volume 8659 of *LNCS*, Springer, 2014, pp. 84–100.

[2] C. Rodríguez, C. Klinkmüller, I. Weber, F. Daniel, F. Casati, Activity matching with human intelligence, in: Business Process Management (BPM) Forum, volume 260 of *LNBIP*, Springer, 2016, pp. 124–140.

[3] E. Scibona, Cost-effective and Scalable Activity Matching using Crowdsourcing, Master's thesis, Politecnico di Milano, Milan, Italy, 2018.
[4] C. Klinkmüller, I. Weber, Every apprentice needs a master: Feedback-based effectiveness improvements for process model matching, Information Systems 95 (2021) 101612.
[5] G. Press, Cleaning big data: Most time-consuming, least enjoyable data science task, survey says, https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=491be206f637, 2016. Accessed 22 Apr 2022.
[6] S. Deterding, M. Sicart, L. E. Nacke, K. O'Hara, D. Dixon, Gamification: Using Game-Design Elements in Non-Gaming Contexts, in: International Conference on Human Factors in Computing Systems (CHI), ACM, 2011, pp. 2425–2428.
[7] van Dongen B.F., Real-life event logs - Hospital log. 4TU.ResearchData. Dataset., https://doi.org/10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffcf54, 2011. Accessed 22 Apr 2022.
[8] van Dongen B.F., BPI Challenge 2015 Municipality 2. 4TU.ResearchData. Dataset., https://doi.org/10.4121/uuid:63a8435a-077d-4ece-97cd-2c76d394d99c, 2015. Accessed 22 Apr 2022.
[9] J. Howe, The rise of crowdsourcing, Wired magazine 14 (2006) 1–4.
[10] S. Oreg, O. Nov, Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values, Computers in Human Behavior 24 (2008) 2055 – 2073.
[11] P. Prasarnphanich, C. Wagner, The role of wiki technology and altruism in collaborative knowledge creation, Journal of Computer Information Systems 49 (2009) 33–41.
[12] H. L. Yang, C. Y. Lai, Motivations of wikipedia content contributors, Computers in Human Behavior 26 (2010) 1377 – 1383.
[13] Y.-K. Chou, Actionable gamification: Beyond points, badges, and leaderboards, Packt Publishing Ltd, 2019.
[14] E. L. Deci, R. M. Ryan, Self-Determination Theory, in: Handbook of theories of social psychology, volume 1, Sage Publications Ltd, 2011, pp. 416–436.
[15] van Dongen B.F., BPI Challenge 2019. 4TU.ResearchData. Dataset., https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1, 2019. Accessed 22 Apr 2022.
[16] P. Sweetser, P. Wyeth, Gameflow: a model for evaluating player enjoyment in games, Computers in Entertainment 3 (2005) 3.