Provenance for Online Decision Making

Amir Sezavar Keshavarz[1], Trung Dong Huynh[1], and Luc Moreau[1]

Electronics and Computer Science (ECS),
University of Southampton, Southampton, United Kingdom
{ask2g10,tdh,L.Moreau}@ecs.soton.ac.uk

1 Introduction

In our IPAW'14 paper [1], we identified a novel application of provenance: Online Provenance-based Decision Making. There, we put forth a generic domain-agnostic methodology whereby a provenance-aware application can use provenance to make online decisions. We evaluated the methodology and demonstrated how a provenance-aware application can use provenance to make quality-based decisions in a timely manner and change its behaviour.

In this paper, we use a scenario (Section 2) to present a tutorial on the methodology, how it can be used in provenance-aware applications, and how provenance-based decisions can change the behaviour of the application (Section 3).

2 Scenario: A Crowd-sourcing Application

CollabMap is a crowd-sourcing application that recruits people to augment existing maps by identifying buildings outline and drawing their evacuation routes to nearby roads. Participants are required to verify tasks by others by providing positive or negative votes on buildings and evacuation routes, helping CollabMap to determine their validity. Meantime, CollabMap records provenance for all the data it generates such as buildings, evacuation routes, and etc. The quality of data generated by a crowd with different backgrounds and expertise is inevitably varied. Therefore, to improve the quality assessment done by CollabMap in an online environment, we set the following requirements for our system.

1. To compute a validity label ("valid", "invalid", or "uncertain") for each data entity (buildings and evacuation routes) by using provenance of data. "valid" data entities are to be included in the final result, while "invalid" data entities are to be discarded.
2. To compute a reliability measure for each user by using provenance of data, so that we can use these measure to increase confidence on validity label by analysing users' reliability.
3. To compute a finish measure for each data entity from its provenance to decide when a task is deemed complete.

3 Online Provenance-based Decision Making Methodology

The methodology by which a provenance-aware application can make provenance-based decisions is domain-agnostic. However, naturally, decisions that are to be made are application-specific. Furthermore, a provenance-aware application may want to use provenance generated in other applications to make decisions. For example, based on scenario, a provenance-aware application is using provenance generated in CollabMap to make quality-based decisions.

The methodology advocates the following three phases to assist provenance-aware applications to make decisions in a timely manner:

1. Use annotations to represent additional data
2. Compute new annotations from existing annotations based on application-specific instructions
3. Comply with the methodology contract to make incremental decisions in a timely manner

3.1 Phase 1: Annotation Assertion

Phase 1 recommends the use of annotation as a generic tool to represent application specific data. For example, the quality-based application annotated User426.12 with *bi* annotation which is an application-specific data representing the reliability of User426 (Figure 1). *bi* is abbreviation for *Building Identification* and represents how reliable a user is in identifying good buildings.

3.2 Phase 2: Annotation Computation

Phase 2 argues annotations could be propagated forward and backward in a provenance graph, resulting in yet more annotations. For example, the quality-based application annotated Building12436.5 with *tcubv*. *tcubv* is abbreviation for *Total Cumulative User Building Verification* which is an application-specific data representing the total reliability of verifiers for that building. This annotation was computed from *bi* annotation of all verifiers. To do so, *bi* annotation was propagated forward through the provenance graph (traversing 4 relations).

Furthermore, the methodology provides a foundation in which provenance-aware applications instruct how annotations are computed. For instance, the quality-based application implements a quality model to define annotations such as *bi* or *tcubv* and instructs how they are computed from other annotations such as *negativeVotes* or *positiveVotes*.

3.3 Phase 3: Methodology Contract

Phase 3 requires provenance-aware applications to comply with a contract that models the structure of provenance graph to assure incremental decision making

in a timely manner. For CollabMap to be able to make incremental decisions, it structures the provenance based on the contract. As new annotations are computed for both User429 and Building12436, new versions of these nodes are created. Figure 1 represents the 12^{th} and 13^{th} versions of User429. Annotations of the 13^{th} version was computed from the 12^{th} version. The same is true for Building12436.

4 Online Decision Making

Online provenance-based decision making methodology allows provenance-aware applications to use provenance to make online decisions while they are executing. By following this methodology, a provenance-aware system can cope with situations as they are evolving and uses provenance of data and annotations to change its behaviour as necessary.

Our preliminary analysis on CollabMap in our IPAW paper [1], shows the methodology can help CollabMap to make better decisions and faster. For example, application-specific annotations for buildings that were computed by using provenance of its data, allow CollabMap to decide to terminate a task earlier if the outline of building did not met a set of requirements (for example having three edges or self-intersecting lines).

The quality-based application also annotates each user with a reliability measure, expressing how reliable they have been performing. CollabMap can use these annotations to allocate more tasks to reliable ones or warns unreliable ones. Furthermore, the quality-based application annotates each building if more verification vote is necessary and CollabMap can decide if and when the crowds contributions to a task are deemed to be complete.

References

1. Amir Sezavar Keshavarz, Trung Dong Huynh, and Luc Moreau. "Provenance for Online Decision Making". *5th Internation Provenance and Annotation Workshop*, 2014.

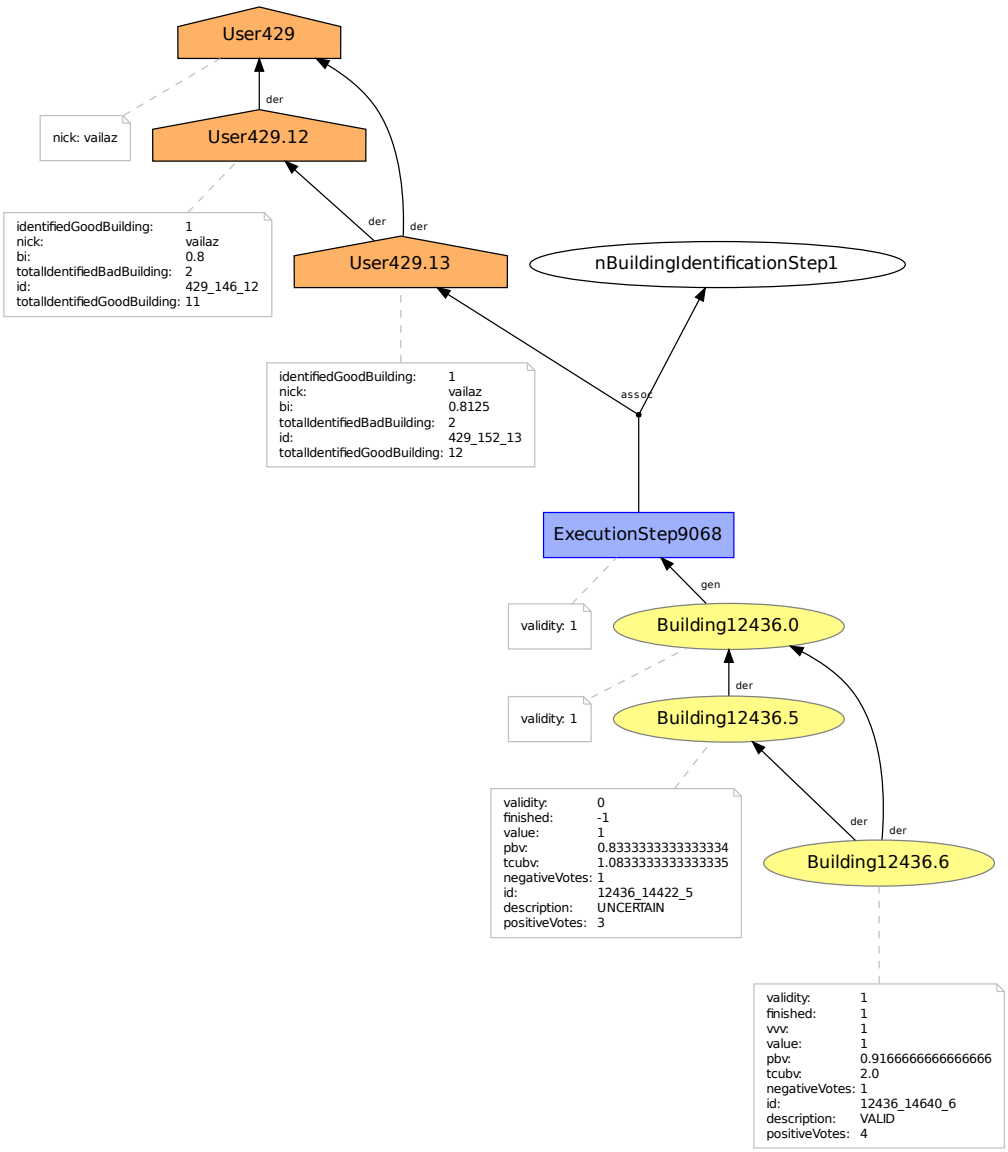

Fig. 1: A sample provenance graph generated in CollabMap and annotated in a provenance quality-based application