Enabling Semantic Understanding of Situations from Contextual Data In A Privacy-Sensitive Manner

Fuming Shih
Computer Science and Artificial Intelligent Lab
Massachusetts Institute of Technology
32 Vassar Street, Cambridge, MA 02139

Vidya Narayanan and **Lukas Kuhn**
Qualcomm, Inc.
5775 Morehouse Dr.
San Diego, CA 92121

Abstract

Mobile applications can be greatly enhanced if they have information about the situation of the user. Situations may be inferred by analyzing several types of contextual information drawn from device sensors, such as location, motion, ambiance and proximity. To capture a richer understanding of users' situations, we introduce an ontology describing the relations between background knowledge about the user and contexts inferred from sensor data. With the right combination of machine learning and semantic modeling, it is possible to create high-level interpretations of user behaviors and situations. However, the potential of understanding and interpreting behavior with such detailed granularity poses significant threats to personal privacy. We propose a framework to mitigate privacy risks by filtering sensitive data in a context-aware way, and maintain provenance of inferred situations as well as relations between existing contexts when sharing information with other parties.

Introduction

With sensing and communication capability, modern mobile devices are becoming the most comprehensive platform for context-aware applications. Such platforms merge user's content from applications on the Web and data created or sensed locally on mobile devices. Smart sensors on the mobile device can easily collect information such as location, movement and environment of a user. Situation awareness can potentially lead to a wealth of applications, including the ability to provide user-specific experiences and improvements. Although not a solved problem, the use of machine learning techniques to derive contextual information about the user, such as motion states or significant places, is very prevalent. There have also been some attempts to approach the problem of situation awareness from the semantic aspects. However, bringing the two fields together to provide a semantic understanding of a situation is promising and powerful. We borrow the use of the terms context and situation as per the definition given by Dey (2001) - *"Context is any information that can be used to characterize the situation of an entity"*.

There are two main challenges in realizing situation awareness: 1) fusing contextual information from disparate sources in a coherent and structured manner for reasoning - the contexts from different sources have varying semantics and relationships; 2) representing the privacy constrains in a situation or context specific manner that allows integration with the reasoning process.

Towards a semantic understanding of situations, we can model the relations between various types of contextual information and content in an upper-level ontology shown in Figure 1. Represented in this ontology are the actions, places and other information produced by a machine learning based analysis on raw sensor or other data. Labels associated with data in machine learning algorithms can be seen as the first level of semantics. For instance, motion may be labeled as walking, running, etc., which are contexts indicated by the Action ontology in Figure 1. The ontologies then capture the relationships that exist across these semantic labels as well as to knowledge bases built from both modeled common sense and learned user knowledge. In addition, we provide a discussion here about how privacy fits into this model and the overall situation awareness subsystem.

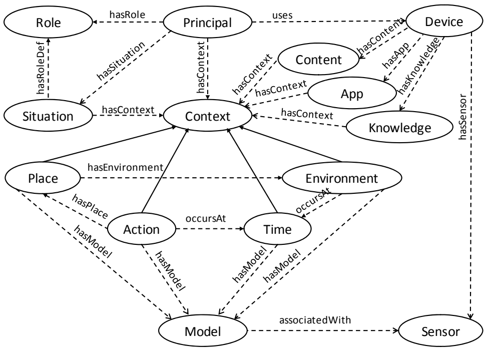

Figure 1: Top level ontology for situation awareness (solid lines show subClassOf)

Motivating Use-Case

We motivate the need for context representation and reasoning for situation-aware applications using a use case from a scheduling application. As we will show, this use case incorporates aspects of context-awareness, reasoning over situations, information sharing, and privacy.

Consider the case where Alice is currently at a doctor's appointment,following which, she has an important meeting at work. Alice is supposed to be presenting at the meeting. Consider a situation-aware application (e.g. a scheduling application) that is able to infer a situation of 'running late for a meeting' from derived context artifacts. Thus, such a system can remind Alice of the meeting in a situation aware manner and notify others that Alice is running late. For example, say the typical travel time from the location of her doctor's office to the location of her meeting is 20 minutes. The system can monitor Alice to track if she is leaving on time. Based on this information, the system can notify other participants of the meeting that Alice is running late.

The system is also aware of the reason for her being late and can provide this to the meeting invitees. However, making the reason known needs to happen in accordance with Alice's privacy preferences. For instance, the reason for being late may be shared only with certain invitees of the meeting and not others. Further, the optimal information sharing strategy may not only come from Alice's situation and settings, but also from the situation of the other participants. For example, if one of the participants is on vacation or if a participant has a conflict that will delay his or her participation in this meeting by 30 minutes, a notification would not do any good.

Upper Level Ontology for Situation Awareness

Towards a collaborative situation-aware application, a first step is to extract context artifacts that lead to a situational understanding. Such artifacts are typically derived from various information sources such as sensors, audio, location, time, application content, user-device interactions, etc. by using a broad range of techniques such as signal processing, machine learning, natural language processing, etc. The contextual information learnt after application of such preprocessing techniques is referred to in this paper as context artifacts.

Given the capability of extracting contextual information, a main challenge is to capture and integrate contextual information in a coherent, structured and semantically meaningful knowledge base (KB) that aids reasoning based on the various context artifacts. This is challenging, as contextual information may be derived from different sources with varying representations and semantic interpretations. Consider location as an example. There are many different ways in which location can be interpreted. A location may refer to an exact position on the globe or a place (e.g. home, office, store). It is important to enable a common understanding of concepts and their relations to enable semantically meaningful information sharing. For this reason, we propose the upper-level ontology for situation awareness illustrated in Figure 1 which is partly motivated from (Chen et al. 2004). The core concepts in this ontology are described below.

- A **principal** is the primary actor who performs actions. A principal can be associated with varying contexts over time and take on multiple roles in the same or different situations.
- A **role** characterizes a principal with respect to an situation. Example roles include guest, waiter, and owner in a restaurant situation or presenter, organizer and invitee in a meeting situation etc.
- A **situation** is a high level concept which is typically defined over a set of context artifacts by tying together various pieces of contextual information. Examples include dining at a restaurant, delivering a lecture, grocery shopping, in a meeting etc.
- **Context** is any information that can characterize a situation. There are many sub-classes in context, only some of which we are able to address explicitly: place, action, environment, and time.
- A **Place** is a location with a semantic meaning. In other words it is a location which is relevant to a situation or an action. Examples of places include home, work, office, meeting room, etc.
- An **action** describes an atomic lower level task performed by a principal with no direct association to a role. Examples include silencing a cell phone, rejecting a call, running, driving, eating, talking, typing, etc. Actions have a temporal aspect and can be seen as context information.
- An **environment** characterizes the ambiance in which a principal is present. Examples include information such as dark, loud, cold, crowded, proximity to other principals or devices, etc.
- **Time** captures temporal information of a context and is itself context.
- A **device** is a system that provides the contextual information that contribute to a situation. In some cases, the device is a system which follows the person (e.g., mobile phone). In some other cases, devices are systems that are present in the environment of the principal.
- The concept **model** captures the characteristic of another concept. Examples include the GPS-model of a place or the motion model of an action.
- **Sensors** on a device include accelerometer, light, proximity, microphone, GPS, WiFi, etc.
- **Content** is any information available from a device, either in the form of structured or unstructured data.
- Applications **(Apps)** are resident on a device; these may produce structured, unstructured or semi-structured data.
- **Knowledge** may be learned or commonsense knowledge obtained from various sources.

The ontology explicitly outlines places, actions and environments as contextual information derived from physical sensors. Other types of contextual information are derived from applications, content and knowledge available in the system. Domain specific ontologies are expected to tie into this upper level ontology for specific types of context information sources. For e.g., in the motivating use case described, a calendar ontology (e.g., iCal RDF vocabulary) may be a type of application context that is used to model the calendar information and the context derived from it. The goal of the ontology is to provide a common formalism to enable reasoning about situations.

Even though context artifacts provide a lot of information, they are typically isolated and do not directly lead to situation-awareness. Fusing context artifacts for inference at a statistical or probabilistic level produces richer context artifacts, but still does not lead to a semantic understanding of the situation. In our use case, examples of context artifacts are: indoor and outdoor location based on GPS or Wifi (Place), motion states such as walking or driving based on inertial sensors (Action), the ambiance of a user based on inertial or audio sensors (Environment), the proximity to other people (Principals) based on audio signatures or network interfaces (Environment), and many more. The gathering of such contextual information is by itself a hard problem which we do not address in the scope of this discussion. Instead, we emphasize that such artifacts are only a stepping stone towards situation-awareness.

Reasoning over Situations

In this section, we focus on reasoning to bridge the gap between isolated context artifacts and situation awareness. Situations are typically defined over a set of context artifacts. A subset of context artifacts may be applicable to multiple situations and this calls for consideration of sufficient artifacts in the reasoning process to infer a situation with high certainty. For example, the situation of being in a meeting is commonly defined as a gathering of two or more people (fact) that has been convened for the purpose (fact) of achieving a common goal (fact) through verbal interaction (fact), such as sharing information or reaching agreement (of Labor Statistics' 2009). However, meetings may occur face to face or virtually, as mediated by communications technology (e.g., a videoconference). Thus, a meeting may be distinguished from other gatherings, such as a chance encounter (not convened), a sports game or a concert (verbal interaction is incidental), a party or the company of friends (no common goal is to be achieved) and a demonstration (whose common goal is achieved mainly through the number of demonstrator present, not verbal interaction).

Given a definition of a situation over contextual artifacts, the challenge remains to infer situations from those contextual artifacts (which are often derived in isolation). We start by formalizing the contextual artifacts and their relations using our ontology. Given this knowledge base, we illustrate rule-based reasoning to infer situations.

Table 1 illustrates an approach to express rules as first order logic expressions and associating with each rule a probabilistic weight that indicates the likelihood that the inference is true given the constraints or assumptions are satisfied. Inference can be performed over all possible worlds to determine the world with the highest likelihood. The variables and predicates over which the constraints and rules are defined are derived from the ontology. First-order logic is used only for illustration and we acknowledge that there are other ways of representing the rules.

Recall the use case from before. Alice is running late to a meeting and invitees should be notified if and only if they are likely to be in the meeting room before she arrives. For this, the system requires an understanding of who should be notified. If Bob is an invitee, the system can infer (without considering any other information) that Bob is at the meeting room based on Rule 1 (in Table 1). However, the likelihood that supports that inferred result is low. Given the location of Bob's mobile phone, Bob can be placed at the same location based on Rule 5. Depending on the accuracy of the indoor location identification, the inference may not have a higher likelihood based on the place information. Acquiring other information such as whether Bob answers his office phone may not be a viable option and therefore, the system continues with reasoning over incomplete knowledge and infers with low confidence that Bob is at the meeting.

Challenges in Reasoning

The main challenges originate from uncertainty in the contextual facts, uncertainty in the inference rules or from missing data. Uncertainty in the contextual artifacts comes from the fact that the artifacts themselves are derived from probabilistic information. Data may also be missing due to unavailability of data for a particular duration or due to known gaps in data for power saving purposes or due to privacy policies or preferences. Further, multiple contextual artifacts may lead to the same situation as well or a subset of the contextual artifacts may be shared across multiple situations. Combined, these aspects present some key challenges in the reasoning process.

Privacy Sensitive Inference and Information Sharing

In any collaborative application, there is an element of information sharing. Designing a strategy for information sharing is a balance between efficient communication and privacy restrictions. Traditionally, this tradeoff is managed with static privacy settings which are not situation sensitive. For instance, in some situations, a user might be willing to disclose his or her location, but may not wish to disclose it all the time. As a result, users end up with static privacy settings which are usually more restrictive and richer functionality cannot be offered as a result. The ability to infer situations allows for a better way of realizing more dynamic privacy settings. However, in a system that is situation-aware, additional challenges arise that require privacy policies to be applied even during the inference process.

Privacy in Situation Awareness

To properly address the privacy aspect in situation awareness, we present a shift in focus from previous researches (Corradi, Montanari, and Tibaldi 2004; Zhang and Parashar 2004; Toninelli et al. 2006), about the concept of privacy. The focus switches from "information access" of sensitive

	English	Inference Knowledge Base	Weight
1	People who are invited to a meeting will go to the meeting.	$Inv(x,m) \Rightarrow Part(x,m)$	0.65
2	People who are located at the meeting location are in the meeting.	$At(m,l) \wedge At(x,l) \Rightarrow Part(x,m)$	0.82
3	People who answer their office phones are in their office.	$Ans(x,p) \wedge At(p,l) \Rightarrow At(x,l)$	0.98
4	People who do not answer their office phone are not in their office.	$\neg Ans(x,p) \wedge At(p,l) \Rightarrow \neg At(x,l)$	0.81
5	People are at the same location as their mobile phone.	$At(m,l) \Rightarrow At(x,l)$	0.94

Table 1: Rule set to infer the location of an invitee of a meeting.

information to the consequences of "information use" (Kagal and Abelson 2010). Recent privacy stories on the Web mainly came from the inappropriate use of personal information with unintended consequences (Shih and Paradesi 2010); the situation in context-aware platform is expected to be even worse. One observation is that the users exhibit different degrees of sensitivity to information about them and this concern is easily affected by the context within where the information is "collected" and where the information is "evaluated" (Sadeh et al. 2009).

Most of the decisions of giving away personal contexts may be driven by the immediate benefit received from situation-aware services. However, the purpose of information use should be clearly defined for both reasoning about situations and sharing inferred context. The reasoning process should also be transparent and provide explanations of the inferred situation as mentioned in (Weitzner et al. 2006). The justifications provided by the program can help the user to assess the quality and appropriateness of the inferences.

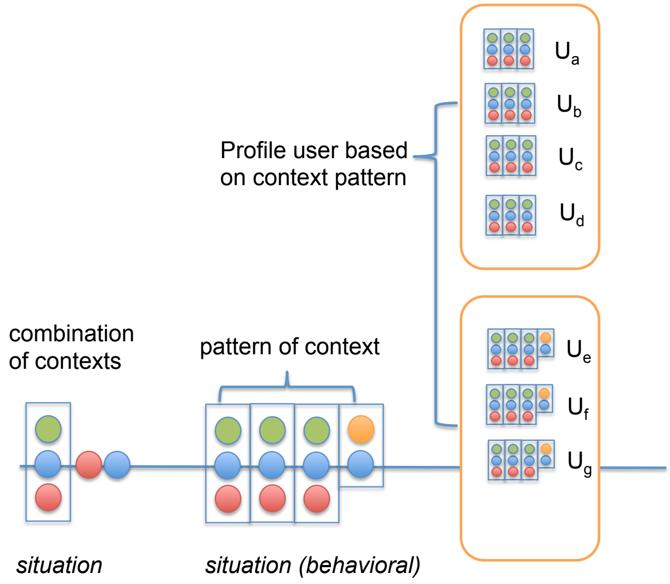

Figure 2: Privacy concerns in different levels of use about context information

We argue that privacy analysis is not limited to a single situation, but also on how each situation plays into a pattern of contexts that can be inferred by the observer. A situation, as shown in 2, may be a snapshot of multiple context artifacts at a particular time. A situation may also be a pattern of context attributes over time - typically, this leads to revealing behaviors of users and groups. From such behavioral patterns, it is possible to construct user profile information, categorizing groups of users. Research work such as the "Reality Mining" effort (Eagle and Pentland 2006) aggregates these behavior-profiles and identifies an individual as belonging to a specific social group. Such an analysis can often lead to loss in privacy, even if individual context attributes or a single situation didn't impact privacy.

Architecture for Privacy in Situation Awareness Platform

Here, we like to address privacy mechanisms in the "contexts" where personal information is collected, inferred, shared, and evaluated. Figure 3 shows the architecture of a privacy-sensitive situation awareness platform. In the development of a situation-aware system that uses data from a variety of sources, three broad privacy issues arise:

a) Privacy of data collected for training to develop context aware systems. Users must be offered privacy so that users who contributed training data cannot be identified. At a minimum, users must be offered the option of limiting when and where the data is collected.

b) Privacy of data used to infer contextual information. User preferences should guide how and what data is used for inference purposes.

c) Privacy in sharing the inferred context. The user should be able to manage permissions on what data and what inference is permitted share between various requesting entities (users or applications).

We illustrate b) and c) in the context of our running example, next.

User's Policy as Privacy Constraints

A user should be able to specify how his or her data can be collected, used and shared by the program in situation

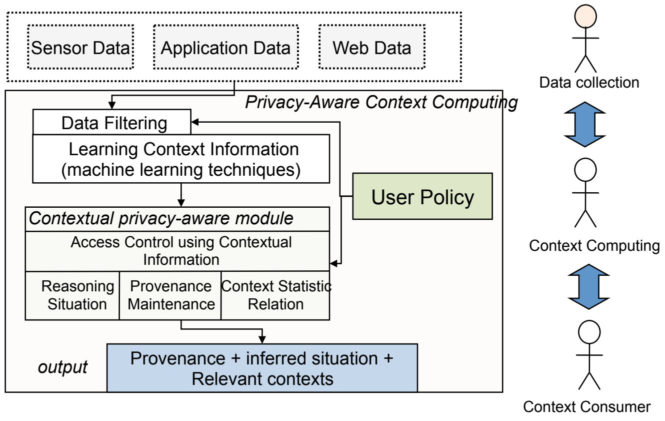
Sensor Data
Application Data
Web Data
Privacy-Aware Context Computing
Data Filtering
Learning Context Information
(machine learning techniques)
User Policy
Contextual privacy-aware module
Access Control using Contextual Information
Reasoning Situation
Provenance Maintenance
Context Statistic Relation
output
Provenance + inferred situation +
Relevant contexts
Data collection
Context Computing
Context Consumer

Figure 3: Architecture of privacy-aware context computing

awareness platform. In addition, a user can use “context artifacts” or ”situations” to specify the conditions under which the collection of data or reasoning about data is valid. For example, a policy from the user may state “do not collect my location when I am not at the work place” or “any contextual information collected after working hours cannot be used for making inferences about my relationship with my colleagues”. In our scenario, Alice can set the rule to prohibit the program from collecting her location while she is away for her doctor’s appointment. With that restriction, the only inference that can be made might be “Alice is currently away from her work place”. On the other hand, Alice can set a policy to constrain the use of collected contexts while away from the office, such that the reasoning engine will skip this fact or it will not be shown in the explanation of the inferred result. For example, if Alice prohibits the use of her location to infer her arrival time, people will receive a notice of “Alice being late to the meeting, but we have no information about why or how we know it”. In this case, the explanation that uses Alice’s location is “eclipsed” by the privacy constraints.

Thus, aside from the context representations, a rich representation of the privacy constraints associated under various contexts also needs to be developed.

Provenance of Inferred Situation and Contextual Information

Provenance information is important for data consumers to evaluate the quality of the information in order to use it appropriately. Two types of provenance information are required in situation awareness, one is the provenance of a inferred situation, another is the provenance of a piece of contextual information. Because inferences can be made from multiple sources, it is important to understand more information about the sources to justify the validity of an inferred situation. For example, Alice may have no problem with an inference of her being late to the meeting due to her current location and her schedule with the doctor. But it may be problematic if the same situation (Alice being late to the meeting) is inferred by a pattern of contexts showing that Alice always has to get a cup of coffee at this time everyday. The provenance of an inferred situation will show whether contexts collected about Alice are used appropriately, and will help the information consumer to evaluate the quality of the rule and the quality the contexts.

Traceability and Explanation

The system should be capable of tracing the data that led to a particular context to be inferred or the sharing of particular data or context and proving an explanation for it. Indirectly, this helps the accountability of consumption of the data under question. While accountability in social situations and data sharing happens in different ways (or sometimes does not happen), traceability is important to understand the implications and diagnosing errors. This may lead to changes in the privacy interface, which will in turn be reflected as changes to privacy policies. The explanations help users to understand how inferences are made in human readable forms with some transformations from the justification tree. A user can create a policy to restrict how much explanation should be generated for an inferred situation, because sometimes even the explanations are sensitive. In our scenario, Alice can set privacy constraints in the use of location context as explanation, thus even when location context is used in the inference, it won’t be shown in the explanation.

Conclusion

In this paper, we motivate and illustrate the boundaries and relations between concepts learned using machine learning techniques and semantic representations. We also describe how constraints arising from privacy preferences can be incorporated in the resulting model. We have shown how rules can be expressed over the context artifacts represented in the ontology to aid reasoning and inference of situations. We have further shown that incorporating the privacy constraints on the model leads to reasoning over missing or hidden data, thereby bringing more challenges to the situation awareness system.

As a next step, we have been collecting datasets and developing machine learning based techniques to derive context artifacts. The data currently being collected include a number of sensors, audio, GPS, WiFi, Bluetooth, etc. and will incorporate application contexts going forward. As a next step, the extracted context artifacts will be used to instantiate the ontology elements and experiments would be done on inferring situations via reasoning using the framework described in this paper.

Acknowledgments

We would like to thank Hal Abelson, Jim Dolter, Lenny Grokop, Jinwon Lee and Sanjiv Nanda for their valuable contributions.

References

Chen, H.; Perich, F.; Finin, T.; and Joshi, A. 2004. Soupa: Standard ontology for ubiquitous and pervasive applications. In *International Conference on Mobile and Ubiquitous Systems: Networking and Services*.

Corradi, A.; Montanari, R.; and Tibaldi, D. 2004. Context-based access control for ubiquitous service provisioning. In *COMPSAC*, 444–451. IEEE Computer Society.

Dey, A. K. 2001. Understanding and using context. *Personal Ubiquitous Comput.* 5(1):4–7.

Eagle, N., and Pentland, A. S. 2006. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.* 10:255–268.

Kagal, L., and Abelson, H. 2010. Access control is an inadequate framework for privacy protection. In *W3C Privacy Workshop*.

of Labor Statistics', U. B. 2009. Meeting and convention planners. In *http://www.bls.gov/oco/ocos298.htm (Retrieved May 4, 2011)*.

Sadeh, N.; Hong, J.; Cranor, L.; Fette, I.; Kelley, P.; Prabaker, M.; and Rao, J. 2009. Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing* 13(6):401–412.

Shih, F., and Paradesi, S. 2010. Saveface: Save george's faces in social networks where contexts collapse. In *IAB/W3C Internet Privacy Workshop*.

Toninelli, A.; Montanari, R.; Kagal, L.; and Lassila, O. 2006. A semantic context-aware access control framework for secure collaborations in pervasive computing environments. In Cruz, I. F.; Decker, S.; Allemang, D.; Preist, C.; Schwabe, D.; Mika, P.; Uschold, M.; and Aroyo, L., eds., *International Semantic Web Conference*, volume 4273 of *Lecture Notes in Computer Science*, 473–486. Springer.

Weitzner, D. J.; Abelson, H.; Berners-Lee, T.; Hanson, C.; Hendler, J.; Kagal, L.; McGuinness, D. L.; Sussman, G. J.; and Waterman, K. K. 2006. Transparent accountable data mining: New strategies forprivacy protection. Technical report, Massachusetts Institute of Technology Computer Science and Arti

cial Intelligence Laboratory.

Zhang, G., and Parashar, M. 2004. Context-aware dynamic access control for pervasive applications. In *Proceedings of the Communication Networks and Distributed Systems Modeling and Simulation Conference*, 21–30.