Multimodality in Pervasive Environment

Marco Anisetti, Valerio Bellandi, Paolo Ceravolo and Ernesto Damiani

Abstract Future pervasive environments are expected to immerse users in a consistent world of probes, sensors and actuators. Multimodal interfaces combined with social computing interactions and high-performance networking can foster a new generation of pervasive environments. However, much work is still needed to harness the full potential of multimodal interaction. In this paper we discuss some short-term research goals, including advanced techniques for joining and correlating multiple data flows, each with its own approximations and uncertainty models. Also, we discuss some longer term objectives, like providing users with a mental model of their own multimodal "aura", enabling them to collaborate with the network infrastructure toward inter-modal correlation of multimodal inputs, much in the same way as the human brain extracts a single self-conscious experience from multiple sensorial data flows.

1 Multimodal systems: Introduction

Humans naturally communicate with each other in a multimodal fashion: we speak, gesture, gaze and move generating a rich, multi-streamed flow of multimedia information. Interacting with machines has traditionally been a much simpler affair, typically generating a single flow of uniform events like the discrete mouse clicks entered sequentially in a graphical user interface. As the global information infrastructure is becoming more and more pervasive, however, digital business transactions are performed in diverse situations, using a variety of mobile devices and across multiple communication channels. Rather than being forced to assume a fixed, pre-set position in front of a machine, users move freely around their work environment, starting and monitoring different transactions. Mobile terminals get

Università degli Studi di Milano,
Dipartimento di Tecnologie dell'informazione,
e-mail: {marco.anisetti, valerio.bellandi, paolo.ceravolo, ernesto.damiani,}@unimi.it

smaller and lighter, yet at the same time the requirements to be able to interact with pervasive applications keep expanding. Terminal devices are increasingly equipped with sensors, such as video cameras or audio microphones, capable of collecting information from the environment. Our own voice, hands, and whole body, once augmented by sensors (e.g. of pressure or acceleration), become the ultimate mobile multimodal input devices. In this new paradigm of multimodal access to networked media, a much richer context representation regarding both users and the resources they access is made available to applications. The outcome of an interaction may well depend of where is the user when a certain application-related event takes place, where is she headed, or even whether is she sitting at her desk alone or walking accompanied by others.

However, multimodal context information is based on sensor data that are hardly ever perfect or certain, especially within unsupervised environments. Still, theories and models proposed so far for representing and managing sensor data are mostly aimed at ensuring semantics-aware interoperability of the sensor infrastructure [21], leaving uncertainty management aside.

Lack of attention for uncertainty is partly due to the idea that different modes can confirm - or disprove -each other's results. Early multimodal systems were based on joint recognition of active modes, such as speech and handwriting, for which there is now a large body of research work. Today, context-aware systems sense and incorporate data about illumination, noise level, location, time, people other than the user, as well as many other pieces of information to adjust their model of the user's environment. However, the emergence of novel pervasive computing applications, which combine active interaction modes with passive modality channels based on perception, context, environment and ambience [1, 2, 3], has raised new challenges linked to the imprecision and time-dependence of multimedia predicates, and to difficulties in conjoining facts coming from different modal streams. We argue that inherent uncertainty of sensor data cannot be hidden simply by representing them as crisp information. Facts representing users position and posture (e.g., as shown in a video feed), for instance, are semantically very different from traditional database records, being both highly dynamic and uncertain. Designing and implementing systems that take the best advantage of recognition-based modalities of interaction and multi-sensory observations is difficult. In pervasive environments, sensors that can capture data about the user's physical state or behavior, have been used to gather cues which can help the system perceive users' status [4, 5, 6]; however, these attempts have only very partially succeeded, due to problems in using different modalities to support or disprove one another. Our lack of understanding of how modalities must be combined in the user interface often leads to flawed understanding of the user's intent. Short-term research objectives include solving well-known technical issues of traditional multimodal interaction (e.g. the one carried out via camera, speech and pen interfaces). Natural modalities of interaction, such as speech and gestures, rely on recognition-based technologies, which are inherently error prone. In pervasive computing applications, where the capture and the analysis of passive modes are key, errors are much greater. Issues to be handled include managing uncertain and error prone sources having heterogeneous uncertainty models in a robust

and consistent way. Today strategies for uncertainty reduction mostly work at the interface level, either by constraining user behavior into less error-prone interaction (i.e. "error reduction by design"), or by exploiting other information coming from other modalities (i.e. "error reduction by cross-modality"). However, it is important to remember that in pervasive environments users may not even be aware that their behavior is monitored by a system. They may also have a wrong understanding of what data the various devices capture, and how it is used. Traditional methods of cross-modality error correction are ill adapted to pervasive computing applications and research is urgently needed to better understand user behavior when faced with errors in this type of application. Mid- and long-term research will address putting the human in the loop of multimodal interaction, not only as a source of sensor data but as an integral part, fully in control of the process of capturing and understanding her own multimodal flows [7]. For instance, users will be supported in forming new mental models of the network and of the networked media they interact with. These mental models will provide them with effortless awareness of what data about them is captured and recorded, and how it is used, enabling the development new user-centric strategies to cope with errors in pervasive computing applications.

Like its sensor-less predecessor, the coming Semantic Web of Sensors is focusing on representation formats, for handling sensor context, (e.g. source of information, temporal location, dependencies and so on) rather than on handling uncertainty. The reason is probably that uncertainty (e.g., about locations or motion parameters) is assumed to be have been successfully handled by some sort of low-level layer. In this paper, we argue that this is not the case. Multimodal interaction involves not only uncertain sensor data, but also uncertain inferences based on these data; and the nature of inference uncertainty (e.g. whether it is frequency-based probability or a belief) depends on application semantics, and cannot be handled by any low-level layer.

Mathematical models for reasoning with uniformly uncertain information have been successfully applied in several situations, but predicates inferred from heterogeneous sensor data exhibit different types of uncertainty (for example, sensor-based predicates like "user accompanied by someone" and "user close to door") and require hybrid reasoning strategies.

An *Ontology of Uncertainty*, like the one proposed by the W3C's UR3W-XG incubator group [1], provides an important first step: a vocabulary to annotate different sources of information with different types of uncertainty. Here we argue that such annotations can be mapped to hybrid reasoning and representation strategies.

As a proof of concept of this approach we present, in this paper, a *Patient Monitoring System* (PMS) implementing a semantics-aware matching strategy, where (i) the semantics of each assertion is represented explicitly as instances of an ontology (ii) the representation of matchings is also based on a specific ontological representation. An uncertainty type is assigned to each relation using SWRL rules, this allows to divide the knowledge base in sub-parts according with the specific uncertainty. The Ontology of Uncertainty, proposed by W3C's UR3W-XG incuba-

---

[1] http://www.w3.org/2005/Incubator/urw3/XGR-urw3-20080331/

tor group, allows an explicit definition of the various types uncertainty. Assigning to each model a reasoner process it then possible to manage different independent sources of information. In our case, the different sources of information derive from sensors producing data of different nature (i.e. video, audio, RFID, positioning sensor).

2 Uncertain Information Representation and Reasoning Issues

Uncertainty falls at a meta-level with respect to truth; it arises when the available knowledge does not provide sufficient information to decide if a statement is true or false. Many researchers have tried to classify uncertainty types, starting from *Epistemic Uncertainty*, that comes from the limited knowledge of the agent that generates the statement , and proceeding to *Aleatory Uncertainty*, which is intrinsic in statements reporting quantitative observations of the world, including scientific ones.Uncertainty can also be classified according to its source: *Objective* if it derives from a repeatable observation and *Subjective*, if the uncertainty in the information is derived from an informal evaluation.
Depending on the "coordinates" of the uncertainty to deal with, a certain model (such as Fuzzy theories, Probabilistic Theories and Possibility theory) can be more suitable than another.
Once we are resigned to (i) establish a unified vocabulary about uncertainty models, such as gradual truth values and probabilities (ii) use it for building meta-assertions describing the type of uncertainty of predicates inferred from sensor data flows, we face the problem of what to do with these meta-assertions, i.e. how to take them into account when composing predicates inferred from sensor data. Unfortunately this is a difficult problem. As stated in [8] probability and possibility theories are not fully compositional with respect to all the logical connectives, without a relevant loss of expressiveness. Some work in this direction has nevertheless been done by imposing restrictions to the expressiveness of the logics. Among the most relevant studies, [13] proposes a definition of possibilistic fuzzy description logics by associating weights, representing degrees of uncertainty, to fuzzy description logic formulas[2]. Other works like [9, 10] define probabilistic description logics programs by assigning probability degrees to fuzzy description logics programs.
In [14] the authors propose a framework for sharing information between three different models of uncertainty, where the fuzzy linguistic truth values are propagated through the three models in a non monotonic way, by exploiting the extension principle [17] and aggregation of linguistic values. . Generally speaking, the choice between uncertainty composition techniques depends on the situation. A shared representation of uncertainty types would indeed facilitate automatic selection of a composition technique. For this reason the URW3-XG1 working group has proposed an ontology (called *Ontology of Uncertainty*) as a generic meta-model for

---

[2] An extension of the Fuzzy Description Logics in the field of Possibility theory has been presented also in [12]

representing the semantics of the uncertainty in various assertions. This ontology is designed for a flexible environment, where different uncertainty types can arise in the same knowledge base, so the selection of the correct model for inference is driven by the information in the ontology. Of course, the URW3-XG incubator group was concerned with representation only, and did not specify how to deal with situations where more than one model is involved in the inference process; this is exactly the open issue we want to address. In the literature we are not aware of hybrid reasoning processes, which can handle the flexible integration of different models. In [16, 14, 15, 11] the interoperability has been studied and defined on a set of selected inference models. Adding new models to the framework can easily result in a revision of the underlying theory.

3 Nursing Home Monitoring Framework

In this Section, we shall briefly present an example of handling heterogenous uncertaintiy types in access control. In the logics-based approach to access control policy evaluation [19], evaluating a policy means computing an inference. In other words, for each access request r to a resource R, the policy evaluation system need to evaluate whether the access policy PR implies the request r logically, taking into account all available predicates representing the context in which the access takes place. If this evaluation terminates correctly, access is granted/denied depending on its outcome. Otherwise, a *negotiation phase* takes place to extract additional information from the user and/or the environment, so that the evaluation can be concluded. As said before, predicates for multimodal pervasive environment can support an extended context of interaction for each user and resource in the environment, modeling their state and spatial-temporal relationships with other users/resources [20]. Here, context is not the static situation of a predefined environment; rather, it is a dynamic part of the process of interacting with a changing environment, composed of mobile users and resources [18]. The environment is monitored by different sensors producing assertions on users and resources that are annotated with the model of uncertainty characterizing the specific sensor. According to the model specified the suitable inference process is activated[3] A reconciliation among the different inferences is computed only at the end of process, allowing to evaluate the conjunction of their results. Several techniques can be envisioned to deal with such reconciliation [22]. To provide an illustration of our method we introduce an example related to a healthcare scenario. The general idea is that the access to specific areas of a Nursing Home are allowed only if a doctor and its patient are together. To evaluate this condition the system is organized in three independent inference models. A first model take in inputs video signals and using a probabilistic model evaluate the identity and the role of the persons inside the focus of the camera. A second model,

---

[3] Note that, in order to avoid the introduction of inconsistencies along the inference process a constraint of *safeness* may be applied to the flow of inference. This constraint implies that if a model A is using the conclusions of a model B, B cannot use conclusions from A in any premises.

using a fuzzy approach, evaluates if the persons inside the focus of the camera are "near" the door. A third model, taking in input the conclusions of the second model plus the data available in a database evaluate if the persons are currently in the right relation on to the other. The conclusions from the different models are conjunctively evaluated to verify if the access request is implied in the access policy.
The access policy describing our example is expressed in assertion 1:

$$Open(door1) \rightarrow Patient(x) \wedge Doctor(y) \wedge nearTo(x,door1) \\ \wedge nearTo(y,door1) \wedge careGiver(y,x). \tag{1}$$

The data expressing the request to be evaluated are obtained by the system from different sources and evaluated according to the specific uncertainty model associated to each. These data can be interpreted as independent set of assertions that any source produces and tags with the appropriate uncertainty model. This way the system activates the right inference service, collects all the conclusions and finally tries to enforce the access policy. In our example the sensors in the environment produces three source of knowledge expressed in the assertions 2, 3, and 4.

$$\rightarrow hasUncertainty(Sentence1, 0.8) \tag{2}$$
$$\rightarrow saidBy(Sentence1, camera1)$$
$$\rightarrow saidAbout(Sentence1, eyes\_distance)$$
$$\rightarrow uncertaintyModel(Sentence1, Probability)$$
$$Sentence1[\rightarrow user1]$$

$$\rightarrow hasUncertainty(Sentence2, 0.9)$$
$$\rightarrow saidBy(Sentence2, camera1)$$
$$\rightarrow saidAbout(Sentence2, iris)$$
$$\rightarrow uncertaintyModel(Sentence2, Probability)$$
$$Sentence2[\rightarrow user1]$$

$$\rightarrow hasUncertainty(Sentence3, 0.3)$$
$$\rightarrow saidBy(Sentence3, camera1)$$
$$\rightarrow saidAbout(Sentence3, eyes_d istance)$$
$$\rightarrow uncertaintyModel(Sentence3, Probability)$$
$$Sentence3[\rightarrow user3]$$

$$\rightarrow hasUncertainty(Sentence4, 0.7)$$
$$\rightarrow saidBy(Sentence4, camera1)$$
$$\rightarrow saidAbout(Sentence4, iris)$$
$$\rightarrow uncertaintyModel(Sentence4, Probability)$$

$$Sentence4[\rightarrow user3]$$

$$\rightarrow hasUncertainty(Sentence5, 0.9)$$
$$\rightarrow saidBy(Sentence5, camera1)$$
$$\rightarrow saidAbout(Sentence5, eyes_distance)$$
$$\rightarrow uncertaintyModel(Sentence5, Probability)$$
$$Sentence5[\rightarrow user2]$$

$$\rightarrow hasUncertainty(Sentence6, 0.9)$$
$$\rightarrow saidBy(Sentence6, camera1)$$
$$\rightarrow saidAbout(Sentence6, iris)$$
$$\rightarrow uncertaintyModel(Sentence6, Probability)$$
$$Sentence6[\rightarrow user2]$$

$$\rightarrow hasUncertainty(Sentence7, 0.7) \qquad (3)$$
$$\rightarrow saidBy(Sentence7, camera1)$$
$$\rightarrow saidAbout(Sentence7, nearTo)$$
$$\rightarrow uncertaintyModel(Sentence7, Fuzzy)$$
$$Sentence7[\rightarrow nearTo(user1, door1)]$$

$$\rightarrow hasUncertainty(Sentence8, 0.8)$$
$$\rightarrow saidBy(Sentence8, camera1)$$
$$\rightarrow saidAbout(Sentence8, nearTo)$$
$$\rightarrow uncertaintyModel(Sentence8, Fuzzy)$$
$$Sentence8[\rightarrow nearTo(user2, door1)]$$

$$\rightarrow hasUncertainty(Sentence9, 1) \qquad (4)$$
$$\rightarrow saidBy(Sentence9, NHDB)$$
$$\rightarrow saidAbout(Sentence9, 12h32m23s)$$
$$\rightarrow uncertaintyModel(Sentence9, Temporal)$$
$$Sentence9[\rightarrow careGiver(user1, user2)]$$

The first source of knowledge is evaluated by a probabilistic reasoner that aggregates the assertions related to the same entity (users in our example) and according to the parameter under analysis (eyes_distance or in our example) computes the

union or intersection of the corresponding probabilities.

$$user1 : p(eyes_distance) \cap p(iris) = 0.8 * 0.9 = 0.72$$

$$user2 : p(eyes_distance) \cap p(iris) = 0.9 * 0.9 = 0.81$$

$$user3 : p(eyes_distance) \cap p(iris) = 0.3 * 0.7 = 0.21$$

Finally a threshold is applied on the assertions, saving only the assertions whose confidence is high enough.
The second source of knowledge is evaluated by a fuzzy reasoner that is configured to apply a threshold not on the single assertions but on the intersection or union of the assertions. In our example we consider the intersection, computed via a t-conorm (the *min* operator).

$$(nearTo(user1, door1) \cap nearTo(user2, door1)) = min(0.7, 0.8) = 0.7$$

If in the example we take 0.7 as a threshold, the assertion in the knowledge base are the following:

$$\rightarrow user1 \tag{5}$$

$$\rightarrow user2 \tag{6}$$

$$\rightarrow nearTo(user1, door1) \tag{7}$$

$$\rightarrow nearTo(user2, door1) \tag{8}$$

The third source of knowledge is not generated by a sensor; rather, it is produced by the system querying the Nursing Home database taking impulse from the assertions generated by the sensors. The goal is to expand the knowledge provided by the sensors with the contextual knowledge of the specific domain in order to verify if the access request (detected directly by observing the environment) is implied in the access rules. The type of uncertainty handled at this level is not related to vagueness or incompleteness as in the previous examples; rather it is related to the fact that the validity of the assertions is temporary, and must be verified within the temporal range of the specific situation. Querying the database one obtains the assertions in 9.

$$user1? :\rightarrow Doctor(user1) \tag{9}$$

$$user2? :\rightarrow Patient(user2) \tag{10}$$

In conclusion the system, computing the conjunction all available sources of knowledge, is able to correctly evaluate assertion [7] that represents the context of the access request, directly derived by observing the environment.

$$\rightarrow Patient(user2) \wedge Doctor(user1) \wedge nearTo(user1, door1)$$
$$\wedge nearTo(user2, door1) \wedge careGiver(user1, user2). \tag{11}$$

4 Conclusions

Like its predecessor, the coming Semantic Web of Sensors promises to be more focused on handling sensor data context, (e.g. source of information, temporal location, dependencies and so on) than on handling uncertainty. The reason is probably that uncertainty (e.g., about locations or motion parameters) is assumed to be successfully handled at the individual sensor level. In this paper, we argue that this is not the case. Multimodal interaction involves not only uncertain sensor data, but also uncertain inferences based on these data. Although mathematical models for reasoning with uncertain information have been successfully applied in several situations, predicates inferred from heterogeneous sensor data exhibit different types of uncertainty (for example, sensor-based predicates like "user accompanied by someone" and "user close to door" involve two very different notions of uncertainty). In this paper, we described a simple Use Case. supporting the idea of using W3C Ontology of Uncertainty to write meta-assertions guiding hybrid reasoning strategies on sensor data.

References

1. G.D. Abowd and E.D. Mynatt: Charting past, present, and future research in ubiquitous computing, ACM Transactions on Computer Human Interaction 7 (1) (2000) 29-58. Germany, July 2004, pp. 292-305.
2. M.A. Feki, S. Renouard, B. Abdulrazak, G. Chollet and M. Mokhtari: Coupling context awareness and multimodality in smart homes concept, Lecture Notes in Computer Science 3118 (2004) pp.906-913
3. A. Oikonomopoulos, I. Patras and M. Pantic, Human action recognition with spatiotemporal salient points, IEEE Transactions on Systems, Man and Cybernetics. Part B 36 (3) (2006) 710-719.
4. M. Anisetti and V. Bellandi, Emotional State Inference Using Face Related Features, New Directions in Intelligent Interactive Multimedia Systems and Services - 2, Pp. 401-411 July 09, 2009
5. E. Damiani, M. Anisetti and V.Bellandi, Toward Exploiting Location-based and Video Information in Negotiated Access Control Policies, International Conference on Information Systems Security , Calcutta, India, 2005.
6. M. Pantic, Affective Computing, in: M. Pagani (Ed.), Encyclopedia of Multimedia Technology and Networking, vol. 1, Hershy, PA, USA, Idea Group Reference, May 2005, pp. 8-14.
7. E. Damiani, R. Khosla and I. K. Sethi, Intelligent multimedia multi-agent systems: a human-centered approach, Kluwer, Sept. 2001
8. Didier Dubois and Henri Prade. Can we enforce full compositionality in uncertainty calculi. In In Proc. of the 11th Nat. Conf. on Artificial Intelligence (AAAI-94), pp. 149-154. AAAI Press / MIT Press, 1994.
9. Thomas Lukasiewicz and Umberto Straccia. Description logic programs under probabilistic uncertainty and fuzzy vagueness. In ECSQARU, pp. 187-198, 2007.

10. Thomas Lukasiewicz and Umberto Straccia. Uncertainty and vagueness in description logic programs for the semantic web, 2007.
11. Grigoris Antoniou and Antonis Bikakis. Dr-prolog: A system for defeasible reasoning with rules and ontologies on the semantic web. IEEE Trans. Knowl. Data Eng., 19(2):233-245, 2007.
12. Fernando Bobillo, Miguel Delgado, and Juan Gomez-Romero: Extending fuzzy description logics with a possibilistic layer. In URSW, volume 327 of CEUR Workshop Proceedings. CEUR-WS.org, 2007.
13. Didier Dubois, Jerome Mengin, and Henri Prade: Possibilistic uncertainty and fuzzy features in description logic. A preliminary discussion. In E. Sanchez, editor, Fuzzy logic and the semantic web, pp. 101-113. Elsevier, 2006. DMenP001.
14. X. Luo, C. Zhang, and N. R. Jennings: A hybrid model for sharing information between fuzzy, uncertain and default reasoning models in multi-agent systems. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(4):401-450, 2002.
15. William James Van Melle: A domain-independent system that aids in constructing knowledge-based consultation programs. PhD thesis, Stanford, CA, USA, 1980.
16. E. H. Shortlife and B. G. Buchanan: A model of inexact reasoning in medicine. pp. 259-275, 1990.
17. L.A. Zadeh: Fuzzy sets. Information Control, pp. 338-353, 1965.
18. J. Coutaz, J.L. Crowley, S. Dobson and D. Garlan : Context is Key. Comm. of the ACM 48 (3), March 2005
19. P. A. Bonatti and P. Samarati : Logics for Authorization and Security. Logics for Emerging Applications of Databases 2003: pp. 277-323
20. E. Damiani, M. Anisetti and V.Bellandi: Toward Exploiting Location-based and Video Information in Negotiated Access Control Policies, International Conference on Information Systems Security (ICISS 05), India, 2005.
21. Amit Sheth and Cory Henson and Satya S. Sahoo: Semantic Sensor Web,IEEE Internet Computing, 2008.
22. Fernando Bobillo, Miguel Delgado and Juan GómezRomero: Crisp Representations and Reasoning for Fuzzy Ontologies, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 17(4):501-530, 2009