Incomplete and uncertain data handling in context-aware rule-based systems with modified certainty factors algebra

Szymon Bobek, Grzegorz J. Nalepa

AGH University of Science and Technology

RuleML 2014
20 August 2014, Prague

http://geist.agh.edu.pl

Outline I

Outline

Mobile context-aware systems (mCAS)

Context

- Where you are, who you are with, what resources are nearby (Schillit)
- Any informaiton that can be used to characterize the situation of an entity (Dey)
- Individuality, activity, location, time, relations (Zimmerman)
- Set of variables that may be of interest for an agent and that influence its actions (Bolchini)

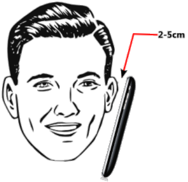
Aware

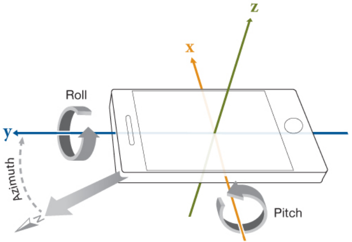
- Artificial intelligence methods

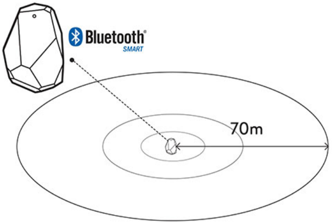
Systems

- Intelligent homes, intelligent cars, robotics
- Ambient intelligence, pervasive environments, ubiquitous computing
- Mobile computing (location aware mobile applicaitons)
- Intelligent software (contextual advertising, etc.)

Mobile environment and uncertainty



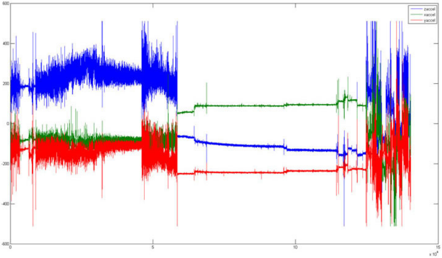


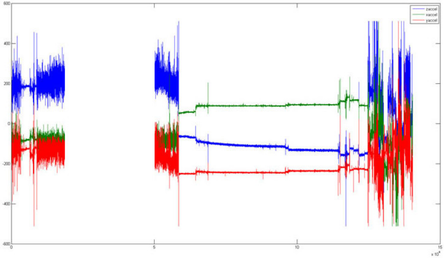


Different types of uncertainty

High-level classification

1. Uncertainty due to lack of knowledge – that comes from incomplete information both at the model level or if the information is not provided by the sensors,
2. Uncertainty due to lack of semantic precision – that may appear due to semantic mismatch in the notion of the information,
3. Uncertainty due to lack of machine precision – which covers machine sensors imprecision and ambiguity.

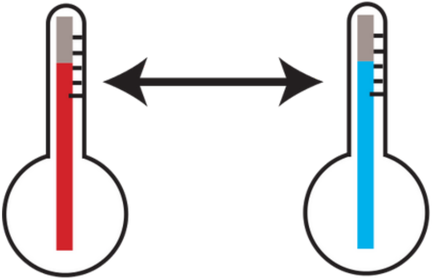




Different types of uncertainty

High-level classification

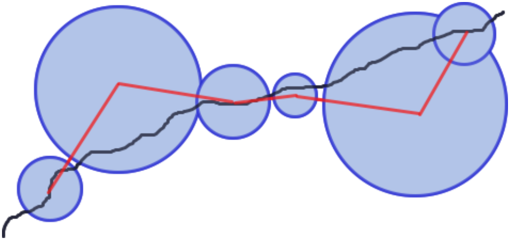
1. Uncertainty due to lack of knowledge – that comes from incomplete information both at the model level or if the information is not provided by the sensors,
2. Uncertainty due to lack of semantic precision – that may appear due to semantic mismatch in the notion of the information,
3. Uncertainty due to lack of machine precision – which covers machine sensors imprecision and ambiguity.



Different types of uncertainty

High-level classification

1. Uncertainty due to lack of knowledge – that comes from incomplete information both at the model level or if the information is not provided by the sensors,
2. Uncertainty due to lack of semantic precision – that may appear due to semantic mismatch in the notion of the information,
3. Uncertainty due to lack of machine precision – which covers machine sensors imprecision and ambiguity.



Different uncertainty modelling and handling mechanisms

	Uncertainty source			
	Lack of knowledge	Semantic imprecision	Machine imprecision	Implementation effort
Probabilistic	◗	❍	●	High
Fuzzy Logic	❍	◗	◗	Medium
Certainty Factors	◗	❍	●	Low
Machine learning	●	❍	●	High

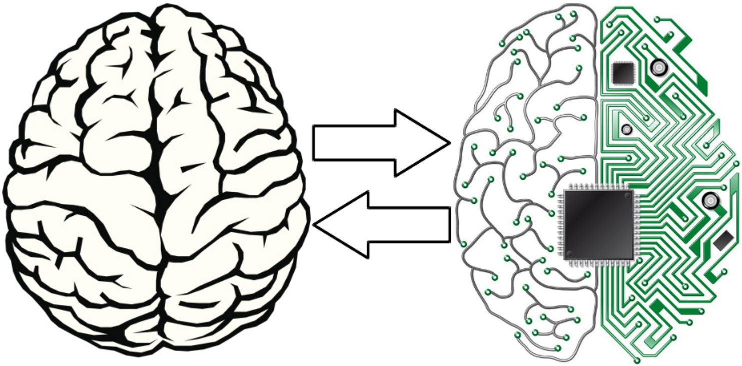
Table : Comparison of uncertainty handling mechanisms. Full circles represent full support, whereas empty circles represent low or no support.

Outline

Mobile environment and uncertainty

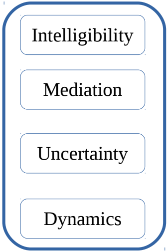
Nature of mCAS

- mCAS are build usually as a user centric systems
- Intelligibility is very important as it may improve users trust to the system
- Mediation may help resolve ambiguity
- The uncertainty is dynamic and such a dynamic should be modelled

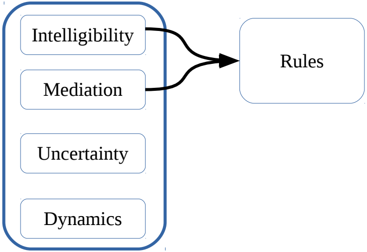


Outline

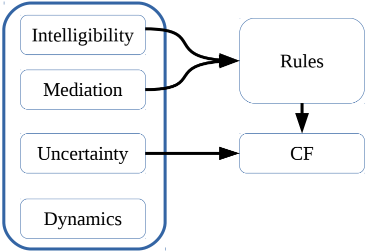
Overall idea



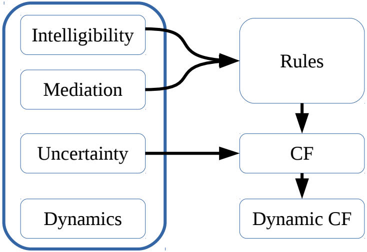
Overall idea



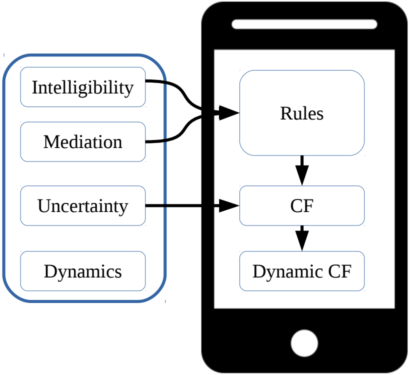
Overall idea



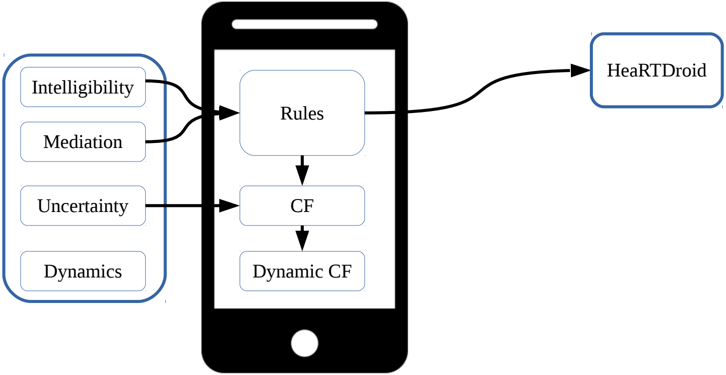
Overall idea



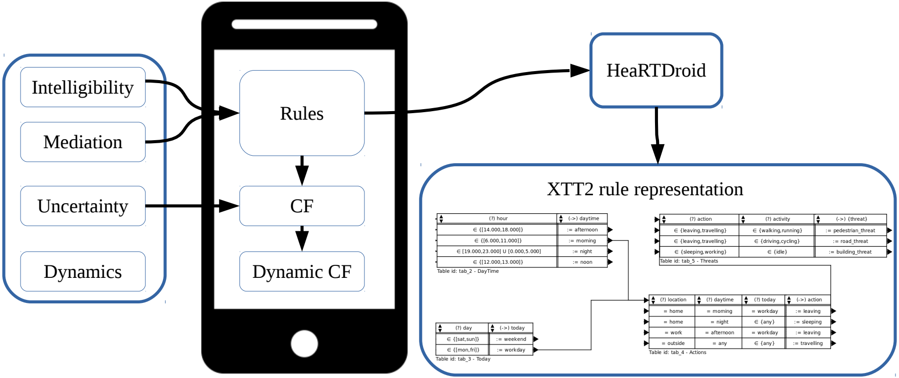
Overall idea



Overall idea



Overall idea



ALSV(FD) logic

XTT2 rule in ALSV(FD) logic

$$(A_i \propto d_i) \wedge (A_j \propto d_j) \wedge \ldots (A_m \propto V_m) \wedge (A_n \propto V_n) \longrightarrow RHS$$

Syntax	Interpretation	Relation
$A_i = d_i$	value of A_i is precisely defined as d_i	eq
$A_i \in V_i$	value of A_i is in V_i	in
$A_i \neq d_i$	shorthand for $A_i \in (\mathbb{D}_i \setminus \{d_i\})$	neq
$A_i \notin V_i$	shorthand for $A_i \in (\mathbb{D}_i \setminus V_i)$	notin

Table : Formulaes for simple attributes

Syntax	Interpretation	Relation
$A_i = V_i$	A_i equal V_i	eq
$A_i \neq V_i$	A_i does not equal V_i	neq
$A_i \subseteq V_i$	A_i is a subset V_i	subset
$A_i \supseteq V_i$	A_i is a superset V_i	supset
$A_i \sim V_i$	A_i has non-empty intersection with V_i	sim
$A_i \not\sim V_i$	A_i has empty intersection with V_i	notsim

Table : Formulaes for generalized attributes

ALSV(FD) logic

XTT2 rule in ALSV(FD) logic

$$(A_i \propto d_i) \wedge (A_j \propto d_j) \wedge \ldots (A_m \propto V_m) \wedge (A_n \propto V_n) \longrightarrow RHS$$

Syntax	Interpretation	Relation
$A_i = d_i$	value of A_i is precisely defined as d_i	`eq`
$A_i \in V_i$	value of A_i is in V_i	`in`
$A_i \neq d_i$	shorthand for $A_i \in (\mathbb{D}_i \setminus \{d_i\})$	`neq`
$A_i \notin V_i$	shorthand for $A_i \in (\mathbb{D}_i \setminus V_i)$	`notin`

Table : Formulae for simple attributes

Example for simple attribute

$$if(activity = driving) \rightarrow$$
$$bluetoothHandset = on$$

Uncertainty in such data

$$activity = walking(30\%)$$
$$activity = cycling(50\%)$$
$$activity = driving(20\%)$$

ALSV(FD) logic

XTT2 rule in ALSV(FD) logic

$$(A_i \propto d_i) \wedge (A_j \propto d_j) \wedge \ldots (A_m \propto V_m) \wedge (A_n \propto V_n) \longrightarrow RHS$$

Example for generalised attribute

$$if\,(language \sim \{en, cz\}) \rightarrow$$
$$recommendConference = RuleML$$

Uncertainty in such data

$$language =$$
$$\{pl(100\%), en(80\%), cz(2\%)\}$$

Syntax	Interpretation	Relation
$A_i = V_i$	A_i equal V_i	`eq`
$A_i \neq V_i$	A_i does not equal V_i	`neq`
$A_i \subseteq V_i$	A_i is a subset V_i	`subset`
$A_i \supseteq V_i$	A_i is a superset V_i	`supset`
$A_i \sim V_i$	A_i has non-empty intersection with V_i	`sim`
$A_i \not\sim V_i$	A_i has empty intersection with V_i	`notsim`

Table : Formulae for generalized attributes

Certainty Factors algebra

Basic notation

- Rule in CF algebra is represented according to formulae:

$$condition_1 \wedge condition_2 \wedge \ldots \wedge condition_k \rightarrow conclusion \tag{1}$$

- Each of the elements of the formulae from equation 1 above can have assigned a certainty factor $cf(element) \in [-1; 1]$
- CF of a conditional part of a rule is determined by the formulae:

$$cf(condition_1 \wedge \ldots \wedge condition_k) = \min_{i \in 1 \ldots k} cf(condition_i)$$

- Certainty factor of conclusion C of a single *i-th* rule is calculated according to a formulae:

$$cf_i(C) = cf(condition_1 \wedge \ldots \wedge condition_k) * cf(rule) \tag{2}$$

Cumulative and disjunctive rules

Disjunctive rules

Disjunctive rules have the same conclusions but are conditionally dependent (i.e. value of any of the conditions determine values of other rules conditions).

$$cf(C) = \max_{i \in 1 \ldots k} \{cf_i(C)\} \tag{3}$$

Cumulative rules

Cumulative rules have the same conclusions and have independent conditions (i.e. value of any of the conditions does not determine values of other rules conditions).

$$cf(C) = \begin{cases} cf_i(C) + cf_j(C) - cf_i(C) * cf_j(C) & \text{if } cf_i(C) \geq 0, cf_j(C) \geq 0 \\ cf_i(C) + cf_j(C) + cf_i(C) * cf_j(C) & \text{if } cf_i(C) \leq 0, cf_j(C) \leq 0 \\ \frac{cf_i(C) + cf_j(C))}{1 - \min\{|cf_i(C)|, |cf_j(C)|\}} & \text{if } cf_i(C) cf_j(C) \notin \{-1, 0\} \end{cases} \tag{4}$$

CF in ALSV(FD)

Disjunctive rules

In particular the formula $A_i \in V_i$ can be translated into a form:

$$(A_i = V_i^0) \vee (A_i = V_i^1) \vee \ldots \vee (A_i = V_i^k)$$

where the V_i^k is a *k-th* element from a subset V_i of domain D_i, and A_i is a simple attribute.

Cumulative rules

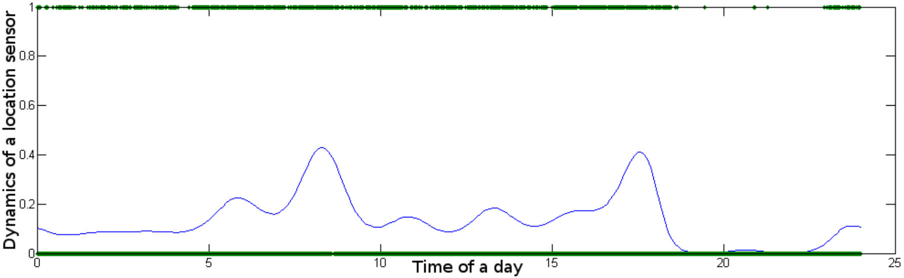
For the general attribute A_i, the formulae of a form $A_i \sim V_i$ can be translated into:

$$(A_i^0 \in V_i) \vee (A_i^1 \in V_i) \vee \ldots \vee (A_i^k \in V_i)$$

where A_i^k is a *k-th* element of a set representing by the general attribute A_i.

Outline

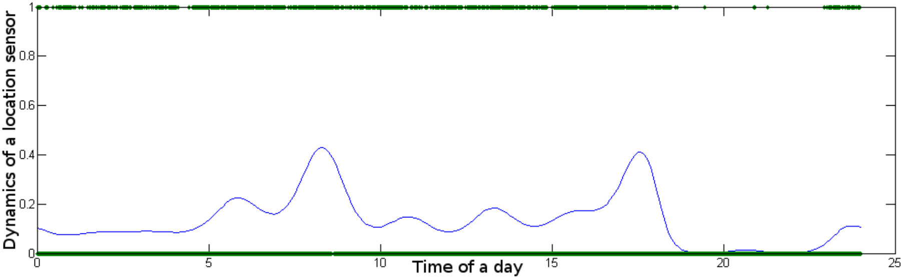
How certainty may change



Expiration time functions

$$cf(V, \Delta t) = \begin{cases} cf(V) * \frac{expiration(A) - \Delta t}{expiration(A)} & \text{if } \Delta t \leq expiration(A) \\ 0 & \text{otherwise} \end{cases}$$

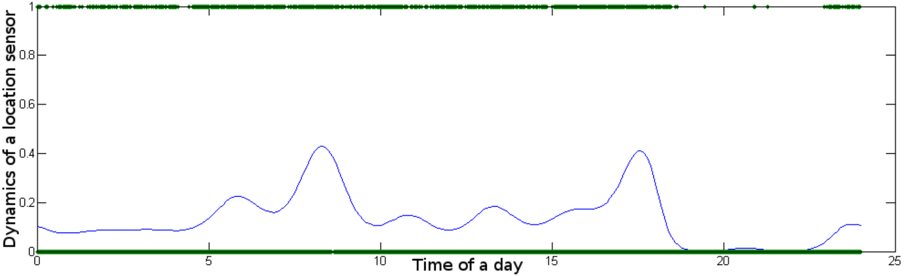
Dynamic expiration times



Dynamics of uncertainty

$$expiration(A, t) = expiration(A) * (1 - dynamic(A, t))$$

Entropy-based expiration times



Entropy-based dynamics

$$expiration(A, n) = expiration(A) * (-log_2 \frac{1}{n} - entropy(A, n))$$

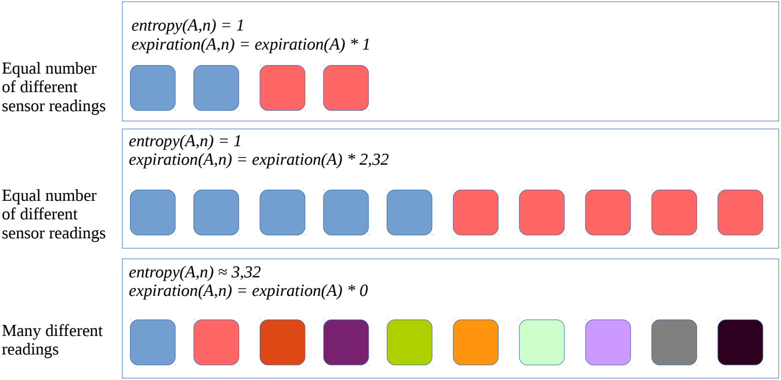
$$entropy(A, n) = -\sum_{x \in X} \frac{x}{n} \log_2 \frac{x}{n}$$

Entropy-based expiration times

Entropy-based dynamics

$$expiration(A,n) = expiration(A) * (-log_2 \frac{1}{n} - entropy(A,n))$$

$$entropy(A,n) = -\sum_{x \in X} \frac{x}{n} \log_2 \frac{x}{n}$$



Outline

POI recommender

(?) weather	(?) {user_profile}	(?) activity	(->) poi
∈ {sunny,cloudy}	~ {eating}	= any	:= outdor-eating
∈ {rainy}	~ {eating}	∈ {walking,running}	:= indoor-eating
∈ {rainy}	~ {eating}	∈ {driving}	:= drivethrough-eating
∈ {rainy,cloudy}	~ {culture,entertainment}	∈ {walking,driving}	:= theatre-cinema
∈ {rainy,cloudy}	= {culture,sighseeing}	∈ {walking,driving}	:= museum
∈ {sunny}	= {sighseeing,culture}	∈ {any}	:= monuments

Table id: tab_2 - Recommendations

Assumed system state

- Weather forecast: *sunny weather* with certainty 0.3, *cloudy* with 0.1, and *rainy* with 0.6.
- How much user is interested inn particular POIs: places for eating – 60%, culture – 20%, entertainment – 80%, sightseeing – 20%.
- the user have been recently walking with certainty 0.8, running with 0.1 certainty and driving with certainty 0.1.

Uncertainties

(?) weather	(?) user_profile	(?) activity	cf(conditions)	*cf(rule)*	*cf(conclusion)*
0.3	0.6	0.8	0.3	1	0.3
0.6	0.6	0.8	0.6	1	0.6
0.6	0.6	0.1	0.1	1	0.1
0.6	0.84	0.8	0.6	1	0.6
0.6	0.36	0.8	0.36	1	0.36
0.3	0.36	0.8	0.3	1	0.3

Assumed system state

- Weather forecast: *sunny weather* with certainty 0.3, *cloudy* with 0.1, and *rainy* with 0.6.
- How much user is interested inn particular POIs: places for eating – 60%, culture – 20%, entertainment – 80%, sightseeing – 20%.
- the user have been recently walking with certainty 0.8, running with 0.1 certainty and driving with certainty 0.1.

Outline

Summary and future work

Summary

- Proposal of uncertainty handling mechanism for mCAS
- Rule-based mechanism ready to easily provide intelligibility and mediation
- Modelling approach for representing dynamic CF

Future work

- Implementation and evaluation on e real-life example
- User feedback for CF updates
- Handling uncertainty on the level of XTT2 model, by translating XTT2 graph into BN
- Gather more data for testing and evaluation: `http://glados.kis.agh.edu.pl`

Summary and future work

(?) hour	(->) daytime
∈ {[14.000,18.000]}	:= afternoon
∈ {[6.000,11.000]}	:= morning
∈ [19.000,23.000] U [0.000,5.000]	:= night
∈ {[12.000,13.000]}	:= noon

Table id: tab_2 - DayTime

(?) action	(?) activity	(->) {threat}
∈ {leaving,travelling}	∈ {walking,running}	:= pedestrian_threat
∈ {leaving,travelling}	∈ {driving,cycling}	:= road_threat
∈ {sleeping,working}	∈ {idle}	:= building_threat

Table id: tab_5 - Threats

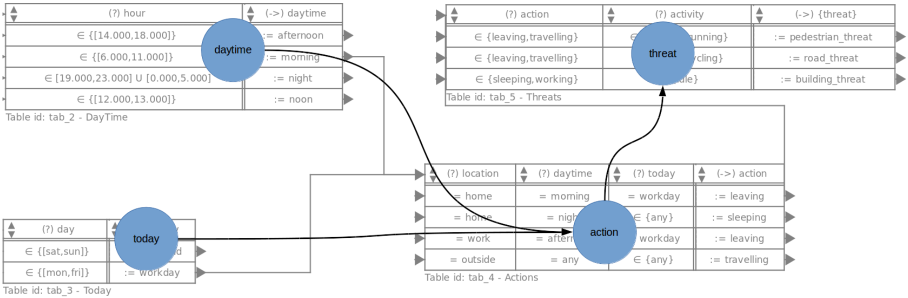
(?) location	(?) daytime	(?) today	(->) action
= home	= morning	= workday	:= leaving
= home	= night	∈ {any}	:= sleeping
= work	= afternoon	= workday	:= leaving
= outside	= any	∈ {any}	:= travelling

Table id: tab_4 - Actions

(?) day	(->) today
∈ {[sat,sun]}	:= weekend
∈ {[mon,fri]}	:= workday

Table id: tab_3 - Today

Summary and future work



Summary and future work

Summary

- Proposal of uncertainty handling mechanism for mCAS
- Rule-based mechanism ready to easily provide intelligibility and mediation
- Modelling approach for representing dynamic CF

Future work

- Implementation and evaluation on e real-life example
- User feedback for CF updates
- Handling uncertainty on the level of XTT2 model, by translating XTT2 graph into BN
- Gather more data for testing and evaluation: `http://glados.kis.agh.edu.pl`

Thank you for your attention!

Do you have any questions?

RuleML 2014

http://geist.agh.edu.pl