Uncertainty Management in Dialogue Systems

Romain Laroche
Microsoft Maluuba

Bernadette Bouchon-Meunier
CNRS, Pierre and Marie Curie University - Par…

Uncertainty Management in Dialogue Systems

Romain Laroche
France Telecom
&
LIP6 * #
{ HYPERLINK "mailto:romain.laroche@orange-ftgroup.com" romain.laroche@orange-ftgroup.com

Bernadette Bouchon-Meunier
* LIP6
104, Président Kennedy avenue,
75016 Paris XVIème
France
Bernadette.Bouchon-Meunier@lip6.fr

Philippe Bretier
France Telecom
2, Pierre Marzin avenue,
22307 Lannion Cedex
France
philippe.bretier@orange-ftgroup.com

Abstract

Dialogue systems find more and more applications but the scientific improvements face difficulties to be transferred into the dialog technologies. France Telecom R&D is designing a new architecture based on uncertainty management in dialogue. Dialogue constraints lead to define a Logical Framework for Probabilistic Reasoning (LFPR) which results are close to the ones of the Evidence Theory and the Assumption-based Truth Maintenance System (ATMS). The paper eventually compares it to the Theory of Hints of Kohlas and Monney.

Keywords: Dialogue System, Uncertainty, Evidence Theory, Theory of Hints, ATMS.

1 Environment and General Considerations

More and more fields require Speech Applications. The Customer Services that can be executed from distance are manifold: Distance Orders, After Sales Services, Shipping Information, … The Customer Services hotlines are very costly and though often overflowed. However, most of the calls concern issues that can be quite easily automated. This explains the rapid development of Speech Applications these last years.

Still, the scientific advances have not been significant during this time because, as soon as a scientific novelty for such systems arises, the development duration grows and even more problematic, the developer must manipulate more and more scientific-level concepts. Another faced difficulty is that the implementations are very little reusable from one service to another, even if they deal with a common field.

In order to cope with these issues, some researchers have tried to automatically process fully data-driven applications [1], [2]. They consider the dialogue cycle (typically Automatic Speech Recognition, Speech Language Understanding, Dialogue Management, Speech Language Generation and Speech Synthesis, see Figure 1) as a Markov Decision Process and optimise each module independently from the others. Aside from the Dialogue Management, each of these modules can be at least partially reused from one application to another and can be learned from annotated corpora provided by the records of the human service it is going to replace. For the automated design of the Dialogue Management, they use a byroad by learning a model of the user on the annotated corpora basis and then by learning the management thanks to this model.

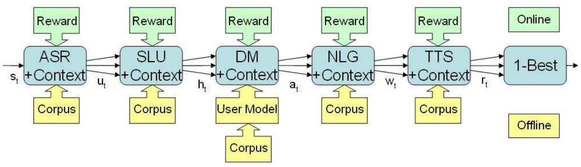


Figure 1: the fully data-driven architecture.

France Telecom R&D approach is quite different because the group needs high quality dialogue services, even at the expense of some more manual design, and, in this scope, we think

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 705–712
Torremolinos (Málaga), June 22–27, 2008

that one should combine expert knowledge with online learning in order to optimise the system.

But as well as the fully data-driven architecture, one of France Telecom R&D's system goals is to develop an end-to-end statistical dialogue architecture with a complete, mathematically precise treatment of uncertainty. This treatment is the topic of this paper.

2 Context

Historically, the dialogue manager of France Telecom R&D's dialogue system is automaton-based. A state of the automaton does not correspond exactly to a state of the system but to a family of states of the system, because some global variables are not represented in the state of the automaton but into the Context module. Most of the choices of transition between one state and another are made after testing the Context. Figure 2 gives an overall picture of the current system.

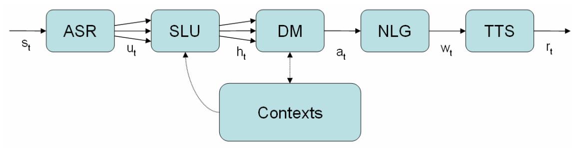


Figure 2: the current system.

The first remark is that the current system is incomplete. The lack of symmetry is obvious and difficult to advocate. The SLU cannot write into the context and once the action determined by the Dialogue Manager, the rest of the architecture is basically a translation of a concept into a string and then a translation of a string into an audio stream.

Figure 3 gives an idea of the target for the system. The automaton of the Dialogue Manager target is the same as the historic one except that it is enhanced by uncertainty management (uncertain context management and stochastic walkthrough of the automaton). For instance, in a state of the automaton, if the transition is determined by the following test: $X < 0$, the state sends a request to the Context Manager which delivers in response an estimation of the probabilities, then the walkthrough continues in every possible branch.

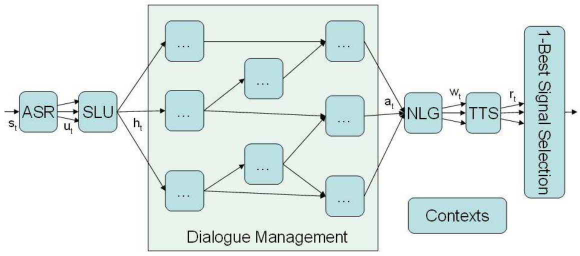


Figure 3: the target system.

3 Uncertainty Management

The main inputs of uncertainty are the statistical models from ASR and SLU. They are supposed to provide probability. Even if at this stage they still fail to be really reliable, a lot of work is done in this direction [3], [4]. As online learning is one of France Telecom R&D goals, it will eventually provide other probability expressed sources of uncertainty. However, probabilistic networks [5], [6] are not suitable for this uncertainty management for the two following reasons.

In a dialogue, the system exchanges signals with the user. Then, the processing of the signal provides the system with derivative information, on the basis of hypotheses. The rules are usually one-way and they do not deliver a full dependency matrix. For instance, the result of the interpretation can be the probability of this proposition: "the user said that he had an internet trouble". But the transition to the probability to the following proposition: "the user has an internet trouble" is not direct because the user might not have said that he had a trouble and still have one.

The dialogue requires a logical framework that enables to deal with rules and to find non trivial inferences and inconsistencies such as "it's unlikely that the user wants to subscribe both ADSL and cable internet connections".

The paper describes the Logical Framework for Probabilistic Reasoning (LFPR). The automaton architecture can be emulated in the LFPR and the existing logical rules of the context component are to be integrated in the LFPR too.

4 Logical Framework for Probabilistic Reasoning

4.1 The Sets of Worlds

The LFPR makes a hard distinction between the probability of a proposition and the probability that this proposition is demonstrated. When the system receives a piece of information about a proposition with a probabilistic reliability, then the proposition is demonstrated with this reliability. However, this reliability is not necessarily the probability of the proposition. For instance, the probabilistic piece of information provided by the interpreter "the user said that he had an internet trouble" provides a demonstration of the proposition "the user has an internet trouble". This demonstration is founded on the hypothesis of the former good interpretation. However, even if the hypothesis is erroneous, it might be true that the user has an internet trouble.

The system is considered as an agent. Among all possible states of the world, only one is right and the agent tries to gather information in order to reduce the set of worlds he considers possible. In the logical framework, a world w is a subpart of the set of the well-formed formula ($\mathbf{WFF}$) such that:

(**i**) w is deductively closed

(**ii**) $\forall p \in \mathbf{WFF}$, if $p \notin w$, then $\neg p \in w$

The set of worlds is noted $\mathbf{W}$. A possible world w is a consistent world, that is to say a world where $\bot \notin w$. As $\bot$ subsumes all the well-formed formula, there is one and only one impossible world that is noted $w_\bot = \mathbf{WFF}$. The probability measure $\mu(\mathbf{S})$ on world sets is considered and the measure $\mu(p)$ of a proposition p is defined as the measure of its support on $\mathbf{W}$. The following results are direct:

(**iii**) $\forall w \in \mathbf{W}, \top \in w$

(**iv**) $\mu(\top) = 1$

(**v**) $\mathbf{W}^\top = \mathbf{W} \setminus \{w_\bot\}$

Where $\top$ is the tautology and $\mathbf{W}^\top$ is the set of possible worlds.

A demonstration d is a tuple of a proposition p and a set of worlds $\mathbf{D}$ where the demonstration stands. A demonstration can be a hypothesis itself, id est directly dependent on a piece of information generated by a module. Or a demonstration can be implied by a set of demonstrations $\{d_k\}$ thanks to a rule. In this case, $\mathbf{D}$ is the intersection of the supports $\mathbf{D}_k$ of the demonstrations d_k.

A demonstration d of a proposition p is said to be reliable in a world w if d is valid in w, which implies necessarily that p is true in w. The statistical rules in the system provide a probabilistic number that can be identified to the probabilistic measure of the support of the reliability of the demonstration they generate.

A demonstration d is said to be applicable in a world w if the semantics allow d to be applied in w. These semantics considerations include every condition that is not part of the reliability.

The support $\mathbf{D}_d$ of a demonstration d is by definition the set of worlds where d is reliable and applicable. If $\mathbf{R}_d$ is the support of the reliability of d and $\mathbf{A}_d$ the support of the applicability d, then the following equality stands: $\mathbf{D}_d = \mathbf{A}_d \cap \mathbf{R}_d$.

The following example illustrates these concepts. In an internet support service, it is supposed that a statistical rule models that most of the connection problems are consequent to a misunderstanding in the hardware connections with the internet box. It is supposed also that whenever the system receives a call, it automatically tests the line and this process can eventually provide the demonstration of another reason for the internet trouble. With a reliability-only based inference system, the system would infer that the internet box has not been properly connected from a line test that revealed a trouble.

I	$\neg line \Rightarrow \neg internet$	$\mathbf{R}_{\mathrm{I}}$	$\mathbf{A}_{\mathrm{I}}$
II	$\neg internet \Rightarrow \neg connectedbox$	$\mathbf{R}_{\mathrm{II}}$	$\mathbf{A}_{\mathrm{II}}$
III	$\neg line \Rightarrow \neg connectedbox$	$\mathbf{R}_{\mathrm{III}}$	$\mathbf{A}_{\mathrm{III}}$

The rule I is a definition rule and it is always reliable: $\mathbf{R}_{\mathrm{I}} = \mathbf{W}$. As a consequence, the demonstration III is reliable on a non empty set of worlds:

$$\mathbf{R}_{\mathrm{III}} = \mathbf{R}_{\mathrm{I}} \cap \mathbf{R}_{\mathrm{II}} = \mathbf{R}_{\mathrm{II}} \neq \varnothing$$

However, everyone agrees that the demonstration `III` is not valid in such a context. The idea is that the fact that rule `I` has been applied conditions the possible set of worlds. If the meta-knowledge information $\mathbf{A}_{\mathrm{I}} \cap \mathbf{R}_{\mathrm{II}} = \varnothing$ is added, then the support of the demonstration `III` is reduced to the empty set:

$$\mathbf{D}_{\mathrm{III}} = \mathbf{R}_{\mathrm{I}} \cap \mathbf{R}_{\mathrm{II}} \cap \mathbf{A}_{\mathrm{I}} \cap \mathbf{A}_{\mathrm{II}} = \varnothing$$

Every time the system receives a new event e, the system is conditioned by the context implied by e. This context $\mathbf{C}_e$ is a set of worlds. The global context $\mathbf{C}_{glob}$ is the context implied by all the events received by the system and by the fact that it is considered that the real world is not the impossible world, which implies the following development of $\mathbf{C}_{glob}$, where $\mathbf{D}_{\perp}$ is the set of worlds where $\perp$ has been demonstrated:

$$\mathbf{C}_{glob} = \overline{\mathbf{D}_{\perp}} \cap \bigcap_{e} \mathbf{C}_e$$

The interesting probability measure for the demonstrations is the measure μ_{glob} of their support conditioned to $\mathbf{C}_{glob}$:

$$\mu_{glob} : 2^{\mathbf{W}} \to [0,1]$$

$$\mu_{glob}(\mathbf{S}) = \frac{\mu(\mathbf{S} \cap \mathbf{C}_{glob})}{\mu(\mathbf{C}_{glob})}$$

As a conclusion to the previous example, the following equality stands: $\mu_{glob}(\mathbf{D}_{\mathrm{III}}) = 0$.

If the propositional framework is fit up with a subsumption relation $\preceq$, then a partial-order relation $\preceq_d$ can be constructed on demonstrations. A demonstration d_1 is subsumed by a second one d_2, if the proposition p_1 which it demonstrates is subsumed by the proposition p_2 that demonstrates d_2, if its reliability set $\mathbf{R}_1$ is included by the reliability set $\mathbf{R}_2$ of d_2 and if its applicability set $\mathbf{A}_1$ is included in the applicability set $\mathbf{A}_2$ of d_2:

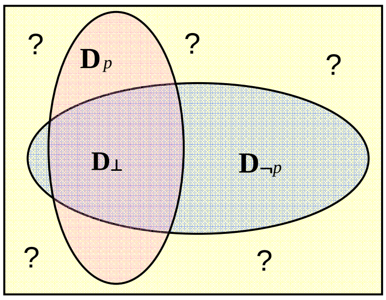
$$d_1 \preceq_d d_2 \Leftrightarrow \begin{cases} p_1 \preceq p_2 \\ \mathbf{R}_1 \subseteq \mathbf{R}_2 \\ \mathbf{A}_1 \subseteq \mathbf{A}_2 \end{cases}$$

This section provided the reader with the concepts and the tools for the computation of the contextual probability measure of demonstrations. These tools are required to give a numerical evaluation of the contextual probabilities of propositions. This is the subject of the next section.

4.2 Evaluation

At this point, the system is supposed to have a logically saturated base of demonstrations. However, the propositions truth cannot be directly evaluated by the system. As the demonstration base might be more or less contradictory, considering that the system might have received contradictory information, the system requires an evaluation process to measure the part of true and false and make its opinion about a proposition p.

The truth probability of p is not influenced by demonstrations that demonstrate a proposition that subsumes p or its contrary $\neg p$. The evaluation consists in making the contextual measure of the sets where p and $\neg p$ are respectively demonstrated. The contextual computation also involves the measure of the set where $\perp$ has been demonstrated.



	$\mathbf{D}_p$	$\overline{\mathbf{D}_p}$
$\mathbf{D}_{\neg p}$	$\perp$	$\neg p$
$\overline{\mathbf{D}_{\neg p}}$	p	???

Figure 5a and 5b: the 4-split of the set of worlds

The inferior probability of a proposition p is the contextual measure of sets where p has been demonstrated:

$$P_{\inf}(p) = \mu_{glob}(\mathbf{D}_p) = \frac{\mu(\mathbf{D}_p \cap \mathbf{C}_{glob})}{\mu(\mathbf{C}_{glob})}$$

The superior probability of a proposition p is the contextual measure of sets where $\neg p$ has not been demonstrated:

$$P_{\sup}(p) = \mu_{glob}\left(\overline{\mathbf{D}_{\neg p}}\right) = \frac{\mu\left(\overline{\mathbf{D}_{\neg p}} \cap \mathbf{C}_{glob}\right)}{\mu(\mathbf{C}_{glob})}$$

As $\overline{\mathbf{D}_{\neg p}} \cap \mathbf{D}_{\perp} \supseteq \mathbf{D}_p \cap \mathbf{D}_{\perp}$, $P_{\sup} \geq P_{\inf}$ stands. It can also be shown that $P_{\inf}$ and $P_{\sup}$ are dual:

$$P_{\inf}(p) + P_{\sup}(\neg p) = \frac{\mu(\mathbf{D}_p \cap \mathbf{C}_{glob}) + \mu(\overline{\mathbf{D}_p} \cap \mathbf{C}_{glob})}{\mu(\mathbf{C}_{glob})}$$
$$= \frac{\mu(\mathbf{C}_{glob})}{\mu(\mathbf{C}_{glob})} = 1$$

4.3 Simplifications

Unfortunately, there is not at disposal enough means to instantiate the different sets that are manipulated with the demonstrations, and even if there were, there would be big difficulties to compute the set intersections and unions. Usually, the best that can be done is to estimate the objective measure of the demonstration reliabilities: $\mu(\mathbf{R}_i) = r_i$, and declare some relations between the demonstration sets: $\mathbf{R}_i \cap \mathbf{A}_j = \varnothing$. In order to get the inferior and superior probabilities computable, the following two default relations are considered:

With no better information, if a global context $\mathbf{C}_{glob}$ implies the use of a demonstration d with applicability $\mathbf{A}_d$, then $\mathbf{C}_{glob} \subseteq \mathbf{A}_d$ stands. In other words, if the context implies the syntactic application of a rule, then, by default, the applicability is granted.

The atomic sets are the sets that are not compounded. Basically, the contexts implied by events and the hypotheses reliabilities and applicabilities are the atomic sets that are considered in this paper. With no better information, the atomic sets are considered independent between them, which corresponds to the maximum entropy. As a consequence, if $\mathbf{G}_k$ are atomic sets that have no specific relation between them, the following relation stands:

$$\mu\left(\bigcap_k \mathbf{G}_k\right) = \prod_k \mu(\mathbf{G}_k)$$

4.4 Evaluation Algorithm

The following fraction has to be calculated:

$$P_{\inf}(p) = \frac{\mu(\mathbf{D}_p \cap \mathbf{C}_{glob})}{\mu(\mathbf{C}_{glob})}$$

The goal of the algorithm is to evaluate the numerator and the denominator of the fraction.

The first stage of the algorithm is the development of the numerator and the denominator in atomic sets. The following well-known relations between sets are sufficient:

$$\mathbf{A} \cup \mathbf{B} = \overline{\overline{\mathbf{A}} \cap \overline{\mathbf{B}}}$$
$$\mu(\overline{\mathbf{A}} \cap \mathbf{B}) = \mu(\mathbf{B}) - \mu(\mathbf{A} \cap \mathbf{B})$$

So, by developing the formula, the following sum (or a difference) of measures is obtained, where $\cap_k$ denotes an intersection of atomic sets:

$$\sum_{\cap_k} \pm \mu(\cap_k)$$

The second stage deals with incompatibilities that show in the $\cap_k$ intersections. If an incompatibility is known as in the **4.1** section example, then:

$$\cap_k \subseteq \{w\bot\} \cap \mathbf{C}_{glob} = \varnothing$$
$$\Rightarrow \quad \mu(\cap_k) = 0$$

The third stage deals with remaining applicability sets, on which there is no information about the measure. The fact that $\mathbf{C}_{glob} \subseteq \mathbf{A}_d$ is used in order to remove them from every $\cap_k$ equation, as justified in the **4.3** section. The $\cap_k$ sets consist of intersections of reliabilities and contexts implied by events only:

$$\cap_k = \bigcap_i \mathbf{R}_i \cap \bigcap_e \mathbf{C}_e$$

At the fourth stage, as explained in the **4.3** section too, the sets that remain in every $\cap_k$ intersection are considered as independent. Then, the fraction reduction can be executed by removing the contextual factors from numerator and denominator. Indeed the context sets are present in every intersection $\cap_k$.

For the fifth stage, only atomic reliability measures remain and their measures are informed by the rules and now, the probability can be numerically evaluated.

4.5 Example

The **4.1** section example is carried and the system learns in addition that there is a problem on the line (automatic test) and that the box may be not connected (the user told he was not sure):

I	$\neg line \Rightarrow \neg internet$	$\mathbf{R}_{\mathrm{I}}$	$\mathbf{A}_{\mathrm{I}}$
II	$\neg internet \Rightarrow \neg cb$	$\mathbf{R}_{\mathrm{II}}$	$\mathbf{A}_{\mathrm{II}}$

III	$\neg line$	$\mathbf{R}_{\mathrm{III}}$	$\mathbf{A}_{\mathrm{III}}$
IV	$\neg cb$	$\mathbf{R}_{\mathrm{IV}}$	$\mathbf{A}_{\mathrm{IV}}$

Where cb denotes that the box is well connected.

The following additional knowledge still stands:

$$\mathbf{A}_{\mathrm{I}} \cap \mathbf{R}_{\mathrm{II}} = \varnothing$$

The following inferences are generated:

V = I + II	$\neg line \Rightarrow \neg cb$	$\mathbf{R}_{\mathrm{V}}$	$\mathbf{A}_{\mathrm{V}}$
VI = I + III	$\neg internet$	$\mathbf{R}_{\mathrm{VI}}$	$\mathbf{A}_{\mathrm{VI}}$
VII = VI + II	$\neg cb$	$\mathbf{R}_{\mathrm{VII}}$	$\mathbf{A}_{\mathrm{VII}}$
VIII = V + III	$\neg cb$	$\mathbf{R}_{\mathrm{VIII}}$	$\mathbf{A}_{\mathrm{VIII}}$

The illustration is made on the evaluation of the inferior and superior probabilities of cb.

There is no demonstration of cb. Therefore, the $P_{\inf}$ calculation is direct:

$$P_{\inf}(cb) = \frac{\mu(\mathbf{D}_{cb} \cap \mathbf{C}_{glob})}{\mu(\mathbf{C}_{glob})} = \frac{\mu(\{w_\perp\} \cap \mathbf{C}_{glob})}{\mu(\mathbf{C}_{glob})} = 0$$

The calculation of the superior probability is more interesting. There are three different demonstrations of $\neg cb$: IV, VII and VIII. Intuitively, the VII and VIII demonstrations are identical and not even valid because of the incompatibilities between $\mathbf{A}_{\mathrm{I}}$ and $\mathbf{R}_{\mathrm{II}}$. The section **4.4** algorithm is applied as follows.

At the first stage, the numerator and the denominator are expanded separately. They are respectively called num and den in the rest of the section. First $\mathbf{C}_{glob}$ and $\mathbf{D}_{\neg cb}$ are expanded:

$$\mathbf{C}_{glob} = \overline{\mathbf{D}_\perp} \cap \bigcap_e \mathbf{C}_e = \bigcap_e \mathbf{C}_e$$

$$\begin{aligned}
\mathbf{D}_{\neg cb} &= \bigcup_{d \in D(\neg cb)} \left(\bigcap_{h \in H(d)} (\mathbf{R}_h \cap \mathbf{A}_h) \right) \\
&= (\mathbf{R}_{\mathrm{IV}} \cap \mathbf{A}_{\mathrm{IV}}) \\
&\quad \cup (\mathbf{R}_{\mathrm{I}} \cap \mathbf{A}_{\mathrm{I}} \cap \mathbf{R}_{\mathrm{II}} \cap \mathbf{A}_{\mathrm{II}} \cap \mathbf{R}_{\mathrm{III}} \cap \mathbf{A}_{\mathrm{III}}) \\
&\quad \cup (\mathbf{R}_{\mathrm{I}} \cap \mathbf{A}_{\mathrm{I}} \cap \mathbf{R}_{\mathrm{II}} \cap \mathbf{A}_{\mathrm{II}} \cap \mathbf{R}_{\mathrm{III}} \cap \mathbf{A}_{\mathrm{III}}) \\
&= \overline{\overline{\mathbf{R}_{\mathrm{I}} \cap \mathbf{A}_{\mathrm{I}} \cap \mathbf{R}_{\mathrm{II}} \cap \mathbf{A}_{\mathrm{II}} \cap \mathbf{R}_{\mathrm{III}} \cap \mathbf{A}_{\mathrm{III}}} \cap \overline{\mathbf{R}_{\mathrm{IV}} \cap \mathbf{A}_{\mathrm{IV}}}}
\end{aligned}$$

Where $D(\neg cb)$ is the set of the demonstrations that demonstrate $\neg cb$ and $H(d)$ is the set of hypotheses used by the demonstration d. As a consequence, the following equations are obtained for num and den.

$$\begin{aligned}
num &= \mu\left(\overline{\mathbf{D}_{\neg cb}} \cap \bigcap_e \mathbf{C}_e \right) \\
&= \mu\left(\bigcap_i \mathbf{C}_i \right) - \mu\left(\mathbf{D}_{\neg cb} \cap \bigcap_i \mathbf{C}_i \right) \\
&= \mu\left(\bigcap_i \mathbf{C}_i \right) - \mu\left(\mathbf{R}_{\mathrm{IV}} \cap \mathbf{A}_{\mathrm{IV}} \cap \bigcap_i \mathbf{C}_i \right) \\
&\quad - \mu\left(\mathbf{R}_{\mathrm{I}} \cap \mathbf{A}_{\mathrm{I}} \cap \mathbf{R}_{\mathrm{II}} \cap \mathbf{A}_{\mathrm{II}} \cap \mathbf{R}_{\mathrm{III}} \cap \mathbf{A}_{\mathrm{III}} \cap \bigcap_i \mathbf{C}_i \right) \\
&\quad + \mu\left(\mathbf{R}_{\mathrm{I}} \cap \mathbf{A}_{\mathrm{I}} \cap \mathbf{R}_{\mathrm{II}} \cap \mathbf{A}_{\mathrm{II}} \cap \mathbf{R}_{\mathrm{III}} \cap \mathbf{A}_{\mathrm{III}} \cap \mathbf{R}_{\mathrm{IV}} \cap \mathbf{A}_{\mathrm{IV}} \cap \bigcap_i \mathbf{C}_i \right)
\end{aligned}$$

$$den = \mu\left(\bigcap_i \mathbf{C}_i \right)$$

The reader can notice that the VII and VIII demonstrations form already one, and that subsumption effort can be made either into the logics or during evaluation.

The second stage deals with incompatibilities. The only one is: $\mathbf{A}_{\mathrm{I}} \cap \mathbf{R}_{\mathrm{II}} = \varnothing$:

$$num = \mu\left(\bigcap_i \mathbf{C}_i \right) - \mu\left(\mathbf{R}_{\mathrm{IV}} \cap \mathbf{A}_{\mathrm{IV}} \cap \bigcap_i \mathbf{C}_i \right)$$

The third stage removes the remaining applicabilities from the equations:

$$num = \mu\left(\bigcap_i \mathbf{C}_i \right) - \mu\left(\mathbf{R}_{\mathrm{IV}} \cap \bigcap_i \mathbf{C}_i \right)$$

The fourth stage uses the independence assumption to get:

$$num = \prod_i \mu(\mathbf{C}_i) - \mu(\mathbf{R}_{\mathrm{IV}}) \times \prod_i \mu(\mathbf{C}_i)$$

$$den = \prod_i \mu(\mathbf{C}_i)$$

Eventually the fraction can be simplified by removing the context factor and it provides the very simple equation for the superior probability:

$$P_{\sup}(cb) = 1 - \mu(\mathbf{R}_{\mathrm{IV}})$$

6 Comparison with State-of-the-Art

6.1 The Evidence Theory

The Evidence Theory [7], [8] does not model any equivalence to the applicability concept. It is the reason why Pearl [5] addressed the criticism about the lack of handling of default rules. Apart from the applicability, there is equivalence with the LFPR.

The demonstration concept of the LFPR is nothing else than a mass assignment where the set $\mathbf{E}$ corresponds to the support in the set of worlds $\mathbf{W}$ of the demonstration proposition p, and where the mass $m(\mathbf{E})$ corresponds to the demonstration reliability $\mathbf{R}$. Analogously, any mass assignment m can be emulated into the LFPR with the following procedure: for every set $\mathbf{E} \in 2^{\mathbf{W}}$ such that $m(\mathbf{E}) \neq 0$, a demonstration $d_{\mathbf{E}}$ has to be created, such that the proposition $p_{\mathbf{E}}$ it demonstrates is the logical description of $\mathbf{E}$, such that its reliability $\mathbf{R}_{\mathbf{E}}$ has a measure equal to $m(\mathbf{E})$ and is separated[1] from any other demonstration reliability that has been generated from the mass assignment m and independent from any other demonstration reliability that has been generated from another mass assignment, and such that its applicability $\mathbf{A}_{\mathbf{E}}$ is equal to $\mathbf{W}$.

The Evidence Theory approach is focused on the mass assignment although the LFPR approach focuses on the logics and the inference. Thus, when LFPR measure μ is constant and does not depend on the observations[2], the Evidence Theory mass assignment is updated after each new observation and the belief functions correspond to the contextual measure μ_{glob}.

6.2 The Theory of Hints

The main idea of de Kleer while inventing the ATMS [9], [10], [11] was to enable the reasoning under assumptions. The ATMS assumptions respectively correspond to the LFPR hypotheses. The role of the ATMS is to compute the set of assumptions that are necessary for the derivation of a given node that is labelled with a proposition. The set of assumptions is actually a union of intersections of assumptions. The union part shows that the node proposition can be inferred from different justifications and the intersection part comes from the combination of assumptions necessary for one justification. There is a lot of similarity in both approaches but this paper studies more precisely the analogy with the ATMS inspired Theory of Hints of Kohlas and Monney

The purpose of the Theory of Hints [12] is to merge the Evidence Theory with ATMS. The main idea was to include the uncertainty through the anomaly concept within the logics. The example of section **4.5** becomes in the Theory of Hints:

I	$\neg line \Rightarrow \neg internet$
II	$\neg internet \Rightarrow \neg cb \vee a_1$
III	$\neg line \vee a_2$
IV	$\neg cb \vee a_3$
V	$\neg line \Rightarrow a_1$

Rule V basically says: $\neg line$ is an exception to rule II. The generated rules are the following ones:

VI = I + II	$\neg line \Rightarrow \neg cb \vee a_1$
VII = VI + III	$\neg cb \vee a_1 \vee a_2$
VIII = III + V	$a_1 \vee a_2$
IX = I + III	$\neg internet \vee a_2$
X = IX + II	$\neg cb \vee a_1 \vee a_2$

Eventually VIII subsumes VII and X. The role of the applicability is the same as the rule V: to model the context implied by the use of the rule I. When the LFPR constraints the context with meta-knowledge, the Theory of Hints does it into the logic, and both frameworks finally come to the same results. The difference between the two theories mainly concerns the approach and the complexity: in the Theory of Hints, the logics are overloaded with anomalies although in the LFPR, one has just to remember the hypotheses that have been done to make every assumption. The complexity on request is greater in the LFPR because the dependencies calculations have not been done during the logical inferences contrary to the Theory of Hints. The interest of the LFPR approach is to compute

[1]. $\mathbf{R}_{\mathbf{E}} \cap \mathbf{R}_{\mathbf{F}} = \varnothing$

[2]. The observations provide the system with measures $\mu(\mathbf{R})$. This is information about the measure, not an evolution of the measure

the complex dependency only when requested. There is almost no raise of complexity between a classical logical inference engine and the LFPR engine during the demonstration generation. The only raise comes from the subsumption tests that are more complex in LFPR than in a basic engine with no uncertainty capabilities.

7 Conclusion

This article presented a new framework, the Logical Framework for Probabilistic Reasoning (LFPR). This framework is based on logics, but it can be considered as an extension of the Evidence Theory, and it obtains the comparable results with the Theory of Hints. We motivated the choice and the interest of this theory in the particular application field: Dialogue Systems.

In the future, we plan to add learning ability to the LFPR, by designing a back propagation of rewards through the inference graph.

Acknowledgements

Special thanks to Irina Zavyalova and Ghislain Putois for their help on this paper.

This work is partially funded by the EC project: "CLASSiC" (FP7-ICT-216594),
see www.classic-project.org

References

[1] O. Lemon, O. Pietquin (2007). Machine Learning for Spoken Dialogue Systems. In *Proceedings of the conference Interspeech 2007*, pages 2685-2688, Antwerp, Belgium, September 2007.

[2] J.D. Williams, S.J. Young (2007). Partially Observable Markov Decision Processes for Spoken Dialogue Systems. In *Computer Speech and Langage*, vol. 21, no. 2, pages 393-422, 2007.

[3] X. Li, J.M. Huerta (2007). How Predicatable is ASR Confidence in Dialogue Applications? In *Proceedings of the conference Interspeech 2007*, pages 1745-1748, Antwerp, Belgium, September 2007.

[4] S. Huet, G. Gravier, P. Sébillot (2007). Morphosyntactic Processing of N-Best Lists for Improved Recognition and Confidence Mesure Computation. In *Proceedings of the conference Interspeech 2007*, pages 1741-1744, Antwerp, Belgium, September 2007.

[5] J. Pearl (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. *Morgan Kaufmann Publishers*, 1988.

[6] R.S. Sutton, A.G. Barto (1998). Reinforcement Learning: An Introduction. *MIT Press*, 1998.

[7] G. Shafer (1976). A Mathematical Theory of Evidence. *Princeton University Press*, 1976.

[8] A.P. Dempster (1968). A generalization of Bayesian inference. In *Journal of the Royal Statistical Society*, Series B, Vol. 30, pages 205-247, 1968.

[9] J. de Kleer (1986). An Assumption-based TMS. In *Artificial Intelligence*, Vol. 28, pages 127-162, 1986.

[10] J. de Kleer (1986). Extending the ATMS. In *Artificial Intelligence*, Vol. 28, pages 163-196, 1986.

[11] J. de Kleer (1986). Problem Solving with the ATMS. In *Artificial Intelligence*, Vol. 28, pages 197-224, 1986.

[12] J. Kohlas, P.A. Monney (1995). A mathematical theory of hints: an approach to the Dempster-Shafer theory of evidence. In *Lecture Notes in Economics and Mathematical Systems*, vol. 425, 1995.