A Probabilistic Abduction Engine for Media Interpretation based on Ontologies

Oliver Gries, Ralf Möller, Anahita Nafissi, Maurice Rosenfeld, Kamil Sokolski, Michael Wessel

Hamburg University of Technology, Germany

Abstract. For multimedia interpretation, and in particular for the combined interpretation of information coming from different modalities, a semantically well-founded formalization is required in the context of an agent-based scenario. Low-level percepts, which are represented symbolically, define the observations of an agent, and interpretations of content are defined as explanations for the observations. We propose an abduction-based formalism that uses description logics for the ontology and Horn rules for defining the space of hypotheses for explanations (i.e., the space of possible interpretations of media content), and we use Markov logic to define the motivation for the agent to generate explanations on the one hand, and for ranking different explanations on the other.[1]

1 Introduction

For multimedia interpretation in the context of an agent-based scenario, and for the combined interpretation of information coming from different modalities in particular, a semantically well-founded formalization is required. Low-level percepts, which are represented symbolically, define the observations of an agent w.r.t. some content, and interpretations of the content are defined as explanations for the observations. In [Castano et al., 2008] we have proposed an abduction-based formalism that uses description logics for the ontology and Horn rules for defining the space of hypotheses for explanations (i.e., the space of possible interpretations of media content). An evaluation of the abduction approach based on description logics and rules is presented in [Espinosa-Peraldi et al., 2010b]. A discussion of related work can be found in [Espinosa-Peraldi et al., 2010a].

In this paper, we propose the use of Markov logic to define the motivation for the agent to generate explanations on the one hand, and for ranking different explanations on the other. Furthermore, we discuss completely how the reasoning process is performed with uncertainty and under inconsistency in the input data. In this paper, we introduce a new approach for ranking interpretation Aboxes. The ranking process is performed based on a probabilistic scoring function (as opposed to the proof-theoretic scoring function used in [Espinosa-Peraldi et al., 2010b]). A termination condition is also defined which determines how long the interpretation process should be performed.

On the one hand, the approach presented in this paper can be used to formalize media interpretation whereas, in the long run, the approach can be seen as a theory for agent behavior for interpreting observations (which are inherently uncertain). We use a probabilistic logic to motivate the media interpretation strategy (or the agent's explanation endeavor) by increasing the belief in the observations. Hence, we define a semantically well-founded utility measure to justify the computational resources spent for generating interpretations. In this paper we focus on the media interpretation scenario.

Based on a presentation of the most important preliminaries in Section 2, the abduction and interpretation procedures are discussed in detail in Section 3. Optimization techniques for the probabilistic abduction engine are pointed out. In Section 4, a complete example is given showing the main approach using intermediate steps. In Section 6 an agenda is described which applies some techniques to improve the performance of the interpretation process. Section 7 summarizes this paper.

[1] This work has been funded by the European Community with the project CASAM (Contract FP7-217061 CASAM) and by the German Science Foundation with the project PRESINT (DFG MO 801/1-1).

2 Preliminaries

In this section, the most important preliminaries are specified in order to make this document self-contained.

2.1 Preliminaries on Description Logics

For specifying the ontology used to describe low-level analysis results as well as high-level interpretation results, a less expressive description logic is applied to facilitate fast computations. We decided to represent the domain knowledge with the DL $\mathcal{ALH}_f{}^-$ (restricted attributive concept language with role hierarchies, functional roles and concrete domains). The motivation to only allow a restricted use of existential restrictions is to support a well-founded integration of the description logic part of the knowledge base with the probabilistic part (based on Markov logic networks, see Section 2.4).

In logic-based approaches, atomic representation units have to be specified. The atomic representation units are fixed using a so-called signature. A DL *signature* is a tuple $\mathcal{S} = (\mathbf{CN}, \mathbf{RN}, \mathbf{IN})$, where $\mathbf{CN} = \{A_1, ..., A_n\}$ is the set of concept names (denoting sets of domain objects) and $\mathbf{RN} = \{R_1, ..., R_m\}$ is the set of role names (denoting relations between domain objects). The signature also contains a component $\mathbf{IN}$ indicating a set of individuals (names for domain objects).

In order to relate concept names and role names to each other (terminological knowledge) and to talk about specific individuals (assertional knowledge), a knowledge base has to be specified. An $\mathcal{ALH}_f{}^-$ *knowledge base* $\Sigma_{\mathcal{S}} = (\mathcal{T}, \mathcal{A})$, defined with respect to a signature $\mathcal{S}$, is comprised of a terminological component $\mathcal{T}$ (called *Tbox*) and an assertional component $\mathcal{A}$ (called *Abox*). In the following we just write Σ if the signature is clear from context. A Tbox is a set of so-called *axioms*, which are restricted to the following form in $\mathcal{ALH}_f{}^-$:

(I)	Subsumption	$A_1 \sqsubseteq A_2,\ R_1 \sqsubseteq R_2$
(II)	Disjointness	$A_1 \sqsubseteq \neg A_2$
(III)	Domain and range restrictions for roles	$\exists R.\top \sqsubseteq A,\ \top \sqsubseteq \forall R.A$
(IV)	Functional restriction on roles	$\top \sqsubseteq (\leq 1\, R)$
(V)	Local range restrictions for roles	$A_1 \sqsubseteq \forall R.A_2$
(VI)	Definitions with value restrictions	$A \equiv A_0 \sqcap \forall R_1.A_1 \sqcap ... \sqcap \forall R_n.A_n$

With axioms of form (I), concept (role) names can be declared to be subconcepts (subroles) of each other. Axioms of form (II) denote disjointness between concepts. Axioms of type (III) introduce domain and range restrictions for roles. Axioms of the form (IV) introduce so-called *functional* restrictions on roles, and axioms of type (V) specify local range restrictions (using value restrictions, see below). With axioms of kind (VI) so-called definitions (with necessary and sufficient conditions) can be specified for concept names found on the left-hand side of the $\equiv$ sign. In the axioms, so-called *concepts* are used. Concepts are concept names or expressions of the form $\top$ (anything), $\bot$ (nothing), $\neg A$ (atomic negation), $(\leq 1\, R)$ (role functionality), $\exists R.\top$ (limited existential restriction), $\forall R.A$ (value restriction) and $(C_1 \sqcap ... \sqcap C_n)$ (concept conjunction).

Knowledge about individuals is represented in the Abox part of Σ. An Abox $\mathcal{A}$ is a set of expressions of the form $A(a)$ or $R(a, b)$ (concept assertions and role assertions, respectively) where A stands for a concept name, R stands for a role name, and a, b stand for individuals. Aboxes can also contain equality $(a = b)$ and inequality assertions $(a \neq b)$. We say that the unique name assumption (UNA) is applied, if $a \neq b$ is added for all pairs of individuals a and b.

In order to understand the notion of logical entailment, we introduce the semantics of $\mathcal{ALH}_f{}^-$. In DLs such as $\mathcal{ALH}_f{}^-$, the semantics is defined with interpretations $\mathcal{I} = (\triangle^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\triangle^{\mathcal{I}}$ is a non-empty set of domain objects (called the domain of $\mathcal{I}$) and $\cdot^{\mathcal{I}}$ is an interpretation function which maps individuals to objects of the domain ($a^{\mathcal{I}} \in \triangle^{\mathcal{I}}$), atomic concepts to subsets of the domain ($A^{\mathcal{I}} \subseteq \triangle^{\mathcal{I}}$) and roles to subsets of the cartesian product of the domain ($R^{\mathcal{I}} \subseteq \triangle^{\mathcal{I}} \times \triangle^{\mathcal{I}}$). The interpretation of arbitrary $\mathcal{ALH}_f{}^-$ concepts is then defined by extending $\cdot^{\mathcal{I}}$ to all $\mathcal{ALH}_f{}^-$ concept constructors:

$$
\begin{array}{ll}
\top^{\mathcal{I}} & = \triangle^{\mathcal{I}} \\
\bot^{\mathcal{I}} & = \emptyset \\
(\neg A)^{\mathcal{I}} & = \triangle^{\mathcal{I}} \setminus A^{\mathcal{I}} \\
(\leq 1\, R)^{\mathcal{I}} & = \{u \in \triangle^{\mathcal{I}} \mid (\forall v_1, v_2)\, [((u, v_1) \in R^{\mathcal{I}} \wedge (u, v_2) \in R^{\mathcal{I}}) \rightarrow v_1 = v_2] \\
(\exists R.\top)^{\mathcal{I}} & = \{u \in \triangle^{\mathcal{I}} \mid (\exists v)\, [(u, v) \in R^{\mathcal{I}}]\} \\
(\forall R.C)^{\mathcal{I}} & = \{u \in \triangle^{\mathcal{I}} \mid (\forall v)\, [(u, v) \in R^{\mathcal{I}} \rightarrow v \in C^{\mathcal{I}}]\} \\
(C_1 \sqcap ... \sqcap C_n)^{\mathcal{I}} & = C_1^{\mathcal{I}} \cap ... \cap C_n^{\mathcal{I}}
\end{array}
$$

In the following, the *satisfiability* condition for axioms and assertions of an $\mathcal{ALH}_f{}^-$-knowledge base Σ in an interpretation $\mathcal{I}$ are defined. A concept inclusion $C \sqsubseteq D$ (concept definition $C \equiv D$) is satisfied in $\mathcal{I}$, if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ (resp. $C^{\mathcal{I}} = D^{\mathcal{I}}$) and a role inclusion $R \sqsubseteq S$ (role definition $R \equiv S$), if $R^{\mathcal{I}} \subseteq S^{\mathcal{I}}$ (resp. $R^{\mathcal{I}} = S^{\mathcal{I}}$). Similarly, assertions $C(a)$ and $R(a, b)$ are satisfied in $\mathcal{I}$, if $a^{\mathcal{I}} \in C^{\mathcal{I}}$ resp. $(a, b)^{\mathcal{I}} \in R^{\mathcal{I}}$. If an interpretation $\mathcal{I}$ satisfies all axioms of $\mathcal{T}$ resp. $\mathcal{A}$ it is called a *model* of $\mathcal{T}$ resp. $\mathcal{A}$. If it satisfies both $\mathcal{T}$ and $\mathcal{A}$ it is called a model of Σ. Finally, if there is a model of Σ (i.e., a model for $\mathcal{T}$ and $\mathcal{A}$), then Σ is called satisfiable. We are now able to define the entailment relation $\models$. A DL knowledge base Σ *logically entails* an assertion α (symbolically $\Sigma \models \alpha$) if α is satisfied in all models of Σ. For an Abox $\mathcal{A}$, we say $\Sigma \models \mathcal{A}$ if $\Sigma \models \alpha$ for all $\alpha \in \mathcal{A}$.

2.2 Substitutions, Queries, and Rules

Sequences, Variable Substitutions and Transformations A *variable* is a name of the form *String* where *String* is a string of characters from {A...Z}. In the following definitions, we denote places where variables can appear with uppercase letters.

Let V be a set of variables, and let $\underline{X}, \underline{Y_1}, \ldots, \underline{Y_n}$ be sequences $\langle \ldots \rangle$ of variables from V. The notation $\underline{z}$ denotes a sequence of individuals. We consider sequences of length 1 or 2 only, if not indicated otherwise, and assume that $(\langle X \rangle)$ is to be read as (X) and $(\langle X, Y \rangle)$ is to be read as (X, Y) etc. Furthermore, we assume that sequences are automatically flattened. A function *as_set* turns a sequence into a set in the obvious way.

A *variable substitution* $\sigma = [X \leftarrow i, Y \leftarrow j, \ldots]$ is a mapping from variables to individuals mentioned in an Abox. The application of a variable substitution σ to a sequence of variables $\langle X \rangle$ or $\langle X, Y \rangle$ is defined as $\langle \sigma(X) \rangle$ or $\langle \sigma(X), \sigma(Y) \rangle$, respectively, with $\sigma(X) = i$ and $\sigma(Y) = j$. In this case, a sequence of individuals is defined. If a substitution is applied to a variable X for which there exists no mapping $X \leftarrow k$ in σ then the result is undefined. A variable for which all required mappings are defined is called *admissible* (w.r.t. the context).

Grounded Conjunctive Queries Let $\underline{X}, \underline{Y_1}, \ldots, \underline{Y_n}$ be sequences of variables, and let $Q_1, \ldots, Q_n$ denote concept or role names. A query is defined by the following syntax: $\{(\underline{X}) \mid Q_1(\underline{Y_1}), \ldots, Q_n(\underline{Y_n})\}$. The sequence $\underline{X}$ may be of arbitrary length but all variables mentioned in $\underline{X}$ must also appear in at least one of the $\underline{Y_1}, \cdots, \underline{Y_n}$: $as\_set(\underline{X}) \subseteq as\_set(\underline{Y_1}) \cup \cdots \cup as\_set(\underline{Y_n})$.

Informally speaking, $\underline{Q_1(\underline{Y_1}), \ldots, Q_n(\underline{Y_n})}$ defines a conjunction of so-called *query atoms* $Q_i(\underline{Y_i})$. The list of variables to the left of the sign $\mid$ is called the *head* and the atoms to the right are called the query *body*. The variables in the head are called distinguished variables. They define the query result. The variables that appear only in the body are called non-distinguished variables and are existentially quantified. Answering a query with respect to a knowledge base Σ means finding admissible variable substitutions σ such that $\Sigma \models \{\sigma(Q_1(\underline{Y_1})), \ldots, \sigma(Q_n(\underline{Y_n}))\}$. We say that a variable substitution $\sigma = [X \leftarrow i, Y \leftarrow j, \ldots]$ introduces *bindings* $i, j, \ldots$ for variables $X, Y, \ldots$. Given all possible variable substitutions σ, the *result* of a query is defined as $\{(\sigma(\underline{X}))\}$. Note that the variable substitution σ is applied before checking whether $\Sigma \models \{Q_1(\sigma(\underline{Y_1})), \ldots, Q_n(\sigma(\underline{Y_n}))\}$, i.e., the query is *grounded* first.

A *boolean* query is a query with $\underline{X}$ being of length zero. If for a boolean query there exists a variable substitution σ such that $\Sigma \models \{\sigma(Q_1(\underline{Y_1})), \ldots, \sigma(Q_n(\underline{Y_n}))\}$ holds, we say that the query is answered with *true*, otherwise the answer is *false*. Later on, we will have to convert query atoms into Abox assertions. This is done with the function *transform*. The function *transform* applied to a set of query atoms $\{\gamma_1, \dots \gamma_n\}$ is defined as $\{transform(\gamma_1, \sigma), \dots, transform(\gamma_n, \sigma)\}$ where $transform(P(\underline{X}), \sigma) := P(\sigma(\underline{X}))$.

Rules A rule r has the following form $P(\underline{X}) \leftarrow Q_1(\underline{Y_1}), \dots, Q_n(\underline{Y_n})$ where P, $Q_1, \dots, Q_n$ denote concept or role names with the additional restriction (safety condition) that $as\_set(\underline{X}) \subseteq as\_set(\underline{Y_1}) \cup \dots \cup as\_set(\underline{Y_n})$. Rules are used to derive new Abox assertions, and we say that a rule r is *applied* to an Abox $\mathcal{A}$. The function call $apply(\Sigma, P(\underline{X}) \leftarrow Q_1(\underline{Y_1}), \dots, Q_n(\underline{Y_n}), \mathcal{A})$ returns a set of Abox assertions $\{\sigma(P(\underline{X}))\}$ if there exists an admissible variable substitution σ such that the answer to the query

$$\{() \mid Q_1(\sigma(\underline{Y_1})), \dots, Q_n(\sigma(\underline{Y_n}))\}$$

is *true* with respect to $\Sigma \cup \mathcal{A}$.[2] If no such σ can be found, the result of the call to $apply(\Sigma, r, \mathcal{A})$ is the empty set. The application of a set of rules $\mathcal{R} = \{r_1, \dots r_n\}$ to an Abox is defined as follows:

$$apply(\Sigma, \mathcal{R}, \mathcal{A}) = \bigcup_{r \in \mathcal{R}} apply(\Sigma, r, \mathcal{A})$$

The result of $forward\_chain(\Sigma, \mathcal{R}, \mathcal{A})$ is defined to be $\emptyset$ if $apply(\Sigma, \mathcal{R}, \mathcal{A}) \cup \mathcal{A} = \mathcal{A}$ holds. Otherwise the result of *forward_chain* is determined by the recursive call
$apply(\Sigma, \mathcal{R}, \mathcal{A}) \cup forward\_chain(\Sigma, \mathcal{R}, \mathcal{A} \cup apply(\Sigma, \mathcal{R}, \mathcal{A}))$.

For some set of rules $\mathcal{R}$ we extend the entailment relation by specifying that $(\mathcal{T}, \mathcal{A}) \models_{\mathcal{R}} \mathcal{A}_0$ iff $(\mathcal{T}, \mathcal{A} \cup forward\_chain((\mathcal{T}, \emptyset), \mathcal{R}, \mathcal{A})) \models \mathcal{A}_0$.

2.3 Probabilistic Knowledge Representation

The basic notion of probabilistic knowledge representation formalisms is the so-called *random experiment*. A *random variable* X is a function assigning a value to the result of a random experiment. The random experiment itself is not represented, so random variables are functions without arguments, which return different values at different points of time. Possible values of a random variable comprise the so-called *domain* of the random variable. In the sequel, we will use *boolean* random variables, whose values can be either 1 or 0 ($true$ or $false$, respectively). Let $\vec{X} = \{X_1, ..., X_n\}$ be the ordered set of all random variables of a random experiment. An *event* (denoted $\vec{X} = \vec{x}$) is an assignment $X_1 = x_1, ..., X_n = x_n$ to all random variables. In case $n = 1$ we call the event simple, otherwise the event is called complex. A certain vector of values $\vec{x}$ is referred to as a *possible world*. A possible world can be associated with a *probability value* or *probability* for short. Hence, the notion of a possible world can be used as a synonym for an event, and depending on the context we use the former or the latter name. In case of an event with a boolean random variable X, we write x as an abbreviation for $X = true$ and $\neg x$ as an abbreviation for $X = false$. Mappings of events to probabilities (or assignment of probabilities to events) are specified with so-called *probability assertions* of the following syntax: $P(\vec{X} = \vec{x}) = p$, where $\vec{X}$ is a vector of random variables, and p is a real value between 0 and 1 (it is assumed that the reader is familiar with Kolmogorov's axioms of probability). In the special case of a simple event (single random variable, $n = 1$) we write $P(X = x) = p$. The probability value p of an event is denoted as $P(\vec{X} = \vec{x})$ (or $P(X = x)$ in the simple case). In its raw form a set of probabilistic assertions is called a *probabilistic knowledge base* (with signature $\vec{X}$). A mapping from the domain of a random variable X to probability values $[0, 1]$ is called a *distribution*. For distributions we use the notation $\mathbf{P}(X)$. Distributions can be defined for (ordered) sets of random variables as well. In this case we use $\mathbf{P}(X_1, \dots, X_n)$ as a denotation for a mapping to the n-dimensional cross product of $[0, 1]$. For specifying a distribution, probability assertions for *all* domain values must be specified, and the values p must sum up to 1. In case all random variables of a random experiment are involved, we speak of a

[2] We slightly misuse notation in assuming $(\mathcal{T}, \mathcal{A}) \cup \Delta = (\mathcal{T}, \mathcal{A} \cup \Delta)$. If $\Sigma \cup \mathcal{A}$ is inconsistent the result is well-defined but useless. It will not be used afterwards.

full joint probability distribution (JPD), otherwise the expression is said to denote a *joint distribution* or a *marginal distribution* (projection of the n-dimensional space of probability values to a lower-dimensional space with m dimensions). The expression $\mathbf{P}(X_1, \ldots, X_m, X_{m+1} = x_{m+1}, \ldots, X_l = x_l)$ denotes an m-dimensional distribution with known values $x_{m+1}, \ldots, x_l$. In slight misuse of notation, we sometimes write $\vec{e}$ for these known values (e stands for evidence). The fragment $\vec{e}$ need not necessarily be written at the end in the parameter list of $\mathbf{P}$. A *conditional probability* for a set of random variables $X_1, ..., X_m$ is denoted with $P(X_1 = x_1, ..., X_m = x_m \mid \vec{e})$ or, in distribution form, we write $\mathbf{P}(X_1, ..., X_m \mid \vec{e})$ (conditional probability distribution). This distribution can also be written as $\frac{\mathbf{P}(\vec{X},\vec{e})}{\mathbf{P}(\vec{e})}$. For a probabilistic knowledge base, formal inference problems are defined. We restrict our attention to the two most convenient probabilistic inference problems: A *conditional probability query* is the computation of the joint distribution of a set of m random variables conditioned on $\vec{e}$ and is denoted with $P_X(x_1 \wedge ... \wedge x_m \mid \vec{e})$ where $vars(x_1, \ldots, x_m) \cap vars(\vec{e}) = \emptyset$ and $vars(x_1, \ldots, x_m) \cup vars(\vec{e}) \subseteq X$ with $vars$ specified in the obvious way. Note that x_i indicates $X_i = x_i$. In the following we have the distribution form of the above query: $\mathbf{P}_X(X_1, ..., X_m \mid \vec{e})$. If the set of random variables X is known from the context, the subscript $_X$ is often omitted. The Maximum A Posteriori (MAP) inference returns the most-likely state of query atoms given the evidence. Based on the MAP inference, the "most probable world" given the evidence is determined as a set of events. The MAP inference problem given a distribution for a set of random variables X is formalized as follows:

$$MAP_X(\vec{e}) := \vec{e} \cup argmax_{\vec{x}} P(\vec{x}|\vec{e}) \tag{1}$$

where $vars(\vec{x}) \cap vars(\vec{e}) = \emptyset$ and $vars(\vec{x}) \cup vars(\vec{e}) = X$. For both inference problems, conditional probability queries as well as the MAP problem, different kinds of algorithms exist, which possibly exploit additional assertions (such as, e.g., conditional independence assumptions in so-called Bayesian networks, or factored probability distribution specifications as in so-called Markov networks). In the next subsection, we focus on the latter formalism.

2.4 Markov Logic

The formalism of Markov logic [Domingos and Richardson, 2007] provides a means to combine the expressivity of first-order logic augmented with the formalism of Markov networks [Pearl, 1988]. The Markov logic formalism uses first-order logic to define "templates" for constructing Markov networks. The basic notion for this is called a Markov logic network.

A Markov logic network $MLN = (\mathcal{F}_{MLN}, \mathcal{W}_{MLN})$ consists of a sequence of first-order formulas $\mathcal{F}_{MLN} = \langle F_1, ..., F_m \rangle$ and a sequence of real number weights $\mathcal{W}_{MLN} = \langle w_1, ..., w_m \rangle$. The association of a formula to its weight is by position in the sequence. For a formula $F \in \mathcal{F}_{MLN}$ with associated weight $w \in \mathcal{W}_{MLN}$ we also write $w\, F$ (weighted formula). Thus, a Markov logic network can also be defined as a set of weighted formulas. Both views can be used interchangeably. As a notational convenience, for ordered sets we nevertheless sometimes write $\vec{X}, \vec{Y}$ instead of $\vec{X} \cup \vec{Y}$.

In contrast to standard first-order logics such as predicate logic, relational structures not satisfying a formula F_i are not ruled out as models. If a relational structure does not satisfy a formula associated with a large weight it is just considered to be quite unlikely the "right" one.

Let $C = \{c_1, ..., c_m\}$ be the set of all constants mentioned in $\mathcal{F}_{MLN}$. A *grounding* of a formula $F_i \in \mathcal{F}_{MLN}$ is a substitution of all variables in the matrix of F_i with constants from C. From all groundings, the (finite) set of grounded atomic formulas (also referred to as *ground atoms*) can be obtained. Grounding corresponds to a domain closure assumption. The motivation is to get rid of the quantifiers and reduce inference problems to the propositional case.

Since a ground atom can either be true or false in an interpretation (or world), it can be considered as a boolean random variable X. Consequently, for each MLN with associated random variables $\vec{X}$, there is a set of possible worlds $\vec{x}$. In this view, sets of ground atoms are sometimes used to denote worlds. In this context, negated ground atoms correspond to *false* and non-negated ones to *true*. We denote worlds using a sequence of (possibly negated) atoms.

When a world $\vec{x}$ violates a weighted formula (does not satisfy the formula) the idea is to ensure that this world is less probable rather than impossible as in predicate logic. Note that weights do not directly correspond to probabilities (see [Domingos and Richardson, 2007] for details).

For each possible world of a Markov logic network $MLN = (\mathcal{F}_{MLN}, \mathcal{W}_{MLN})$ there is a probability for its occurrence. Probabilistic knowledge is required to obtain this value. As usual, probabilistic knowledge is specified using a probability distribution. In the formalism of Markov networks the full joint probability distribution of a Markov logic network MLN is specified in symbolic form as $\mathbf{P}_{MLN}(\vec{X}) = (P(\vec{X} = \vec{x}_1), \ldots, P(\vec{X} = \vec{x}_n))$, for every possible $\vec{x}_i \in \{true, false\}^n$, $n = |\vec{X}|$ and $P(\vec{X} = \vec{x}) := log\_lin_{MLN}(\vec{x})$ (for a motivation of the log-linear form, see, e.g., [Domingos and Richardson, 2007]), with log_lin being defined as

$$log\_lin_{MLN}(\vec{x}) = \frac{1}{Z}\, exp\,(\sum_{i=1}^{|\mathcal{F}_{MLN}|} w_i n_i(\vec{x}))$$

According to this definition, the probability of a possible world $\vec{x}$ is determined by the exponential of the sum of the number of true groundings (n_i) formulas $F_i \in \mathcal{F}_{MLN}$ in $\vec{x}$, multiplied with their corresponding weights $w_i \in \mathcal{W}_{MLN}$, and finally normalized with

$$Z = \sum_{\vec{x} \in \vec{X}} \exp\,(\sum_{i=1}^{|\mathcal{F}_{MLN}|} w_i n_i(\vec{x})), \tag{2}$$

the sum of the probabilities of all possible worlds. Thus, rather than specifying the full joint distribution directly in symbolic form as we have discussed before, in the Markov logic formalism, the probabilistic knowledge is specified implicitly by the weights associated with formulas. Determining these formulas and their weights in a practical context is all but obvious, such that machine learning techniques are usually employed for knowledge acquisition.

A *conditional probability query for a Markov logic network MLN* is the computation of the joint distribution of a set of m events involving random variables conditioned on $\vec{e}$ and is denoted with

$$P_{MLN}(x_1 \wedge \ldots \wedge x_m \mid \vec{e})$$

The semantics of this query is given as:

$$P_{rand\_vars(MLN)}(x_1 \wedge \ldots \wedge x_m \mid \vec{e}) \text{ } w.r.t. \text{ } \mathbf{P}_{MLN}(rand\_vars(MLN))$$

where $vars(x_1, \ldots, x_m) \cap vars(\vec{e}) = \emptyset$ and $vars(x_1, \ldots, x_m) \subseteq rand\_vars(MLN)$. The function $rand\_vars$ is defined as follows: $rand\_vars((\mathcal{F}, \mathcal{W})) := \{A(\underline{C}) \mid A(\underline{C})$ is mentioned in some grounded formula $F \in \mathcal{F}\}$. Grounding is accomplished w.r.t. all constants that appear in $\mathcal{F}$ where A denotes atomic concept or atomic role. An algorithm for answering queries of the above form is investigated in [Gries and Möller, 2010]. In the case of Markov logic, the definition of the MAP problem given in (1) can be rewritten as follows. The conditional probability term $P(\vec{x}|\vec{e})$ is replaced with the Markovian formula:

$$MAP_{MLN}(\vec{e}) := \vec{e} \cup argmax_{\vec{x}} \frac{1}{Z_e} \exp\left(\sum_i w_i n_i\,(\vec{x}, \vec{e})\right) \tag{3}$$

Thus, for describing the most-probable world, MAP returns a set of events, one for each random variable used in the Markov network derived from MLN. In the above equation, $\vec{x}$ denotes the hidden variables, and Z_e denotes the normalization constant which indicates that the normalization process is performed over possible worlds consistent with the evidence $\vec{e}$. In the next equation, Z_e is removed since it is constant and it does not affect the $argmax$ operation. Similarly, in order to optimize the MAP computation the exp function is left out since it is a monotonic function and only its argument has to be maximized:

$$MAP_{MLN}(\vec{e}) := \vec{e} \cup argmax_{\vec{x}} \sum_i w_i n_i\,(\vec{x}, \vec{e}) \tag{4}$$

The above equation shows that the MAP problem in Markov logic formalism is reduced to a new problem which maximizes the sum of weights of satisfied clauses. Since the MAP determination in Markov networks is an **NP**-hard problem [Domingos and Richardson, 2007], it is performed by exact and approximate solvers. The most commonly used approximate solver is theMaxWalkSAT algorithm, a weighted variant of the WalkSAT local-search satisfiability solver. The MaxWalkSAT algorithm attempts to satisfy clauses with positive weights and keeps clauses with negative weights unsatisfied.

2.5 Combining Markov Logic and Description Logics

Since $\mathcal{ALH}_f{}^-$ is a fragment of first-order logic, its extension to the Markovian style of formalisms is specified in a similar way as for predicate logic in the section before. The formulas in Markov logic correspond to Tbox axioms and Abox assertions. Weights in Markov description logics are associated with axioms and assertions. Groundings of Tbox axioms are defined analogously to the previous case.[3] Abox assertions do not contain variables and are already grounded. Note that due to the restricted use of existential restrictions in $\mathcal{ALH}_f{}^-$ there always exists a model with a domain whose elements correspond the individuals mentioned in an Abox (no "unnamed domain objects" are required).

For appropriately representing domain knowledge, weights might be used only for a subset of the axioms of the domain ontology. The remaining axioms can be assumed to be *strict*, i.e., assumed to be true in any case. A consequence of specifying strict axioms is that lots of possible worlds $\vec{x}$ can be ruled out (i.e., will have probability 0 by definition). This has a direct consequence for implementing sampling approaches to answer probabilistic queries. Using deterministic knowledge (rather than high weights) can speed up Gibbs sampling significantly. Since lots of possible worlds do not have to be considered because their probability is known to be 0, probabilistic reasoning will be significantly faster, given one can show the ergodic character of the corresponding Markov chains. In [Gries and Möller, 2010] we show that Gibbs sampling with deterministic dependencies specified in a fragment of $\mathcal{ALH}_f^-$ remains correct, i.e., probability estimations approximate the correct probabilities. The advantage of this approach is that initial ontology engineering is done as usual with standard reasoning support and with the possibility to add weighted axioms and weighted assertions on top of the strict fundament.

3 Probabilistic Interpretation Engine

At the beginning of this section, the most important preliminaries to the abduction process are specified. Afterwards, functions are introduced for the abduction procedure, interpretation procedure, and the media interpretation agent.

3.1 Computing Explanations

In general, abduction is formalized as $\Sigma \cup \Delta \models_{\mathcal{R}} \Gamma$ where background knowledge (Σ), rules ($\mathcal{R}$), and observations (Γ) are given, and explanations (Δ) are to be computed. In terms of DLs, Δ and Γ are Aboxes and Σ is a pair of Tbox and Abox. Abox abduction is implemented as a non-standard retrieval inference service in DLs. In contrast to standard retrieval inference services where answers are found by exploiting the ontology, Abox abduction has the task of acquiring what should be added to the knowledge base in order to answer a query. Therefore, the result of Abox abduction is a set of hypothesized Abox assertions. To achieve this, the space of abducibles has to be defined. We do this in terms of rules. We assume that a set of rules $\mathcal{R}$ as defined above (see Section 2.2) are specified, and define a non-deterministic function *compute_explanation* as follows.[4]

- $compute\_explanation(\Sigma, \mathcal{R}, \mathcal{A}, P(\underline{z})) = transform(\Phi, \sigma)$ if there exists a rule

$$r = P(\underline{X}) \leftarrow Q_1(\underline{Y_1}), \ldots, Q_n(\underline{Y_n}) \in \mathcal{R}$$

[3] For this purpose, the variable-free syntax of axioms can be first translated to predicate logic.
[4] The funcion $transform$ is defined in Section 2.2.

that is applied to an Abox $\mathcal{A}$ such that a minimal set of atoms Φ and an admissible variable substitution σ with $\sigma(\underline{X}) = \underline{z}$ can be found such that the query $Q := \{() \mid expand(P(\underline{z}), r, \mathcal{R}, \sigma) \backslash \Phi\}$ is answered with *true*. Note that σ might introduce mappings to individuals not mentioned in $\mathcal{A}$ (new individuals). The number of new individuals is bounded by the number of variables.
- If no such rule r exists in $\mathcal{R}$ it holds that $compute\_explanation(\Sigma, \mathcal{R}, \mathcal{A}, P(\underline{z})) = \emptyset$.

The goal of the function *compute_explanation* is to determine what must be added ($\sigma(\Phi)$) such that an entailment $\Sigma \cup \mathcal{A} \cup \Phi \models_{\mathcal{R}} P(\underline{z})$ holds. Hence, for *compute_explanation*, abductive reasoning is used. The set of assertions $\Phi \subseteq expand(P(\underline{X}), r, \mathcal{R}, \sigma)$ represents what needs to be hypothesized in order to answer the query Q with *true*. The definition of *compute_explanation* is non-deterministic due to several possible choices for Φ.

Assuming a fresh name $fresh\_prefix$ for each application of rule $r = P(\underline{X}) \leftarrow Q_1(\underline{Y_1}), \ldots, Q_n(\underline{Y_n})$, the function application $expand(P(\underline{z}), P(\underline{X}) \leftarrow Q_1(\underline{Y_1}), \ldots, Q_n(\underline{Y_n}), \mathcal{R}, \sigma)$ is also defined in a non-deterministic way as

$$expand'(Q_1(\sigma'_{fresh\_prefix}(\underline{Y_1})), \mathcal{R}, \sigma) \cup \cdots \cup expand'(Q_n(\sigma'_{fresh\_prefix}(\underline{Y_n})), \mathcal{R}, \sigma)$$

with $expand'(P(\underline{z}), \mathcal{R}, \sigma)$ being $expand(P(\underline{z}), r, \mathcal{R}, \sigma)$ if there exists a rule $r = (P(\underline{X}) \leftarrow \ldots) \in \mathcal{R}$ and $\langle P(\underline{X}) \rangle$ otherwise. The variable substitution σ' is an extension of σ such that $\sigma'_{prefix}(x) = \sigma(x)$ if $x \in as\_set(\underline{X})$ and, otherwise, $\sigma'_{prefix}(x) = concat(prefix, x)$ where *concat* is a function for concatenating $prefix$ and x.

Applying the expand proedure, we say the set of rules is backward-chained, and since there might be multiple rules in $\mathcal{R}$, backward-chaining is non-deterministic. Thus, multiple explanations are generated.

3.2 The Abduction Procedure

In the following, we devise an abstract computational engine for "explaining" Abox assertions in terms of a given set of rules. Explanation of Abox assertions w.r.t. a set of rules is meant in the sense that using the rules some high-level explanations are constructed such that the Abox assertions are entailed. The explanation of an Abox is again an Abox. For instance, the output Abox represents results of the content interpretation process. The presentation is slightly extended compared to the one in [Castano et al., 2008]. Let the agenda $\mathfrak{A}$ be a set of Aboxes Γ and let Γ be an Abox of observations whose assertions are to be explained. The goal of the explanation process is to use a set of rules $\mathcal{R}$ to derive "explanations" for elements in Γ. The explanation algorithm implemented in the Conceptual Abduction Engine (CAE) works on a set of Aboxes $\mathfrak{I}$.

The complete explanation process is implemented by the CAE function:

Function CAE(Ω, Ξ, Σ, $\mathcal{R}$, S, $\mathfrak{A}$):
Input: a strategy function Ω, a termination function Ξ, a knowledge base Σ, a set of rules $\mathcal{R}$, a scoring function S, and an agenda $\mathfrak{A}$
Output: a set of interpretation Aboxes $\mathfrak{I}'$
$\mathfrak{I}' := \{assign\_level(l, \mathfrak{A})\}$;
repeat
  $\mathfrak{I} := \mathfrak{I}'$;
  $(\mathcal{A}, \alpha) := \Omega(\mathfrak{I})$;
  $l = l + 1$;
  $\mathfrak{I}' := (\mathfrak{A} \setminus \{\mathcal{A}\}) \cup assign\_level(l, explanation\_step(\Sigma, \mathcal{R}, S, \mathcal{A}, \alpha))$;
until $\Xi(\mathfrak{I})$ *or no* $\mathcal{A}$ *and* α *can be selected such that* $\mathfrak{I}' \neq \mathfrak{I}$;
return $\mathfrak{I}'$

where $assign\_level(l, \mathfrak{A})$ is defined as follows:

$$assign\_level(l, \mathfrak{A}) = map(\lambda(\mathcal{A}) \bullet assign\_level(l, \mathcal{A}), \mathfrak{A}) \tag{5}$$

$assign\_level(l, \mathfrak{A})$ takes as input a superscript l and an agenda $\mathfrak{A}$. In the following, $assign\_level(l, \mathcal{A})$ is defined which superscripts each assertion α of the Abox $\mathcal{A}$ with l if the assertion α does not already have a superscript:

$$assign\_level(l, \mathcal{A}) = \left\{\alpha^l \mid \alpha \in \mathcal{A}, \alpha \neq \beta^i, i \in \mathbb{N}\right\} \tag{6}$$

The motivation for adding levels to assertions is to support different strategies Ω. Note that l is a global variable, its starting value is zero, and it is incremented in the CAE function. The *map*[5] function is defined as follows:

$$map(f, X) = \bigcup_{x \in X} \{f(x)\} \tag{7}$$

It takes as parameters a function f and a set X and returns a set consisting of the values of f applied to every element x of X. The CAE function applies the strategy function Ω in order to decide which assertions to explain, uses a termination function Ξ in order to check whether to terminate due to resource constraints and a scoring function S to valuate an explanation. The function Ω for the explanation strategy and Ξ for the termination condition are used as an oracle and must be defined in an application-specific way. The function *explanation_step* is defined as follows.

$explanation\_step(\Sigma, \mathcal{R}, S, \mathcal{A}, \alpha)$:

$$\bigcup_{\Delta \in compute\_all\_explanations(\Sigma, \mathcal{R}, S, \mathcal{A}, \alpha)} consistent\_completed\_explanations(\Sigma, \mathcal{R}, \mathcal{A}, \Delta).$$

We need two additional auxiliary functions.

$consistent\_completed\_explanations(\Sigma, \mathcal{R}, \mathcal{A}, \Delta)$:

$$\{\Delta' \mid \Delta' = \Delta \cup \mathcal{A} \cup forward\_chain(\Sigma, \mathcal{R}, \Delta \cup \mathcal{A}), consistent_{\Sigma}(\Delta')\}$$

$compute\_all\_explanations(\Sigma, \mathcal{R}, S, \mathcal{A}, \alpha)$:

$$maximize(\Sigma, \mathcal{R}, \mathcal{A}, \{\Delta \mid \Delta = compute\_explanation(\Sigma, \mathcal{R}, \alpha), consistent_{\Sigma \cup \mathcal{A}}(\Delta)\}, S).$$

The function $maximize(\Sigma, \mathcal{R}, \mathcal{A}, \Delta s, S)$ selects those explanations $\Delta \in \Delta$s for which the score $S(\Sigma, \mathcal{R}, \mathcal{A}, \Delta)$ is maximal, i.e., there exists no other $\Delta' \in \Delta s$ such that $S(\Sigma, \mathcal{R}, \mathcal{A}, \Delta') > S(\Sigma, \mathcal{R}, \mathcal{A}, \Delta)$. The function $consistent_{(\mathcal{T}, \mathcal{A})}(\mathcal{A}')$ determines if the Abox $\mathcal{A} \cup \mathcal{A}'$ has a model which is also a model of the Tbox $\mathcal{T}$. Note the call to the nondeterministic function *compute_explanation*. It may return different values, all of which are collected. In the next Section we explain how probabilistic knowledge is used to (i) formalize the effect of the "explanation", and (ii) formalize the scoring function S used in the CAE algorithm explained above. In addition, it is shown how the termination condition (represented with the parameter Ξ in the above procedure) can be defined based on the probabilistic conditions.

3.3 The Interpretation Procedure

The interpretation procedure is completely discussed in this section by explaining the interpretation problem and presenting a solution to this problem. The solution is presented by a probabilistic interpretation algorithm which calls the CAE function described in the previous section. In the given algorithm, a termination function, and a scoring function are defined. The termination function determines if the interpretation process can be stopped since at some point during the interpretation process it makes no sense to continue the process. The reason for stopping the interpretation process is that no significant changes can be seen in the results. The defined scoring function in this section assigns probabilistic scores to the interpretation Aboxes.

[5] Please note that in this report, the expression *map* is used in two different contexts. The first one MAP denotes the Maximum A Posteriori approach which is a sampling method whereas the second one *map* is a function used in the $assign\_level(l, \mathfrak{A})$ function.

Problem The objective of the interpretation component is the generation of interpretations for the observations. An interpretation is an Abox which contains high level concept assertions. Since in we adopt the view that agents are used for solving the problems while acquiring information, in the following the same problem is formalized in the perspective of an agent: Consider an agent given some percepts in an environment where the percepts are the analysis results of the multimedia documents.[6] The objective of this agent is finding explanations for the existence of percepts. The question is how the interpretation Aboxes are determined and how long the interpretation process must be performed by the agent. The functionality of this Media Interpretation Agent is presented in the *MI_Agent* algorithm in Section 3.4.

Solution In the following, an application for a probabilistic interpretation algorithm is presented which gives a solution to the mentioned problem. This solution illustrates a new perspective to the interpretation process and the reason why it is performed. In this approach, we define a probabilistic scoring function which assigns probabilities to the interpretation Aboxes. Additionally, we define a termination function which determines whether the interpretation process can be terminated. The central idea is to check whether interpretation results computed by a call to CAE substantially increase the probability the the observations are true. If there is no significant increase (due to a threshold ϵ, possible interpretations are considered as irrelevant for the agent.[7] Another important idea is that, given a "current" interpretation, the agent should be able to compute what must be added due to new percepts and what must be retracted (for this purpose, an Abox difference operator is used).

We are now ready to define the algorithm. Assume that the media interpretation component receives a weighted Abox $\mathcal{A}$ which contains observations. In the following, the applied operation $P(\mathcal{A}, \mathcal{A}', \mathcal{R}, \mathcal{WR}, \mathcal{T})$ in the algorithm is explained:

The $P(\mathcal{A}, \mathcal{A}', \mathcal{R}, \mathcal{WR}, \mathcal{T})$ function determines the probability of the Abox $\mathcal{A}$ with respect to the Abox $\mathcal{A}'$, a set of rules $\mathcal{R}$, a set of weighted rules $\mathcal{WR}$, and the Tbox $\mathcal{T}$ where $\mathcal{A} \subseteq \mathcal{A}'$. Note that $\mathcal{R}$ is a set of forward and backward chaining rules. The probability determination is performed based on the Markov logic formalism as follows:

$$P(\mathcal{A}, \mathcal{A}', \mathcal{R}, \mathcal{WR}, \mathcal{T}) = P_{MLN(\mathcal{A}, \mathcal{A}', \mathcal{R}, \mathcal{WR}, \mathcal{T})}(\vec{Q}(\mathcal{A}) \mid \vec{e}(\mathcal{A}')) \tag{8}$$

$\vec{Q}(\mathcal{A})$ denotes an event composed of the conjunction of all assertions which appear in the Abox $\mathcal{A}$. Assume that the Abox $\mathcal{A}$ contains n assertions $\alpha_1, \ldots, \alpha_n$. Consequently, the event for the Abox $\mathcal{A}$ is defined as follows:

$$\vec{Q}(\mathcal{A}) = \langle \alpha_1 = true \wedge \ldots \wedge \alpha_n = true \rangle \tag{9}$$

Consider Abox $\mathcal{A}$ contains m assertions $\alpha_1, \ldots, \alpha_m$. Then, the evidence vector $\vec{e}(\mathcal{A})$ is defined by:

$$\vec{e}(\mathcal{A}) = \langle \alpha_1 = true, \ldots, \alpha_m = true \rangle \tag{10}$$

Note that $\alpha_1, \ldots, \alpha_n$ denote the boolean random variables of the MLN. In order to answer the query $P_{MLN(\mathcal{A}, \mathcal{A}', \mathcal{R}, \mathcal{WR}, \mathcal{T})}(\vec{Q}(\mathcal{A}) \mid \vec{e}(\mathcal{A}'))$ the function $MLN(\mathcal{A}, \mathcal{A}', \mathcal{R}, \mathcal{WR}, \mathcal{T})$ is called. This function returns a Markov logic network $MLN = (\mathcal{F}_{MLN}, \mathcal{W}_{MLN})$ where $\mathcal{F}_{MLN}$ and $\mathcal{W}_{MLN}$ are ordered sets initialized as follows: $\mathcal{F}_{MLN} = \emptyset$ and $\mathcal{W}_{MLN} = \emptyset$. In the following, it is described how the MLN is built based on the Aboxes $\mathcal{A}$ and $\mathcal{A}'$, the rules $\mathcal{R}$ and $\mathcal{WR}$ and the Tbox $\mathcal{T}$:[8]

$$MLN(\mathcal{A}, \mathcal{A}', \mathcal{R}, \mathcal{WR}, \mathcal{T}) = \begin{cases} \mathcal{F}_{MLN} = \mathcal{F}_{MLN} \cup \{\alpha\}; \quad \mathcal{W}_{MLN} = \mathcal{W}_{MLN} \cup \{w\} & \text{if } w\,\alpha \in \mathcal{A} \\ \mathcal{F}_{MLN} = \mathcal{F}_{MLN} \cup \{\alpha\}; \quad \mathcal{W}_{MLN} = \mathcal{W}_{MLN} \cup \{\infty\} & \text{if } \alpha \in \mathcal{A} \\ \mathcal{F}_{MLN} = \mathcal{F}_{MLN} \cup \{\alpha\}; \quad \mathcal{W}_{MLN} = \mathcal{W}_{MLN} \cup \{w\} & \text{if } w\,\alpha \in \mathcal{A}' \\ \mathcal{F}_{MLN} = \mathcal{F}_{MLN} \cup \{\alpha\}; \quad \mathcal{W}_{MLN} = \mathcal{W}_{MLN} \cup \{\infty\} & \text{if } \alpha \in \mathcal{A}' \\ \mathcal{F}_{MLN} = \mathcal{F}_{MLN} \cup \{\alpha\}; \quad \mathcal{W}_{MLN} = \mathcal{W}_{MLN} \cup \{\infty\} & \text{if } \alpha \in \mathcal{R} \\ \mathcal{F}_{MLN} = \mathcal{F}_{MLN} \cup \{\alpha\}; \quad \mathcal{W}_{MLN} = \mathcal{W}_{MLN} \cup \{w\} & \text{if } w\,\alpha \in \mathcal{WR} \\ \mathcal{F}_{MLN} = \mathcal{F}_{MLN} \cup \{FOL(\alpha)\}; \quad \mathcal{W}_{MLN} = \mathcal{W}_{MLN} \cup \{\infty\} & \text{if } \alpha \in \mathcal{T} \end{cases}$$

[6] The analysis might also be carried out by the agent.

[7] Obviously, there is a horizon problem, which we neglect for the time being.

[8] $FOL(\phi)$ represents the GCI ϕ is first-order notation.

where w and α denote a weight and an assertion, respectively. In the following, the interpretation algorithm *Interpret* is presented:

Function Interpret($\mathfrak{A}$, $CurrentI$, Γ, $\mathcal{T}$, $\mathcal{FR}$, $\mathcal{BR}$, $\mathcal{WR}$, ϵ)
Input: an agenda $\mathfrak{A}$, a current interpretation Abox $CurrentI$, an Abox of observations Γ, a Tbox $\mathcal{T}$, a set of forward chaining rules $\mathcal{FR}$, a set of backward chaining rules $\mathcal{BR}$, a set of weighted rules $\mathcal{WR}$, and the desired explanation significance threshold ϵ
Output: an agenda $\mathfrak{A}'$, a new interpretation Abox $NewI$, and Abox differences for additions Δ_1 and omissions Δ_2
$i := 0$;
$p_0 := P(\Gamma, \Gamma, \mathcal{R}, \mathcal{WR}, \mathcal{T})$;
$\Xi := \lambda(\mathfrak{A}) \bullet \left\{ i := i + 1; p_i := \max_{\mathcal{A} \in \mathfrak{A}} P(\Gamma, \mathcal{A} \cup \mathcal{A}_0, \mathcal{R}, \mathcal{WR}, \mathcal{T}); \textbf{return} \mid p_i - p_{i-1} \mid < \frac{\epsilon}{i} \right\}$;
$\Sigma := (\mathcal{T}, \emptyset)$;
$\mathcal{R} := \mathcal{FR} \cup \mathcal{BR}$;
$S := \lambda((\mathcal{T}, \mathcal{A}_0)), \mathcal{R}, \mathcal{A}, \Delta) \bullet P(\Gamma, \mathcal{A} \cup \mathcal{A}_0 \cup \Delta, \mathcal{R}, \mathcal{WR}, \mathcal{T})$;
$\mathfrak{A}' := CAE(\Omega, \Xi, \Sigma, \mathcal{R}, S, \mathfrak{A})$;
$NewI = argmax_{\mathcal{A} \in \mathfrak{A}'}(P(\Gamma, \mathcal{A}, \mathcal{R}, \mathcal{WR}, \mathcal{T}))$;
$\Delta^+ = AboxDiff(NewI, CurrentI)$; `// additions`
$\Delta^- = AboxDiff(CurrentI, NewI)$; `// omissions`
return $(\mathfrak{A}', NewI, \Delta^+, \Delta^-)$;

In the above algorithm, the termination function Ξ and the scoring function S are defined by lambda calculus terms. The termination condition Ξ of the algorithm is that no significant changes can be seen in the successive probabilities p_i and p_{i-1} (scores) of the two successive generated interpretation Aboxes in two successive levels $i-1$ and i. In this case, the current interpretation Abox $CurrentI$ is preferred to the new interpretation Abox $NewI$. The CAE function is called which returns agenda $\mathfrak{A}'$. Afterwards, the interpretation Abox $NewI$ with the maximum score among the Aboxes $\mathcal{A}$ of $\mathfrak{A}'$ is selected. Additionally, the Abox differences Δ^+ and Δ^-, respectively, for additions and omissions among the interpretation Aboxes $CurrentI$ and $NewI$ are computed.

For the time being, we formalize AboxDiff as set difference, knowing that a semantic approach would be desirable.

In the following, the strategy condition Ω is defined which is one of the parameters of CAE function:

Function $\Omega(\mathfrak{I})$
Input: a set of interpretation Aboxes $\mathfrak{I}$
Output: an Abox $\mathcal{A}$ and a fiat assertion α
$\mathfrak{A} := \left\{ \mathcal{A} \in \mathfrak{I} \mid \neg \exists \mathcal{A}' \in \mathfrak{I}, \mathcal{A}' \neq \mathcal{A} : \exists {\alpha'}^{l'} \in \mathcal{A}' : \forall \alpha^l \in \mathcal{A} : l' < l \right\}$;
$\mathcal{A} := random\_select(\mathfrak{A})$;
$min\_\alpha s = \left\{ \alpha^l \in \mathcal{A} \mid \neg \exists {\alpha'}^{l'} \in \mathcal{A}', {\alpha'}^{l'} \neq \alpha^l, l' < l \right\}$;
return $(\mathcal{A}, random\_select(min\_\alpha s))$;

In the above strategy function Ω, the agenda $\mathfrak{A}$ is a set of Aboxes $\mathcal{A}$ such that the assigned superscripts to their assertions are minimum. In the next step, an Abox $\mathcal{A}$ from $\mathfrak{A}$ is randomly selected. Afterwards, the min_α_s set is determined which contains the assertions α from $\mathcal{A}$ whose superscripts are minimal. These are the assertions which require explanations. The strategy function returns as output an Abox $\mathcal{A}$ and an assertion α which requires explanation.

3.4 The Media Interpretation Agent

In the following, the MI_Agent function is presented which calls the *Interpret* function:

Function MI_Agent(Q, $Partners$, Die, $(\mathcal{T}, \mathcal{A}_0), \mathcal{FR}, \mathcal{BR}, \mathcal{WR}, \epsilon$)
Input: a queue of percept results Q, a set of partners $Partners$, a function Die, a background knowledge base $(\mathcal{T}, \mathcal{A}_0)$, a set of forward chaining rules $\mathcal{FR}$, a set of backward chaining rules $\mathcal{BR}$, a set of weighted rules $\mathcal{WR}$, and the desired precision of the results ϵ
Output: –
$CurrentI = \emptyset$;
$\mathfrak{A}'' = \{\emptyset\}$;
repeat
$\quad \Gamma := extractObservations(Q)$;
$\quad W := MAP(\Gamma, \mathcal{WR}, \mathcal{T})$;
$\quad \Gamma' := select(W, \Gamma)$;
$\quad \mathfrak{A}' := filter(\lambda(\mathcal{A}) \bullet consistent_{\Sigma}(\mathcal{A}), map(\lambda(\mathcal{A}) \bullet \Gamma' \cup \mathcal{A} \cup \mathcal{A}_0 \cup forward\_chain(\Sigma, \mathcal{FR}, \Gamma' \cup \mathcal{A} \cup \mathcal{A}_0),$
$\qquad \{select(MAP(\Gamma' \cup \mathcal{A} \cup \mathcal{A}_0, \mathcal{WR}, \mathcal{T}), \Gamma' \cup \mathcal{A} \cup \mathcal{A}_0) \mid \mathcal{A} \in \mathfrak{A}''\}))$;
$\quad (\mathfrak{A}'', NewI, \Delta^+, \Delta^-) := Interpret(\mathfrak{A}', CurrentI, \Gamma', \mathcal{T}, \mathcal{FR}, \mathcal{BR}, \mathcal{WR} \cup \Gamma, \epsilon)$;
$\quad CurrentI := NewI$;
$\quad Communicate(\Delta^+, \Delta^-, Partners)$;
$\quad \mathfrak{A}'' := manage\_agenda(\mathfrak{A}'')$;
until *Die()* ;

The body of MI_Agent uses a set of auxiliary functions, which are defined as follows.

$$filter(f, X) = \bigcup_{x \in X} \begin{cases} \{x\} & \text{if } f(x) = true \\ \emptyset & \text{else} \end{cases} \tag{11}$$

The function *filter* takes as parameters a function f and a set X and returns a set consisting of the values of f applied to every element x of X. In the MI_Agent function, the current interpretation $CurrentI$ is initialized to empty set and the agenda $\mathfrak{A}''$ to a set containing empty set. Since the agent performs an incremental process, it is defined by a repeat-loop. The percept results Γ are sent to the queue Q. In order to take the observations Γ from the queue Q, the MI_Agent calls the *extractObservations* function.

In the following we assume that Γ represents an ordered set. The $MAP(\Gamma, \mathcal{WR}, \mathcal{T})$ determines the most probable world of observations Γ with respect to a set of weighted rules $\mathcal{WR}$ and the Tbox $\mathcal{T}$. This function performs the MAP process defined in Section 2. It returns a vector W which consists of zeros and ones assigned to indicate whether the ground atoms of the considered world are true (positive) and false (negative), respectively. The function $select(W, \Gamma)$ then selects the positive assertions in the input Abox Γ using the bit vector W. The selected positive assertions are the assertions which require explanations. The *Select* operation returns as output an Abox Γ' which has the following characteristic: $\Gamma' \subseteq \Gamma$. The determination of the most probable world by the MAP function and the selection of the positive assertions is also carried out on $\Gamma' \cup \mathcal{A} \cup \mathcal{A}_0$.

In the next step, a set of forward chaining rules $\mathcal{FR}$ is applied to $\Gamma' \cup \mathcal{A} \cup \mathcal{A}_0$. The generated assertions in this process are added to the to $\Gamma' \cup \mathcal{A} \cup \mathcal{A}_0$. In the next step, only the consistent Aboxes are selected and the inconsistent Aboxes are removed. Afterwards, the *Interpret* function is called to determine the new agenda $\mathfrak{A}''$, the new interpretation Abox $NewI$ and the Abox differences Δ_1 and Δ_2 for additions and omissions among $CurrentI$ and $NewI$. Afterwards, the $CurrentI$ is set to the $NewI$ and the MI_Agent function communicates the Abox differences Δ_1 and Δ_2 to the $Partners$. The *manage_agenda* function is also called. This function is explained in Section 6. The termination condition of the MI_Agent function is that the $Die()$ function is true. Note that the MI_Agent waits in the function call $extractObservations(Q)$ if $Q = \emptyset$.

After presenting the above algorithms, the mentioned unanswered questions can be discussed. A reason for performing the interpretation process and explaining the fiat assertions is that the probability of $P(\mathcal{A}, \mathcal{A}', \mathcal{R}, \mathcal{WR}, \mathcal{T})$ will increase through the interpretation process. In other words, by explaining the observations the agent's belief to the percepts will increase. This shows a new perspective for performing the interpretation process. The answer to the question whether there is any measure for stopping the interpretation process, is indeed positive. This is expressed by $| p_i - p_{i-1} | < \frac{\epsilon}{i}$ which is the termination condition Ξ of the algorithm. The reason for selecting $\frac{\epsilon}{i}$ and not ϵ as the upper limit for the termination condition is to terminate the oscillation behaviour of the results. In other words, the precision interval is tightened step by step during the interpretation process. In Section 4, we discuss an example for interpreting a single video shot.

4 Preference-Based Video Shot Interpretation

One of the main innovations introduced in the previous section, namely the introduction of a probabilistic preference measure to control the space of possible interpretations, is demonstrated here using examples from an environmental domain.

We have to mention that this example is not constructed to show the possible branchings through the interpretation process. The purpose of this example is to show how the probabilities of the most probable world of observations $P(\mathcal{A}_0, \mathcal{A}, \mathcal{R}, \mathcal{WR}, \mathcal{T})$ behave during the interpretation process.

At the beginning of this example, the **signature** of the knowledge base is presented. The set of all concept names **CN** is divided into two disjoint sets **Events** and **PhysicalThings** such that

CN = **Events** ∪ **PhysicalThings** where these two sets are defined as follows:

Events = $\{CarEntry, EnvConference, EnvProt, HealthProt\}$

PhysicalThings = $\{Car, DoorSlam, Building, Environment, Agency\}$

$EnvConference$, $EnvProt$ and $HealthProt$ denote respectively environmental conference, environmental protection and health protection. The set of role names **RN** is defined as follows:

RN = $\{Causes, OccursAt, HasAgency, HasTopic, HasSubject, HasObject, HasEffect,$
$HasSubEvent, HasLocation\}$

In the following, the set of individual names **IN** is given:

IN = $\{C_1, DS_1, ES_1, Ind_{42}, Ind_{43}, Ind_{44}, Ind_{45}, Ind_{46}, Ind_{47}, Ind_{48}\}$

In the following, the set of the forward chaining rules $\mathcal{FR}$ is defined:

$\mathcal{FR} = \{\forall x \;\; CarEntry(x) \rightarrow \exists y \;\; Building(y), OccursAt(x, y),$
$\forall x \;\; EnvConference(x) \rightarrow \exists y \;\; Environment(y), HasTopic(x, y),$
$\forall x \;\; EnvProt(x) \rightarrow \exists y \;\; Agency(y), HasAgency(x, y)\}$

Similarly, the set of backward chaining rules $\mathcal{BR}$ is given as follows:

$\mathcal{BR} = \{Causes(x, y) \leftarrow CarEntry(z), HasObject(z, x), HasEffect(z, y), Car(x), DoorSlam(y),$
$OccursAt(x, y) \leftarrow EnvConference(z), HasSubEvent(z, x), HasLocation(z, y), CarEntry(x), Building(y),$
$HasTopic(x, y) \leftarrow EnvProt(z), HasSubEvent(z, x), HasObject(z, y), EnvConference(x), Environment(y),$
$HasAgency(x, y) \leftarrow HealthProt(z), HasObject(z, x), HasSubject(z, y), EnvProt(x), Agency(y)\}$

In the following, a set of weighted rules $\mathcal{WR}$ is defined where the weight of each rule is 5:

$\mathcal{WR} = \{5 \; \forall x, y, z \; CarEntry(z) \wedge HasObject(z, x) \wedge HasEffect(z, y) \rightarrow Car(x) \wedge DoorSlam(y) \wedge Causes(x, y),$
$5 \; \forall x, y, z \; EnvConference(z) \wedge HasSubEvent(z, x) \wedge HasLocation(z, y) \rightarrow CarEntry(x) \wedge Building(y) \wedge OccursAt(x, y),$
$5 \; \forall x, y, z \; EnvProt(z) \wedge HasSubEvent(z, x) \wedge HasObject(z, y) \rightarrow EnvConference(x) \wedge Environment(y) \wedge HasTopic(x, y),$
$5 \; \forall x, y, z \; HealthProt(z) \wedge HasObject(z, x) \wedge HasSubject(z, y) \rightarrow EnvProt(x) \wedge Agency(y) \wedge HasAgency(x, y)\}$

The selected value for ϵ in this example is 0.05. In the following, Δ_1 and Δ_2 denote respectively the set of assertions hypothesized by a forward chaining rule and the set of assertions generated by a backward chaining rule at each interpretation level. Let us assume that the media interpretation agent receives the following weighted Abox $\mathcal{A}$:

$\mathcal{A} = \{1.3 \; Car(C_1), 1.2 \; DoorSlam(DS_1), -0.3 \; EngineSound(ES_1), Causes(C_1, DS_1)\}$

The first applied operation to A is the MAP function which returns the bit vector $W = \langle 1, 1, 0, 1 \rangle$. By applying the *select* function to W and the input Abox $\mathcal{A}$, the assertions from the input Abox $\mathcal{A}$ are selected that correspond to the positive events in W. Additionally, the assigned weights to the positive assertions are also taken from the input Abox $\mathcal{A}$. In the following, Abox A_0 is depicted which contains the positive assertions:

$\mathcal{A}_0 = \{1.3 \; Car(C_1), 1.2 \; DoorSlam(DS_1), Causes(C_1, DS_1)\}$

At this step, $p_0 = P(\mathcal{A}_0, \mathcal{A}_0, \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.755$. Since no appropriate forward chaining rule from $\mathcal{FR}$ is applicable to Abox $\mathcal{A}_0$, $\Delta_1 = \emptyset$ and as a result $\mathcal{A}_0 = \mathcal{A}_0 \cup \emptyset$. The next step is the performance of *backward_chain* function where the next backward chaining rule from $\mathcal{BR}$ can be applied to Abox $\mathcal{A}_0$:

$Causes(x, y) \leftarrow CarEntry(z), HasObject(z, x), HasEffect(z, y), Car(x), DoorSlam(y)$

Consequently, by applying the above rule the next set of assertions is hypothesized:

$\Delta_2 = \{CarEntry(Ind_{42}), HasObject(Ind_{42}, C_1), HasEffect(Ind_{42}, DS_1)\}$

which are considered as strict assertions. Consequently, $\mathcal{A}_1$ is defined as follows: $A_1 = \mathcal{A}_0 \cup \Delta_2$. In the above Abox, $p_1 = P(\mathcal{A}_0, \mathcal{A}_1, \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.993$. As it can be seen, $p_1 > p_0$ i.e. $P(\mathcal{A}_0, \mathcal{A}_i, \mathcal{R}, \mathcal{WR}, \mathcal{T})$ increases by adding the new hypothesized assertions. This shows that the new assertions are considered as additional support. The termination condition of the algorithm is not fulfilled therefore the algorithm continues processing. At this level, it is still not known whether Abox $\mathcal{A}_1$ can be considered as the final interpretation Abox. Thus, this process is continued with another level. Consider the next forward chaining rule:

$\forall x \;\; CarEntry(x) \rightarrow \exists y \;\; Building(y), OccursAt(x, y)$

By applying the above rule, the next set of assertions is generated namely:

$\Delta_1 = \{Building(Ind_{43}), OccursAt(Ind_{42}, Ind_{43})\}$

The generated assertions are also considered as strict assertions. In the following, the expanded Abox $\mathcal{A}_1$ is defined as follows: $\mathcal{A}_1 := \mathcal{A}_1 \cup \Delta_1$.

Let us assume the next backward chaining rule from $\mathcal{BR}$:

$OccursAt(x, y) \leftarrow EnvConference(z), HasSubEvent(z, x), HasLocation(z, y), CarEntry(x), Building(y)$

Consequently, by applying the above abduction rule the next set of assertions is hypothesized:

$\Delta_2 = \{EnvConference(Ind_{44}), HasSubEvent(Ind_{44}, Ind_{42}), HasLocation(Ind_{44}, Ind_{43})\}$

which are considered as strict assertions. Consequently, $A_2 = A_1 \cup \Delta_2$.

In the above Abox, $p_2 = P(\mathcal{A}_0, \mathcal{A}_2, \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.988$. As it can be seen, $p_2 < p_1$ i.e. $P(\mathcal{A}_0, \mathcal{A}_i, \mathcal{R}, \mathcal{WR}, \mathcal{T})$ decreases slightly by adding the new hypothesized assertions. Since the termination condition of the algorithm is fulfilled, Abox $\mathcal{A}_1$ can be considered as the final interpretation Abox. To realize how the further behaviour of the probabilities is, this process is continued for the sake of illustration. Consider the next forward chaining rule from $\mathcal{FR}$:

$\forall x \;\; EnvConference(x) \rightarrow \exists y \;\; Environment(y), HasTopic(x, y)$

By applying the above rule, new assertions are generated.

$\Delta_1 = \{Environment(Ind_{45}), HasTopic(Ind_{44}, Ind_{45})\}$

In the following, the expanded Abox $\mathcal{A}_2$ is defined: $\mathcal{A}_2 = \mathcal{A}_2 \cup \Delta_1$.

Consider the next backward chaining rule from $\mathcal{BR}$:

$HasTopic(x, y) \leftarrow EnvProt(z), HasSubEvent(z, x), HasObject(z, y), EnvConference(x), Environment(y)$

By applying the above abduction rule, the following set of assertions is hypothesized:

$\Delta_2 = \{EnvProt(Ind_{46}), HasSubEvent(Ind_{46}, Ind_{44}), HasObject(Ind_{46}, Ind_{45})\}$

which are considered as strict assertions. In the following, $\mathcal{A}_3$ is defined as follows $A_3 = A_2 \cup \Delta_2$. In the above Abox $\mathcal{A}_3$, $p_3 = P(\mathcal{A}_0, \mathcal{A}_3, \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.99$. As it can be seen, $p_3 > p_2$, i.e.

$P(\mathcal{A}_0, \mathcal{A}_i, \mathcal{R}, \mathcal{WR}, \mathcal{T})$ increases slightly by adding the new hypothesized assertions.

Consider the next forward chaining rule:

$\forall x \;\; EnvProt(x) \rightarrow \exists y \;\; Agency(y), HasAgency(x, y)$

By applying the above rule, the next assertions are generated:

$\Delta_1 = \{Agency(Ind_{47}), HasAgency(Ind_{46}, Ind_{47})\}$

As a result, the expanded Abox $\mathcal{A}_3$ is presented as follows: $\mathcal{A}_3 = \mathcal{A}_3 \cup \Delta_1$.

Let us consider the next backward chaining rule from $\mathcal{BR}$:

$HasAgency(x, y) \leftarrow HealthProt(z), HasObject(z, x), HasSubject(z, y), EnvProt(x), Agency(y)$

Consequently, new assertions are hypothesized by applying the above abduction rule, namely:

$\Delta_2 = \{HealthProt(Ind_{48}), HasObject(Ind_{48}, Ind_{46}), HasSubject(Ind_{48}, Ind_{47})\}$

which are considered as strict assertions. Consequently, $\mathcal{A}_4$ is defined as follows: $A_4 = A_3 \cup \Delta_2$.

In the above Abox, $p_4 = P(\mathcal{A}_0, \mathcal{A}_4, \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.985$. As it can be seen, $p_4 < p_3$, i.e.

$P(\mathcal{A}_0, \mathcal{A}_i, \mathcal{R}, \mathcal{WR}, \mathcal{T})$ decreases slightly by adding the new hypothesized assertions.

Discussion of the Results:

The determined probability values $P(\mathcal{A}_0, \mathcal{A}_i, \mathcal{R}, \mathcal{WR}, \mathcal{T})$ of this example are summarized in the next table which shows clearly the behaviour of the probabilities stepwise after performing the interpretation process. In this table, variable i denotes the successive levels of the interpretation process.

i	*Abox* $\mathcal{A}_i$	$p_i = P(\mathcal{A}_0, \mathcal{A}_i, \mathcal{R}, \mathcal{WR}, \mathcal{T})$
0	$\mathcal{A}_0$	$p_0 = 0.755$
1	$\mathcal{A}_1$	$p_1 = 0.993$
2	$\mathcal{A}_2$	$p_2 = 0.988$
3	$\mathcal{A}_3$	$p_3 = 0.99$
4	$\mathcal{A}_4$	$p_4 = 0.985$

In this example, the interpretation process is consecutively performed four times. As it can be seen, through the first interpretation level the probability p_1 increases strongly in comparison to p_0. By performing the second, third and the forth interpretation levels, the probability values decrease slightly in comparison to p_1. This means no significant changes can be seen in the results. In other words, the determination of $\mathcal{A}_3$ and $\mathcal{A}_4$ were not required at all. But the determination of $\mathcal{A}_2$ was required to realize the slight difference $|p_2 - p_1| < \frac{\epsilon}{2}$. Consequently, Abox $\mathcal{A}_1$ is considered as the final interpretation Abox.

5 Preference-based Scene Interpretation

In this example, we discuss how an interpretation process is performed by considering the analysis results of two consecutive video shots. For the interpretation of each video shot we require information about the previous video shots, otherwise the interpretation process does not work as intended. The question is which assertions have to be considered from the previous video shots. As was discussed in this paper we would like to consider the assertions from the previous video shots which increase $P(\mathcal{A}_0, \mathcal{A}_i, \mathcal{R}, \mathcal{WR}, \mathcal{T})$. At the beginning of this example, the signature of the knowledge base is presented. The set of the concept names **CN** is divided into two disjoint sets **Events** and **PhysicalThings** which are described as follows:

Events $= \{CarEntry, CarExit, CarRide\}$

PhysicalThings $= \{Car, DoorSlam\}$

Additionally, the set of the role names **RN** and the set of the individual names **IN** are represented as follows:

RN $= \{Causes, HasObject, HasEffect, Before, HasStartEvent, HasEndEvent\}$

IN $= \{C_1, C_2, DS_1, DS_2, Ind_{41}, Ind_{42}, Ind_{44}\}$

The Tbox $\mathcal{T}$ contains the axiom $CarEntry \sqsubseteq \neg CarExit$. In the following, the set of the forward chaining rules $\mathcal{FR}$ is given:

$$
\begin{aligned}
\mathcal{FR} = \{ \\
\forall x, xl, y, yl, w, z\ & AudioSeg(x), HasSegLoc(x, xl), VideoSeg(y), HasSegLoc(y, yl), IsSmaller(xl, yl), \\
& Depicts(x, w), Depicts(y, z), CarEntry(w), CarEntry(z) \rightarrow Before(z, w), \\
\forall x, xl, y, yl, w, z\ & AudioSeg(x), HasSegLoc(x, xl), VideoSeg(y), HasSegLoc(y, yl), IsSmaller(xl, yl), \\
& Depicts(x, w), Depicts(y, z), CarEntry(w), CarExit(z) \rightarrow Before(z, w), \\
\forall x, xl, y, yl, w, z\ & AudioSeg(x), HasSegLoc(x, xl), VideoSeg(y), HasSegLoc(y, yl), IsSmaller(xl, yl), \\
& Depicts(x, w), Depicts(y, z), CarExit(w), CarEntry(z) \rightarrow Before(z, w), \\
\forall x, xl, y, yl, w, z\ & AudioSeg(x), HasSegLoc(x, xl), VideoSeg(y), HasSegLoc(y, yl), IsSmaller(xl, yl), \\
& Depicts(x, w), Depicts(y, z), CarExit(w), CarExit(z) \rightarrow Before(z, w)\}
\end{aligned}
$$

where $AudioSeg$, $HasSegLoc$ and $VideoSeg$ denote $AudioSegment$, $HasSegmentLocator$ and $VideoSegment$ respectively. Note that the concepts and roles in $\mathcal{FR}$ which are not given in **CN** and **RN** appear only in the multimedia content ontology. The multimedia content ontology determines the structure of the multimedia document. Additionally, it determines whether the concpts are originated from video, audio or text. The above rules mean that the concept assertion $CarEntry$ or $CarExit$ from the first video shot appear chronologically before the concept assertion $CarEntry$ or $CarExit$ from the second video shot. The set of the backward chaining rules $\mathcal{BR}$ is presented as follows:

$$
\begin{aligned}
\mathcal{BR} = \{ & Causes(x, y) \leftarrow CarEntry(z), HasObject(z, x), HasEffect(z, y), Car(x), DoorSlam(y), \\
& Causes(x, y) \leftarrow CarExit(z), HasObject(z, x), HasEffect(z, y), Car(x), DoorSlam(y),
\end{aligned}
$$

$$Before(x, y) \leftarrow CarRide(z), HasStartEvent(z, x), HasEndEvent(z, y), CarEntry(x), CarExit(y)\}$$

Additionally, the set of the weighted rules is defined as follows:

$$\begin{aligned}\mathcal{WR} = \{&5\ \forall x, y, z\ CarEntry(z) \wedge HasObject(z, x) \wedge HasEffect(z, y) \Rightarrow Car(x) \wedge DoorSlam(y) \wedge Causes(x, y),\\ &5\ \forall x, y, z\ CarExit(z) \wedge HasObject(z, x) \wedge HasEffect(z, y) \Rightarrow Car(x) \wedge DoorSlam(y) \wedge Causes(x, y),\\ &5\ \forall x, y, z, k, m\ \ CarRide(z) \wedge HasStartEvent(z, x) \wedge HasEndEvent(z, y) \wedge HasObject(x, k) \wedge\\ &\qquad HasObject(y, m) \Rightarrow CarEntry(x) \wedge CarExit(y) \wedge Car(k) \wedge Car(m) \wedge k = m\}\end{aligned}$$

The selected value for ϵ in this example is 0.05. Consider the next figure as the first video shot of a video.

Let us assume that the analysis results of the first video shot represented in the Abox $\mathcal{A}_1$ are sent to the queue Q:

$\mathcal{A}_1 = \{1.3\ Car(C_1), 1.2\ DoorSlam(DS_1), Causes(C_1, DS_1)\}$

For the interpretation of the first video shot, we will call the function $MI\_Agent(Q, Partners, Die, (\mathcal{T}, \mathcal{A}_0), \mathcal{FR}, \mathcal{BR}, \mathcal{WR}, \epsilon)$. At the beginning of this function, there are initializations for some variables, namely $CurrentI = \emptyset$ and $\mathfrak{A}'' = \{\emptyset\}$. Afterwards extracting observations from the queue Q is performed, which leads to $\Gamma = \mathcal{A}_1$. Determination of the most probable world $W = \langle 1, 1, 1\rangle$ is performed in the next step and selecting the positive assertions and their related weights determines $\Gamma' = \Gamma$. At this step, $\mathcal{A} = \emptyset$ since $\mathfrak{A}'' = \{\emptyset\}$. Additionally, $\mathcal{A}_0 = \emptyset$. Consequently, $MAP(\Gamma', \mathcal{WR}, \mathcal{T}) = W$ and $select(W, \Gamma') = \Gamma'$. $forward\_Chain(\Sigma, \mathcal{FR}, \Gamma') = \emptyset$ since there is no forward chaining rule applicable to Γ'. $\mathfrak{A}' = \Gamma'$. The $Interpret(\mathfrak{A}', CurrentI, \Gamma', \mathcal{T}, \mathcal{FR}, \mathcal{BR}, \mathcal{WR} \cup \Gamma, \epsilon)$ is called in the next step which determines $p_0 = P(\Gamma', \Gamma', \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.733$. The $Interpret$ function calls CAE function which returns $\mathfrak{A}' = \{\Gamma' \cup \Delta_1, \Gamma' \cup \Delta_2\}$ where the two possible explanations Δ_1 and Δ_2 are defined as follows:

$\Delta_1\ = \{CarEntry(Ind_{41}), HasObject(Ind_{41}, C_1), HasEffect(Ind_{41}, DS_1)\}$

$\Delta_2 = \{CarExit(Ind_{41}), HasObject(Ind_{41}, C_1), HasEffect(Ind_{41}, DS_1)\}$

Each of the above interpretation Aboxes have scoring values:

$p_1 = P(\Gamma', \Gamma' \cup \Delta_1, \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.941$ and $p_1 = P(\Gamma', \Gamma' \cup \Delta_2, \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.935$. $NewI = \Gamma' \cup \Delta_1$ since this is the interpretation Abox with the maximum scoring value. The termination condition is not fulfilled since $p_1 - p_0 = 0.208 > 0.05$. The Abox difference for additions is defined as follows: $\Delta^+ = NewI - CurrentI = NewI - \emptyset = NewI$. Simiarly, $\Delta^- = \emptyset$ is the Abox difference for the omissions. The CAE function returns $NewI$, $\mathfrak{A}'$ and the Abox differences Δ^+ and Δ^- to the $Interpret$ function. Consider the next figure depicts the second video shot. Assume that the analysis results of the second video shot given in the next Abox are sent to the queue Q:

$\mathcal{A}_2 = \{1.3\ Car(C_2), 1.2\ DoorSlam(DS_2), Causes(C_2, DS_2)\}$

Similarly, for the interpretation of the second video shot we will call the function $MI\_Agent(Q, Partners, Die, (\mathcal{T}, \mathcal{A}_0), \mathcal{FR}, \mathcal{BR}, \mathcal{WR}, \epsilon)$. The observation extraction process from Q leads to $\Gamma = \mathcal{A}_2$. Afterwards, the most probable world $W = \langle 1, 1, 1\rangle$ is determined and applying $select$ function on W gives $\Gamma' = \mathcal{A}_2$. Consider $\mathcal{A} \in \mathfrak{A}''$ where $\mathfrak{A}'' = \{\mathcal{A}_1 \cup \Delta_1, \mathcal{A}_1 \cup \Delta_2\}$. $\Gamma' \cup \mathcal{A} = \{\mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_1, \mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_2\}$. Applying $MAP(\Gamma' \cup \mathcal{A}, \mathcal{WR}, \mathcal{T})$ gives $W = \langle 1, \ldots, 1\rangle$ and applying the $select(W, \Gamma' \cup \mathcal{A})$ function gives $\{\mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_1, \mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_2\}$. Since no forward chaining rule is applicable to the above set and this set contains consistent Aboxes, $\mathfrak{A}' = \{\mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_1, \mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_2\}$. In the next step, the function $Interpret(\mathfrak{A}', CurrentI, \Gamma', \mathcal{T}, \mathcal{FR}, \mathcal{BR}, \mathcal{WR} \cup \Gamma, \epsilon)$ is called which determines $P(\Gamma', \Gamma', \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.733$. Afterwards, the CAE function is called which determines the next exaplanations:

$\Delta_3 = \{CarEntry(Ind_{42}), HasObject(Ind_{42}, C_2), HasEffect(Ind_{42}, DS_2)\}$

$\Delta_4 = \{CarExit(Ind_{42}), HasObject(Ind_{42}, C_2), HasEffect(Ind_{42}, DS_2)\}$

The CAE function generates the following agenda which contains all possible interpretation Aboxes $\{I_1, I_2, I_3, I_4\}$ where:

$I_1 = \mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_1 \cup \Delta_3$ $\qquad$ $I_2 = \mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_1 \cup \Delta_4$

$I_3 = \mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_2 \cup \Delta_3$ $\qquad$ $I_4 = \mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_2 \cup \Delta_4$

Afterwards applies the forward chaining rules on the above agenda. A new assertion $\mathit{Before}(Ind_{41}, Ind_{42})$ is generated and added to the four interpretation Aboxes. In the following, the possible four interpretation Aboxes are given:

$$I_1 = \mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_1 \cup \Delta_3 \cup \{\mathit{Before}(Ind_{41}, Ind_{42})\}$$
$$I_2 = \mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_1 \cup \Delta_4 \cup \{\mathit{Before}(Ind_{41}, Ind_{42})\}$$
$$I_3 = \mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_2 \cup \Delta_3 \cup \{\mathit{Before}(Ind_{41}, Ind_{42})\}$$
$$I_4 = \mathcal{A}_2 \cup \mathcal{A}_1 \cup \Delta_2 \cup \Delta_4 \cup \{\mathit{Before}(Ind_{41}, Ind_{42})\}$$

Afterwards the backward chaining rule is applied which generates the following set only for the interpretation Abox I_2:

$$\Delta = \{CarRide(Ind_{44}), HasStartEvent(Ind_{44}, Ind_{41}), HasEndEvent(Ind_{44}, Ind_{42})\}$$

Consequently $I_2 = I_2 \cup \Delta$. The interpretation Aboxes have the next scoring values:

$$P(\mathcal{A}_1 \cup \mathcal{A}_2, I_1, \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.964$$
$$P(\mathcal{A}_1 \cup \mathcal{A}_2, I_2, \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.978$$
$$P(\mathcal{A}_1 \cup \mathcal{A}_2, I_3, \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.952$$
$$P(\mathcal{A}_1 \cup \mathcal{A}_2, I_4, \mathcal{R}, \mathcal{WR}, \mathcal{T}) = 0.959$$

The above values show that the interpretation Abox I_2 has a higher scoring value than the other interpretation Aboxes. Therefore the final interpretation Abox is $NewI = I_2$. The Abox differences for additions and omissions are defined as follows:

$$\Delta^+ = \mathcal{A}_2 \cup \Delta_4 \cup \Delta \cup \{\mathit{Before}(Ind_{41}, Ind_{42})\} \qquad \Delta^- = \emptyset$$

For the next interpretation steps, the agenda can continue with I_2 and eliminate the other interpretation Aboxes since this Abox has a higher scoring value.

6 Manage Agenda

The *manage_agenda*($\mathfrak{A}$) function is called in the *MI_Agent* function to improve its performance. In this section, we briefly introduce some techniques which are applied by the *manage_agenda* function to an agenda $\mathfrak{A}$ which contains multiple interpretation Aboxes.

– Elimination of the interpretation Aboxes: This technique is applied if there are multiple interpretation Aboxes with different scoring values where one of the Aboxes has a higher scoring value. At this step, we can select this Abox, eliminate the remaining interpretation Aboxes and continue the interpretation process with the selected Abox.
– Combining the interpretation Aboxes: Consider the interpretation Aboxes $I_1, \ldots, I_n$. In order to determine the final interpretation Abox, the MAP process can be applied to the union of all interpretation Aboxes $I_1 \cup \ldots \cup I_n$. The MAP process determines the most probable world based on the Tbox $\mathcal{T}$ and the set of weighted rules $\mathcal{WR}$.
– Shrinking the interpretation Aboxes: This approach determines which assertions from the previous video shots have to be considered for the interpretation process of the next video shots since considering all assertions of the previous video shots will slow down the interpretation process. We believe that only the high level concept assertions from the previous video shots play an important role and not the low level concept assertions.

7 Summary

For multimedia interpretation, a semantically well-founded formalization is required. In accordance with previous work, a well-founded abduction-based approach is pursued. Extending previous work, abduction is controlled by probabilistic knowledge, and it is done in terms of first-order logic. Rather than merely using abduction for computing explanation with which observations are entailed, the approach presented in this paper, uses a probabilistic logic to motivate the explanation endeavor by increasing the belief in the observations. Hence, there exists a certain utility for an agent for the computational resources it spends for generating explanations. Thus, we have presented a first attempt to more appropriately model a media interpretation agent. Additionally, we have discussed how the video shot interpretation process is perfomed. A manage agenda is also introduced which improves the interpretation process by applying some techniques. Describing the mentioned techniques is our future work.

References

[Castano et al., 2008] Castano, S., Espinosa, S., Ferrara, A., Karkaletsis, V., Kaya, A., Möller, R., Montanelli, S., Petasis, G., and Wessel, M. (2008). Multimedia interpretation for dynamic ontology evolution. In *Journal of Logic and Computation*. Oxford University Press.

[Domingos and Richardson, 2007] Domingos, P. and Richardson, M. (2007). Markov logic: A unifying framework for statistical relational learning. In Getoor, L. and Taskar, B., editors, *Introduction to Statistical Relational Learning*, pages 339–371. Cambridge, MA: MIT Press.

[Espinosa-Peraldi et al., 2010a] Espinosa-Peraldi, S., Kaya, A., and Möller, R. (2010a). *BOEMIE: State of the Art in Ontology-based Multimedia Interpretation*, chapter Logical Formalization of Multimedia Interpretation. Springer.

[Espinosa-Peraldi et al., 2010b] Espinosa-Peraldi, S., Kaya, A., and Möller, R. (2010b). Formalizing multimedia interpretation based on abduction over description logic aboxes. In Cuena-Grau, B., Horrocks, I., and Motik, B., editors, *Proc. of the 22nd International Workshop on Description Logics (DL2009)*.

[Gries and Möller, 2010] Gries, O. and Möller, R. (2010). Gibbs sampling in probabilistic description logics with deterministic dependencies. In *Proc. of the First International Workshop on Uncertainty in Description Logics, Edinburgh*.

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