DESCRIPTION PLAUSIBLE LOGIC PROGRAMS FOR STREAM REASONING

Ioan Alfred Letia and Adrian Groza

Department of Computer Science, Technical University of Cluj-Napoca, Cluj-Napoca, Romania

Keywords: Stream reasoning, Description logic, Plausible logic, Lazy evaluation, Sensors.

Abstract: Stream reasoning is defined as real time logical reasoning on large, noisy, heterogeneous data streams, aiming to support the decision process of large numbers of concurrent querying agents. In this research we exploit nonmonotonic rule-based systems for handling inconsistent or incomplete information and also ontologies to deal with heterogeneity. Data is aggregated from distributed streams in real time and plausible rules fire when new data is available. This study also investigates the advantages of lazy evaluation on data streams.

1 INTRODUCTION

Sensor networks are estimated to drive the formation of a new Web, by 2015 (Le-Phuoc et al., 2010). The value of the Sensor Web is related to the capacity to aggregate, analyse and interpret this new source of knowledge. Currently, there is a lack of systems designed to manage rapidly changing information at the semantic level (Valle et al., 2009). The solution given by data-stream management systems (DSMS) is limited mainly by the incapacity to perform complex reasoning tasks.

Stream reasoning is defined as real time logical reasoning on huge, possible infinite, noisy data streams, aiming to support the decision process of large numbers of concurrent querying agents. In order to handle blocking operators on infinite streams (like min, mean, average, sort), the reasoning process is restricted to a certain window of concern within the stream, whilst the previous information is discharged (Barbieri et al., 2010). This strategy is applicable only for applications where recent data have higher relevance (e.g. average water debit in the last 10 minutes). In some reasoning tasks, tuples need to be joined arbitrarily far apart from different streams. Stream Reasoning adopts the continuous processing model, where reasoning goals are continuously evaluated against a dynamic knowledge base. This leads to the concept of transient queries, opposite to the persistent queries in a database. Typical applications of stream reasoning are: traffic monitoring, urban computing, patient monitoring, weather monitoring from satellite data, monitoring financial transactions (Valle et al., 2009) or stock market. Real time events analysis is conducted in domains like seismic incidents, flu outbreaks, or tsunami alert based on a wide range of sensor networks starting from the RFID technology to the Twitter dataflow (Savage, 2011). Decisions should be taken based on plausible events. Waiting to have complete confirmation of an event might be too risky action.

Streams of sensor data are often characterised by heterogeneity, noise and contradictory data. In this research we exploit nonmonotonic rule-based systems for handling inconsistent or incomplete information and also ontologies to deal with heterogeneity. Data is aggregated from distributed streams in real time and plausible rules fire when new data is available. This study investigates the advantages of lazy evaluation on data streams, as well.

2 INTEGRATING PLAUSIBLE RULES WITH ONTOLOGIES

2.1 Plausible Logic

Plausible logic is an improvement of defeasible logic (Rock, 2010; Billington and Rock, 2001). A clause $\vee a_1, a_2, ..., a_n$ is the disjunction of positive or negative atoms $a_i$. If both an atom and its negation appear, the clause is a tautology. A contingent clause is a clause which is neither empty nor a tautology (Rock, 2010).
Definition 1. A plausible description of a situation is a tuple $PD = (Ax, R_p, R_d, \triangleright)$, where $Ax$ is a set of contingent clauses, called axioms, characterising the aspects of the situation that are certain, $R_p$ is a set of plausible rules, $R_d$ is a set of defeater rules, and $\triangleright$ is a priority relation on $R_p \cup R_d$.

A plausible theory is computed from a plausible description by deriving the set $R$, of strict rules from the definite facts $Ax$. Thus, a plausible knowledge base consists of strict rules ($\rightarrow$), plausible rules ($\Rightarrow$), defeater (warning) rules ($\dashv$), and a priority relation on the rules ($\triangleright$). Strict rules are rules in the classical theory and a proof function $P$ is defined as $\lambda f . \mu \alpha. \{ \alpha \cap \beta, \gamma \}$, one is monotonic and four are non-monotonic: $\mu$ monotonic, strict, like classical logic; $\alpha = \beta \land \pi, \pi$ plausible, propagating ambiguity; $\beta$ plausible, blocking ambiguity; and $\gamma = \pi \lor \beta$.

2.2 Translating from DL to Plausible Logic Programs

Facing the challenge to reason on huge amount of noise and heterogeneous data, the OWL fragment corresponding to Horn clauses, known as Description Logic Programs (Grosif et al., 2003), can be a suitable choice. This section exploits the work in (Gomez et al., 2010) in order to translate description logic based ontologies into plausible logic axioms.

Conjunctions and universal restrictions in the right hand side of inclusion axioms are converted into rule heads ($L_0$ classes), whilst conjunction, disjunction and existential restriction appearing in the left-hand side are translated into rule bodies ($L_0$ classes). Figure 1 presents the mapping function $T$ from DL to strict rules in a plausible knowledge base, where $A$, $C$, and $D$ are concepts such that $A \in L_0$, $D \in L_0$, $A$ is an atomic concept, $X$, $Y$, $Z$ variables, and $P$, $Q$ roles.

3. DATA STREAM MANAGEMENT SYSTEM IN HASKELL

This section details the system architecture, as depicted in figure 2. The user is responsible to define the priorities and the plausible rules in order to handle contradictory data for the problem in hand.

3.1 The Haskell Platform

The advantages which Haskell brings in this landscape, lazy evaluation and implicit parallelism, are significant features when dealing with huge data streams which are parallel in nature. The parallel performing of reasoning tasks is of significant importance in order to provide answers in due time (Valle et al., 2009). The Haskell’s polymorphism allows to write generic code to process streams, which is particularly useful due to the different exploitation of the same data stream. The absence of side effects, means that the order of expression evaluation is of no importance, which is extremely desirable in the context of data streams coming from different sources. One challenge when answering in real-time to many continuously queries is query optimisation. Allowing equational reasoning, can be exploited for automatic program and query optimisation. A premise specified as lazy is matched only when its variables are antici-
The continuous semantics of data streams assumes that: i) streams are volatile - they are consumed on the fly and not stored forever; and ii) continuous processing - queries are registered and produce answers continuously (Barbieri et al., 2010). In our case, the rules are triggered continuously in order to produce streams of consequents. The stateless feature of pure Haskell facilitates the conceptual model of networks of stream reasoners as envisaged in (Stuckenschmidt et al., 2010), where data is processed on the fly, without being stored. The lazy evaluation in Haskell provides answers perpetually, when the queries are executed against infinite streams. One does not have to specify the timesteps when the query should be executed. By default, the tuples are consumed when they become available, and only in case they contribute to a query answer.

The computational efficiency is supported by the fact that a function is not forced to wait for a data to arrive - the possible computation are executed instead. Moreover one can use the about-to-come data by borrowing it from the future, as long as no function tries to change its value. The non strict semantics of Haskell, allows the functions to not produce errors in case these errors can be avoided. Consequently, some noise data can be avoided, without disturbing the computations.

There is no constraint on the nature of data fed by a stream. The functions can be applied on RDF streams as follows: A triple object is created by the triple function

\[
data\text{Triple} = \text{triple} !\text{Node} !\text{Node} !\text{Node}
\]
\[
\text{triple} :: \text{Subject} \rightarrow \text{Predicate} \rightarrow \text{Object} \rightarrow \text{Triple}
\]

**Definition 2.** An RDF stream is an infinite list of tuples of the form \((\text{subj}, \text{pred}, \text{obj})\) annotated with their timestamps \(\tau\).

\[
\text{type RDFStream} = \left\{((\text{subj}, \text{pred}, \text{obj}), \tau)\right\}
\]

**Example 1.** An RDF stream of auction bids states the bidder agent, its action, and the bid value:

\[
[(\text{a}_1, \text{sell} , 30E), 14.32), (\text{a}_2, \text{sell} , 28E), 14.34), (\text{a}_3, \text{buy} , 26E), 14.35), (\text{a}_4, \text{sell} , 27E), 14.36)]
\]

### 3.2 Streams Module

Table 1 illustrates the operators provided by Haskell to manipulate infinite streams. Considering one wants to add the corresponding values from two financial data streams \(s_1\) and \(s_2\), expressed by two different currencies:

\[
\text{zipWith } + \text{ s}_1 (\text{map conversion } s_2)
\]

where the conversion function is applied on each element from \(s_2\). For computing at each step the sum of a string of transactional data, the following expression can be used: \(\text{scan } + \text{ 0 [2, 4, 5, 3, ...]}\), providing as output the infinite stream \([0, 2, 6, 11, 14, ...]\).

The aggregation of two streams takes place according to an aggregation policy, depending on the time or the configuration of the new tuples. Here, the policy is a function provided as input argument for the high order function \(\text{zipWith policy stream stream}\). Similarly, generating new stream is done based on a policy. The incoming streams can be dynamically split into two streams, based on a predicate \(p\).

### 3.3 The Mapping Module

The ontologies are translated based on the conceptual instrumentation introduced in section 2.2. Two sources of knowledge are exploited to reason on data
### Table 1: Stream operators in Haskell (S stands for the Stream datatype).

<table>
<thead>
<tr>
<th>Type</th>
<th>Function</th>
<th>Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>constructor, extract first element, takes a stream and returns all its</td>
<td>map :: (a -&gt; b) -&gt; s a -&gt; s b</td>
</tr>
<tr>
<td></td>
<td>prefixes, takes a stream and returns all its suffixes</td>
<td>inter :: Stream a -&gt; Stream a -&gt; s a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>scan :: (a -&gt; b -&gt; a) -&gt; a -&gt; s b -&gt; s a</td>
</tr>
<tr>
<td>Transformation</td>
<td>applies a function over all elements, interleaves 2 streams, yields a</td>
<td>transp :: s (s a) -&gt; s (s a)</td>
</tr>
<tr>
<td></td>
<td>stream of successive reduced values, computes the transposition of a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>stream of streams</td>
<td></td>
</tr>
<tr>
<td>Building</td>
<td>repeated applications of a function</td>
<td>iterate :: (a -&gt; a) -&gt; a -&gt; s a</td>
</tr>
<tr>
<td>streams</td>
<td></td>
<td>repeat :: a -&gt; s a</td>
</tr>
<tr>
<td></td>
<td>returns the infinite repetition of a set of values</td>
<td>cycle :: [a] -&gt; s a</td>
</tr>
<tr>
<td>Extracting</td>
<td>takes the first elements, drops the first elements, returns the longest</td>
<td>take :: Int -&gt; [a] -&gt; s a</td>
</tr>
<tr>
<td>sublists</td>
<td>prefix for which the predicate p holds</td>
<td>drop :: Int -&gt; s a -&gt; s a</td>
</tr>
<tr>
<td></td>
<td>returns the suffix remaining after takeWhile</td>
<td>dropWhile :: (a -&gt; Bool) -&gt; s a -&gt; [a]</td>
</tr>
<tr>
<td></td>
<td>removes elements that do not satisfy p</td>
<td>filter :: a -&gt; [a] -&gt; s a</td>
</tr>
<tr>
<td>Index</td>
<td>return the element of the stream at index n</td>
<td>elemIndex :: Eq a -&gt; a -&gt; s a -&gt; Int</td>
</tr>
<tr>
<td></td>
<td>return the index of the first element equal to the query element</td>
<td>findIndex :: (a -&gt; Bool) -&gt; s a -&gt; Int</td>
</tr>
<tr>
<td></td>
<td>return the index of the first element satisfying p</td>
<td></td>
</tr>
<tr>
<td>Aggregation</td>
<td>return a list of corresponding pairs from 2 streams</td>
<td>zip :: s a -&gt; s b -&gt; s (a,b)</td>
</tr>
<tr>
<td></td>
<td>combine two streams based on a given function</td>
<td>ZipWith :: (a -&gt; b z) -&gt; c) -&gt; [a,b] -&gt; s (a,b)</td>
</tr>
</tbody>
</table>

```
Sensor ⊑ ∀measure.PhysicalQuality
Sensor ⊑ ∀hasLatency.Time
Sensor ⊑ ∀hasLocation.Location
Sensor ⊑ ∀hasFrequency.Frequency
Sensor ⊑ ∀hasAccuracy.MeasureUnit
WirelessSensor ⊑ Sensor
RFIDSensor ⊑ WirelessSensor
ActiveRFID ⊑ RFIDSensor
Sensor(X), Measures(X,Y) → PhysicalQuality(Y)
Sensor(X), HasLatency(X,Y) → Time(Y)
Sensor(X), HasLocation(X,Y) → Location(Y)
Sensor(X), HasFrequency(X,Y) → Frequency(Y)
Sensor(X), HasAccuracy(X,Y) → MeasureUnit(Y)
WirelessSensor(X) → Sensor(X)
RFIDSensor(X) → WirelessSensor(X)
ActiveRFID(X) → WirelessSensor(X)
```

Figure 3: Translating the sensor ontology.

The main advantage consists in the possibility to dynamically include new background knowledge in the system.

### 3.4 Efficiency

The system incorporates the Decisive Plausible Logic tool\(^1\). A Haskell glue module that exports functions requesting proofs (Rock, 2010) is used to make the connection with the other modules. The efficiency is mandatory when one needs to reason on huge data in real time. The efficiency of the proposed solution is based on the following vectors: i) The implementation of a family of defeasible logic is polynomial (Maher et al., 2001). Plausible logic being a particular case of defeasible reasoning belongs to this efficiency class. The possibility to select the current inference algorithm among \(\mu, \alpha, \pi, \beta, \gamma\) can be exploited to adjust the reasoning task to the complexity of problem in hand. ii) DLP are subfragments of Horn logics and their complexity is polynomial, as reported in (Krötzsch et al., 2007).

Milk \sqsubseteq \text{Item}

Item \forall \text{HasPeak.Time}

WholeMilk \sqsubseteq \text{Milk}

LowFatMilk \sqsubseteq \text{Milk}

f_{m1} : \text{WholeMilk}

s_{m1} : \text{LowFatMilk}

s_{m2} : \text{LowFatMilk}

Figure 4: Domain Knowledge for the Milk Monitoring.

\begin{align*}
  r_1 & : \text{Milk}(X) \rightarrow \text{Item}(X) \\
  r_2 & : \text{Item}(X), \text{HasPeak}(X,Y) \rightarrow \text{Time}(Y) \\
  r_3 & : \text{WholeMilk}(X) \rightarrow \text{Milk}(X) \\
  r_4 & : \text{LowFatMilk}(X) \rightarrow \text{Milk}(X) \\
  f_1 & : \text{WholeMilk}(fm1) \\
  f_2 & : \text{LowFatMilk}(sm1) \\
  f_3 & : \text{LowFatMilk}(sm2) \\
  r_{10} & : \text{Milk}(X), \text{Stock}(X,Y), \text{Less}(Y,c1) \Rightarrow \\
  & \text{NormalSupply}(X,c2) \\
  r_{11} & : \text{HasPeak}(X,Y) \rightarrow \text{NormalSupply}(X,c2) \\
  r_{12} & : \text{Milk}(X), \text{Stock}(X,Y), \text{Less}(Y,c1), \\
  & \text{hasPeak}(X,Z), \text{now}(Z) \Rightarrow \text{PeakSupply}(X,c3) \\
  r_{13} & : \text{AlterativeItem}(X,Z), \text{Milk}(X), \text{Stock}(Z,Y), \\
  & \text{Greater}(Y,c4) \Rightarrow \text{~PeakSupply}(X,c3) \\
  r_{14} & : \text{LastMeasurement}(S,Y), \text{HasLatency}(S,Z), \\
  & \text{Greater}(Y,Z) \Rightarrow \text{BrokenSensor}(S) \\
  r_{15} & : \text{BrokenSensor}(S), \text{Measure}(S,X) \rightarrow \text{Stock}(X,c) \\
  r_{13} & \succ r_{12}
\end{align*}

Figure 5: Plausible Knowledge Base.

4 RUNNING SCENARIO

The scenario regards supporting real-time supply chain decisions based on RFID streams. Consider the stock management of a retailer. RFID sensors are used to count the items entering on the shelves from two locations. The clients leave the supermarket from three payment points, corresponding to three output streams. Monitoring an item like Milk implies monitoring several subcategories like WholeMilk and LowFatMilk. The retailer sells a specific item fm1 of whole milk, and two types of low fat milk sm1 and sm2. Some peak periods are associated to each commercialised item. This background knowledge is formalised in figure 4. The corresponding strict rules are depicted in the upper part of the figure 5. During peak periods for an item the usual supply action is blocked by the defeater r_{11}.

The plausible rule r_{10} says that if the milk stock Y is below the alert threshold c1, the normalSupply action should be executed. NormalSupply assures a stock value of c2. Instead, the PeakSupply action is derived by the rule r_{11}.

If there is an alternative item Z for the Milk product and the stock of the alternative is larger than the threshold c4, this implies not to supply the higher quantity c2 (the rule r_{12}). Depending on the priority relation between the rules r_{12} and r_{13}, the action is executed or not.

The sensor related information can be integrating when reasoning. If the sensor S seems not to function according to the specifications in the ontology, it is plausible to be broken (the rule r_{14}). A broken sensor defeats the stock information asserted in the knowledge base related to the measured item (the defeater r_{18}).

The merchandise flow is simulated by generating infinite input and output streams. Assuming that the function randomItem :: [Item] -> Item, based on the list of available items returns a random item. The output stream for the payment point out_1 would be:

```
out_1 = (randomItem l) : out_1
```

where l represents the available items in the simulation. Assume a stream of sold items and the time of measurement s_1 = [(sm1,1), (m1,2), (fm1,3), (m2,4), (m3,5), (sm2,6), (m4,7), ...].

The updateStock function continuously computes the current stocks based on the s_1 stream. Based on the fact f_1 and the rule r_{13}, one can conclude that fm1 is a milk item. Similarly, based on the facts f_2 and f_3, the rule r_4 categorises the items sm1 and sm2 as milk items. The filter function is used to monitor each milk item, either low fat or not:

```
milkItems = filter milk (map first s_1)
```

Here, the predicate milk returns true if the input is of type Milk according to the rules r_3 or r_4. The map function is used to select only the first element from the tuples (item, time) from the stream s_1. The stream milkItems collects all the items of type milk, and everytime an item occurs, the updateStock :: Item -> Stream -> Int function is activated to compute the available stock for a specific category. Thus, by combining ontological knowledge with plausible rules one can reason with generic products (Milk), even if the streams report data regarding instances of specific products (WholeMilk and LowFatMilk), minimising the number of business rules that should be added within the system.

5 DISCUSSION AND RELATED WORK

Stream integration is considered an ongoing challenge for the stream management systems (Valle
et al., 2009; Calbimonte et al., 2010; Le-Phuoc et al., 2010; Palopoli et al., 2003). There are several tools available to perform stream reasoning.

DyKnow (Fredrik Heintz and Doherty, 2009) introduces the knowledge processing language KPL to specify knowledge processing applications on streams. We exploit the Haskell stream operators to handle streams and list comprehension for querying these streams. The SPARQL algebra is extended in (Bolles et al., 2008) with time windows and pattern matching for stream processing. In our approach we exploit the existing list comprehension and pattern matching in Haskell, aiming at the same goal of RDF streams processing. Comparing to C-SPARQL, Haskell provides capabilities to aggregate streams before querying them. E talis tool performs reasoning tasks over streaming events with respect to background knowledge (Anicic et al., 2010). In our case the background knowledge is obtained from ontologies; translated as strict rules in order to reason over a unified space.

The research conducted here can be integrated into the larger context of Semantic Sensor Web, where challenges like abstraction level, data fusion, application development (Corcho and Garcia-Castro, 2010) are addressed by several research projects like Asp i re2 or Sensei3. By encapsulating domain knowledge as description logic programs, the level of abstraction can be adapted for the application in hand by importing a more refined ontology into DLP.

Streams being approximate, omniscient rationality is not assumed when performing reasoning tasks on streams. Consequently, we argue that plausible reasoning for real time decision making is adequate. One particularity of our system consists of applying an efficient non-monotonic rule based system (Maher et al., 2001) when reasoning on gradually occurring stream data. The inference is based on several algorithms, which is in line with the proof layers defined in the Semantic Web. Moreover, all the Haskell language is available to extend or adapt the existing code. The efficiency of data driven computation in functional reactive programming is supported by the lazy evaluation mechanism which allows to use values before they can be known.

The strength of plausibility of the consequents is given by the superiority relation among rules. One idea of computing the degree of plausibility is to exploit specific plausible reasoning patterns like epagoge: "If A is true, then B is true. Therefore, A becomes more plausible." If A is true, then B becomes more plausible. B is true. Therefore, A becomes more plausible.”

6 CONCLUSIONS

Our semantic based stream management system is characterised by: i) continuous situation awareness and capability to handle theoretically infinite data streams due to the lazy evaluation mechanism, ii) aggregating heterogeneous sensors based on the ontologies translated as strict rules, iii) handling noise and contradictory information inherently in the context of many sensors, due to the plausible reasoning mechanism. Ongoing work regards conducting experiments to test the efficiency and scalability of the proposed framework, based on the results reported in (Maher et al., 2001) and on the reduced complexity of description logic programs (Krotzsch et al., 2007).

ACKNOWLEDGEMENTS

We are grateful to the anonymous reviewers for their useful comments. The work has been co-funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/89/1.5/S/62557 and PN-11-Ideas-170.

REFERENCES


2http://www.fp7-aspire.eu/
3http://www.ict-sensei.org/


