Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

Scalable Maintenance of Knowledge Discovery in an Ontology Stream

Freddy Lécué
IBM Research - Ireland
freddy.lecue@ie.ibm.com

Abstract

In dynamic settings where data is exposed by streams, knowledge discovery aims at learning associations of data across streams. In the semantic Web, streams expose their meaning through evolutive versions of ontologies. Such settings pose challenges of scalability for discovering (a posteriori) knowledge. In our work, the semantics, identifying knowledge similarity and rarity in streams, together with incremental, approximate maintenance, control scalability while preserving accuracy of streams associations (as semantic rules) discovery.

1 Introduction and Related Work

In the semantic Web, the meaning of data streams is represented as ontology streams [Huang and Stuckenschmidt, 2005]. Such streams are dynamic and evolutive versions of ontologies where OWL (Web Ontology Language), which is underpinned by Description Logics (DL) [Baader *et al.*, 2003], is used as a rich description language. From knowledge materialization [Barbieri *et al.*, 2010], to diagnosis or predictive reasoning [Lécué and Pan, 2013], all are inferences where dynamics, semantics of data are exploited for deriving a priori knowledge from pre-established (certain) statements.

From a data-driven perspective, knowledge can be gained a posteriori from uncertainty by learning probabilistic rules or associations across streams [Agrawal *et al.*, 1993]. When a priori and a posteriori knowledge are reunified, the problem of knowledge discovery can be revisited as the problem of discovering stochastic rules with semantic representation of their premises, conclusions [Lécué and Pan, 2015]. In a dynamic context, with highly changing data, it is crucial to ensure scalable knowledge discovery by adjusting and updating all learnt semantic rules rather than re-elaborating potential associations from scratch (exponential in the size of data).

Most of techniques in Database e.g., [Cheung *et al.*, 1996], adapting `Apriori` [Agrawal *et al.*, 1996] for streams, focus on syntactic representation of data to iteratively identify frequent associations. [Lee *et al.*, 2003] improved its scalability by partitioning all streams using a sliding-window filtering. Approaches in Machine Learning e.g., [Gama and Kosina, 2011] focus on learning decision rules, as a subset of association rules, for classifying data from streams in real-time. Rules are incrementally learnt through approximate specialization. Although such models fits (raw) data streams exposing (syntactic) numeric/symbolic values, they could not cope with semantics captured by ontology streams. Indeed they all fail in interpreting the underlying semantics of data, making knowledge discovery highly subject to changes, and not necessarily accurate, scalable. Facing these limitations, techniques from Knowledge Representation and Reasoning explored ontology-based learning for generating DL axioms [Völker and Niepert, 2011], Horn rules [Galárraga *et al.*, 2013], DL rules [Lécué and Pan, 2015] from semantic data. All rules are mined from scratch for each update, which is ineffective since it (i) restricts the dynamics of (stream) learning, (ii) neglects previously discovered rules, (iii) intolerably limits its scalability, adaptability and reactivity.

We address "*scalable maintenance of knowledge discovery in dynamic semantic data*". Given ontology streams, how to ensure scalable identification of dynamic, knowledge associations? This poses problems of dynamically maintaining associations and their interestingness up-to-date in a window sliding over streams. Key contributions include: (1) By exploiting the semantics of ontology streams, we introduce the notions of knowledge similarity, rarity over time, which are logically based and do not only rely on the syntax of axioms. They enable approximation of knowledge discovery while maintaining accuracy. (2) We design the first incremental algorithm which controls the scalability of learning by efficiently maintaining DL rules, as associations across streams.

Next section reviews the adopted logic, ontology stream together with rules representation. In Section 3 we study knowledge similarity and rarity in streams. Section 4 presents how knowledge discovery is incrementally maintained. Finally, we report experimental results on scalability and accuracy with data from Dublin City and draw some conclusions.

2 Background

The semantics of data is represented using an ontology. We focus on DL to define ontologies since this logic offers good reasoning support for most of its expressive families and compatibility to W3C standards e.g., OWL 2. Since our work requires DLs supporting polynomial time reasoning when combined with rules, $\mathcal{EL}^{++}$ [Baader *et al.*, 2005] will be considered for illustration. We review (i) DL basics of $\mathcal{EL}^{++}$, (ii) ontology stream, (iii) $\mathcal{EL}^{++}$ atomsets and association rules.

$SocialEvent \sqcap \exists type.Music \sqsubseteq Event \sqcap \exists disruption.High$ (1)					
$Incident \sqcap \exists impact.Serious \sqsubseteq Event \sqcap \exists disruption.High$ (2)					
$Road \sqcap \exists adj.(\exists occur.\exists disruption.High) \sqsubseteq DisruptedRoad$ (3)					
$BusRoad \sqcap \exists travel.Long \sqsubseteq Road \sqcap \exists with.CongestedBus$ (4)					
$Road \sqcap \exists with.Bus \sqsubseteq BusRoad$	(5)			$Road(r_0)$	(6)
$Road(r_1)$	(7)			$Stop \sqsubseteq Long \sqsubseteq Abnormal$	(8)
$Road(r_2)$	(9)	$Bus(b_1)$	(10)	$Bus(b_2)$	(11)
$Bus(b_3)$	(12)	$adj(r_0, r_1)$	(13)	$adj(r_0, r_2)$	(14)

Figure 1: $\mathcal{O} \doteq \langle \mathcal{T}, \mathcal{A} \rangle$. Sample of TBox $\mathcal{T}$ and ABox $\mathcal{A}$.

2.1 Description Logics $\mathcal{EL}^{++}$

A signature Σ, noted $(\mathcal{N}_C, \mathcal{N}_R, \mathcal{N}_I)$ consists of 3 disjoint sets of (i) atomic concepts $\mathcal{N}_C$, (ii) atomic roles $\mathcal{N}_R$, and (iii) individuals $\mathcal{N}_I$. Given a signature, the top concept $\top$, the bottom concept $\bot$, an atomic concept A, an individual a, an atomic role expression r, $\mathcal{EL}^{++}$ concept expressions C and D in $\mathcal{C}$ can be composed with the following constructs:

$$\top \mid \bot \mid A \mid C \sqcap D \mid \exists r.C \mid \{a\}$$

The DL ontology $\mathcal{O} \doteq \langle \mathcal{T}, \mathcal{A} \rangle$ is composed of TBox $\mathcal{T}$, ABox $\mathcal{A}$. A TBox is a set of concept, role axioms. $\mathcal{EL}^{++}$ supports General Concept Inclusion axioms (GCIs e.g. $C \sqsubseteq D$), Role Inclusion axioms (RIs e.g., $r \sqsubseteq s$). An ABox is a set of concept assertion axioms e.g., $C(a)$, role assertion axioms e.g., $R(a, b)$, individual in/equality axioms e.g., $a \neq b$, $a = b$.

Example 1. ***(TBox and ABox Concept Assertion Axioms)*** *Figure 1 presents (i) a TBox $\mathcal{T}$ where $DisruptedRoad$ (**??**) denotes the concept of "roads which are adjacent to an event causing high disruption", (ii) concept assertions (**??**-**??**) denoting the individual r_0 having $r_{i,1 \leq i \leq 2}$ as adjunct roads.*

All completion rules, which are used to classify $\mathcal{EL}^{++}$ TBox $\mathcal{T}$ and entail subsumption, are described in [Baader *et al.*, 2005]. Reasoning with such rules is PTime-Complete.

2.2 Ontology Stream and its Evolution

We represent knowledge evolution by a dynamic, evolutive version of ontologies [Huang and Stuckenschmidt, 2005]. Data (ABox), its inferred statements (entailments) are evolving over time while its schema (TBox) remains unchanged.

Definition 1. ***(DL $\mathcal{L}$ Ontology Stream)*** *A DL $\mathcal{L}$ ontology stream $\mathcal{P}_m^n$ from point of time m to point of time n is a sequence of (sets of) Abox axioms $(\mathcal{P}_m^n(m), \mathcal{P}_m^n(m+1), \cdots, \mathcal{P}_m^n(n))$ with respect to a static TBox $\mathcal{T}$ in a DL $\mathcal{L}$ where $m, n \in \mathbb{N}$ and $m < n$.*

$\mathcal{P}_m^n(i)$ is a snapshot of an ontology stream (stream for short) $\mathcal{P}_m^n$ at time i, referring to ABox axioms with respect to a TBox in $\mathcal{L}$. We denote by $\mathcal{P}_m^n[i, j]$ i.e., $\bigcup_{k=i}^{j} \mathcal{P}_m^n(k)$ a windowed stream of $\mathcal{P}_m^n$ between time i and j with $i < j$. All windows $[i, j]$ have fixed length. We consider streams $\mathcal{P}_0^n$ with $[\alpha] \doteq [i, j]$, $[\beta] \doteq [k, l]$ as windows in $[0, n]$ and $i < k$.

Example 2. ***(DL $\mathcal{EL}^{++}$ Ontology Stream)*** *Figure 2 illustrates $\mathcal{EL}^{++}$ streams $\mathcal{P}_0^9$, $\mathcal{Q}_0^9$, $\mathcal{R}_0^9$, related to events, travel time, buses, through snapshots at time $i \in \{5, 6, 7\}$ (i.e., a view on window $[5, 7]$). Their dynamic knowledge is captured by evolutive ABox axioms e.g., (**??**) captures e_1 as "a social music event occurring in r_1" at time 5 of $\mathcal{P}_0^9$.*

$\mathcal{P}_0^9(5) : (SocialEvent \sqcap \exists type.Music)(e_1),\ \ occur(r_1, e_1)$	(18)
$\mathcal{Q}_0^9(5) : (Road \sqcap \exists travel.Abnormal)(r_1)$	(19)
$\mathcal{R}_0^9(5) : with(r_1, b_1)$	(20)
$\mathcal{P}_0^9(6) : (Event \sqcap \exists disruption.High)(e_2),\ \ occur(r_2, e_2)$	(21)
$\mathcal{Q}_0^9(6) : (Road \sqcap \exists travel.Long)(r_2)$	(22)
$\mathcal{R}_0^9(6) : with(r_2, b_2)$	(23)
$\mathcal{P}_0^9(7) : (Incident \sqcap \exists impact.Serious)(e_3),\ \ occur(r_2, e_3)$	(24)
$\mathcal{Q}_0^9(7) : (Road \sqcap \exists travel.Stop)(r_2)$	(25)
$\mathcal{R}_0^9(7) : with(r_2, b_3)$	(26)

Figure 2: Ontology Streams $\mathcal{P}_0^9(i), \mathcal{Q}_0^9(i), \mathcal{R}_0^9(i)_{i \in \{5,6,7\}}$.

Windowed Stream Changes	$(\mathcal{Q} \cup \mathcal{R})_0^9[6, 7] \nabla (\mathcal{Q} \cup \mathcal{R})_0^9[5, 6]$		
	new	*invariant*	*obsolete*
$\exists travel.Abnormal(r_1)$			✓
$\exists with.CongestedBus(r_2)$		✓	
$\exists travel.Stop(r_2)$	✓		

Table 1: ABox Entailment-based Stream Changes.

By applying completion rules on static knowledge $\mathcal{T}$ and ontology streams $\mathcal{P}_0^n$, snapshot-specific axioms are inferred.

The evolution of a stream is captured along its changes i.e., *new*, *obsolete* and *invariant* ABox entailments from one windowed stream to another one in Definition 2.

Definition 2. ***(ABox Entailment-based Stream Changes)*** *Let $\mathcal{S}_0^n$ be a stream; $[\alpha]$, $[\beta]$ be windows in $[0, n]$; $\mathcal{T}$ be axioms, $\mathcal{G}$ its ABox entailments. The changes occurring from $\mathcal{S}_0^n[\alpha]$ to $\mathcal{S}_0^n[\beta]$, denoted by $\mathcal{S}_0^n[\beta] \nabla \mathcal{S}_0^n[\alpha]$, are ABox entailments in $\mathcal{G}$ being new (1), $obsolete$ (2), $invariant$ (3).*

$$\mathcal{G}_{new}^{[\alpha],[\beta]} \doteq \{g \in \mathcal{G} \mid \mathcal{T} \cup \mathcal{S}_0^n[\beta] \models g \ \wedge \mathcal{T} \cup \mathcal{S}_0^n[\alpha] \not\models g\} \quad (1)$$

$$\mathcal{G}_{obs}^{[\alpha],[\beta]} \doteq \{g \in \mathcal{G} \mid \mathcal{T} \cup \mathcal{S}_0^n[\beta] \not\models g \ \wedge \mathcal{T} \cup \mathcal{S}_0^n[\alpha] \models g\} \quad (2)$$

$$\mathcal{G}_{inv}^{[\alpha],[\beta]} \doteq \{g \in \mathcal{G} \mid \mathcal{T} \cup \mathcal{S}_0^n[\beta] \models g \ \wedge \mathcal{T} \cup \mathcal{S}_0^n[\alpha] \models g\} \quad (3)$$

(1) reflects knowledge we gain by sliding window from $[\alpha]$ to $[\beta]$ while (2) and (3) denote respectively lost and stability of knowledge. All duplicates are supposed removed. Definition 2 provides basics, through ABox entailments, for understanding how knowledge is changing among windows.

Example 3. ***(ABox Entailment-based Stream Changes)*** *Table 1 illustrates changes occurring from $(\mathcal{Q} \cup \mathcal{R})_0^9[5, 6]$ to $(\mathcal{Q} \cup \mathcal{R})_0^9[6, 7]$ through ABox entailements. For instance "r_2 as a road with (at least) one congested bus" in window $[6, 7]$ of $(\mathcal{Q} \cup \mathcal{R})_0^9$ is invariant with respect to knowledge in $[5, 6]$. It is entailed using DL completion rules on (**??**), (**??**) and (**??**).*

2.3 $\mathcal{EL}^{++}$ Atomsets and Association Rule

We consider $\mathcal{EL}^{++}$ with (i) concept expressions $\mathcal{C}$, role names $\mathcal{N}_R$, individual names $\mathcal{N}_I$, and (ii) a countable set of first-order variables $\mathcal{V}$.

Atomset: Given terms $x_1, x_2 \in \mathcal{V} \cup \mathcal{N}_I$, a concept (role) atom is a formula $C(x_1)$ $(R(x_1, x_2))$ with $C \in \mathcal{C}$ $(R \in \mathcal{N}_R)$. We use finite sets (*atomsets*) $\mathbb{B}$ of (concepts, roles) atoms for representing conjunction $\forall \vec{x}. \bigwedge B$ where $\vec{x} \doteq x_1, \cdots, x_n \in \mathcal{V}$ are variables of atoms $B \in \mathbb{B}$ which could be shared. $\xi(\mathcal{A})$ denotes the set of all atomsets generated from atoms in $\mathcal{A}$ w.r.t $\mathcal{T}$. By abuse of notation, we continue to write A for $\xi(\mathcal{A})$.

Atomset Binding: Atomsets can be seen as conjunctive queries [Glimm *et al.*, 2007] without non-distinguished variables. We write $\mathcal{T}, \mathcal{A} \models \mathbb{B}[\vec{a}]$ to denote that $\vec{a} \in \mathcal{N}_I$ is a binding (answer) to atomset (query) $\mathbb{B}$.

Example 4. ***(Atomset and Binding)***
r_0 is a binding to atomset $\{adj(x, r_1)\}$ w.r.t. $\mathcal{O}$ in Figure 1.

$\mathcal{EL}^{++}$ **Rules** [Krötzsch *et al.*, 2008] extends $\mathcal{EL}^{++}$ expressivity with rules while maintaining polynomial time reasoning. Given atomsets $\mathbb{B}$, $\mathbb{H}$, and all variables $\vec{x} \in \mathcal{V}$ of atomset $\mathbb{B} \cup \mathbb{H}$, an $\mathcal{EL}^{++}$ *rule* is a formula $\mathbb{B} \twoheadrightarrow \mathbb{H}$, such that $\mathbb{B}$ is cycle free and does not contain atom of the form $R(x, x)$.

Example 5. ***($\mathcal{EL}^{++}$ Rule)***
Below rule denotes "if x_3 is adjacent to a x_2 where a highly disruptive event x_1 occurs then buses are congested in x_3".

$$(Event \sqcap \exists disruption.High)(x_1) \wedge \quad (4)$$

$$occur(x_2, x_1) \ \wedge \ adj(x_3, x_2) \quad (5)$$

$$\twoheadrightarrow (\exists with.CongestedBus)(x_3) \quad (6)$$

Association $\mathcal{EL}^{++}$ Rules [Lécué and Pan, 2015] are $\mathcal{EL}^{++}$ rules defined across some associations of axioms in streams. They extend Database association rules [Agrawal *et al.*, 1993] as knowledge from different streams is associated and combined by learning rules. Such rules are modeled in $\mathcal{EL}^{++}$.

Example 6. ***(Association $\mathcal{EL}^{++}$ Rule)***
$\{(4), (5)\} \twoheadrightarrow \{(6)\}$ is an association rule from $\mathcal{P}_0^9$ to $\mathcal{Q}_0^9 \cup \mathcal{R}_0^9$. (4-5) are defined in $\mathcal{P}_0^9$, (6) is inferred from $\mathcal{Q}_0^9 \cup \mathcal{R}_0^9$.

3 Significance of Knowledge Evolution

We introduce knowledge *similarity* and *rarity*, as basis for (i) measuring the significance of knowledge evolution in streams and (ii) controlling when to operate knowledge discovery.

3.1 Knowledge Similarity in An Ontology Stream

Definition 3 revisits stream correlation [Lécué and Pan, 2013] to capture knowledge similarity over windowed streams. It captures what knowledge, through ABox entailments, is common or modified (new or obsolete) over two windows.

Definition 3. ***(Knowledge Similarity in Ontology Stream)***
Let $\mathcal{S}_0^n$ be a stream; $[\alpha]$, $[\beta]$ be windows in $[0, n]$. The (symmetric) knowledge similarity between $\mathcal{S}_0^n[\alpha]$ and $\mathcal{S}_0^n[\beta]$ is:

$$\Phi(\mathcal{S}_0^n[\alpha], \mathcal{S}_0^n[\beta]) \doteq \frac{|\mathcal{G}_{inv}^{[\alpha],[\beta]}|}{|\mathcal{G}_{new}^{[\alpha],[\beta]}| + |\mathcal{G}_{inv}^{[\alpha],[\beta]}| + |\mathcal{G}_{obs}^{[\alpha],[\beta]}|} \quad (7)$$

where the expressions in between | refer to its cardinality i.e., the number of new, obsolete and invariant ABox entailments obtained from $\mathcal{S}_0^n[\alpha]$ to $\mathcal{S}_0^n[\beta]$ using DL completion rules.

(7) captures the knowledge similarity of windowed streams $\mathcal{S}_0^n[\alpha]$, $\mathcal{S}_0^n[\beta]$ in $[0, 1]$. The number of *invariant* ABox entailments emphasizes "common" knowledge while the number of *new* and *obsolete* ABox entailments is capturing differentiators in knowledge evolution from window $[\alpha]$ to $[\beta]$. Therefore the higher $\Phi(\mathcal{S}_0^n[\alpha], \mathcal{S}_0^n[\beta])$ the more similarity between knowledge captured in windows $[\alpha]$ and $[\beta]$ of $\mathcal{S}_0^n$.

Evaluating (7) is in worst case polynomial time with respect to acyclic $\mathcal{T}$ and $\mathcal{S}_0^n$ in $\mathcal{EL}^{++}$ since DL concepts expansion, unfolding and subsumption required from Definition 2 are all solvable in polynomial time [Baader *et al.*, 2005].

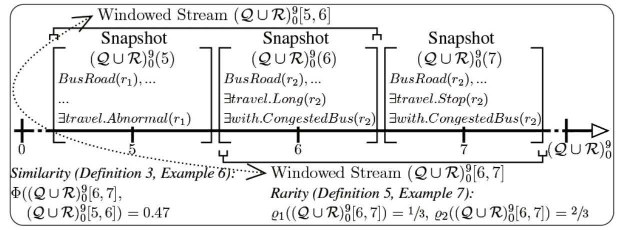

Figure 3: Knowledge Similarity and Rarity in $(\mathcal{Q} \cup \mathcal{R})_0^9$.

Example 7. ***(Knowledge Similarity in Ontology Stream)***
For $(\mathcal{Q} \cup \mathcal{R})_0^9[6, 7] \nabla (\mathcal{Q} \cup \mathcal{R})_0^9[5, 6]$, all entailments in $(\mathcal{Q} \cup \mathcal{R})_0^9[5, 6]$ and $(\mathcal{Q} \cup \mathcal{R})_0^9[6, 7]$, derived using DL completion rules, static knowledge (Figure 1), dynamic knowledge (Figure 2), are required. By applying their resulting numbers of 8 invariant, 4 new, 5 obsolete entailments in (7), knowledge similarity of $(\mathcal{Q} \cup \mathcal{R})_0^9$ in $[6, 7]$ with $[5, 6]$ is 0.47 cf. Figure 3.

3.2 Knowledge Rarity in An Ontology Stream

Knowledge rarity or the state of knowledge of being infrequent, is a strong indicator of (i) knowledge representativity and thus of (ii) significance for knowledge discovery. Definition 5 revisits the concept of rarity for ontology streams as the proportion of its rare knowledge (Definition 4) i.e., knowledge occurring over a pre-determined number of snapshots.

Definition 4. ***(ε-Rare Knowledge in Ontology Stream)***
Let $\mathcal{S}_0^n$ be a stream; $\mathcal{T}$, $\mathcal{A}$ be static axioms; $[\alpha]$ be a window in $[0, n]$. An ABox assertion a is ε-rare in $[\alpha]$ for integer ε if:

$$|\{i \in [\alpha] \mid \mathcal{T} \cup \mathcal{A} \cup \mathcal{S}_0^n(i) \models a\}| = \varepsilon \quad (8)$$

where i is an instant in $[\alpha]$ and $\mathcal{S}_0^n(i)$ is a snapshot of $\mathcal{S}_0^n[\alpha]$.

Assertion a is ε-rare if a can be entailed over ε snapshots of $[\alpha]$. We denote by ε-rare$(\mathcal{S}_0^n[\alpha])$ the set of ABox assertions which are ε-rare in $[\alpha]$ of $\mathcal{S}_0^n$, by $|\varepsilon$-rare$(\mathcal{S}_0^n[\alpha])|$ its cardinality. Let $|$distinct$(\mathcal{S}_0^n[\alpha])|$ be the number of distinct ABox assertions entailed in $[\alpha]$ of $\mathcal{S}_0^n$. Thus, knowledge rarity is characterized by the ratio of rare to distinct ABox assertions.

Definition 5. ***(ε-Rarity of Knowledge in Ontology Stream)***
Let $\mathcal{S}_0^n$ be a stream; $[\alpha]$ be a window in $[0, n]$ of $\mathcal{S}_0^n$. The ε-rarity of knowledge in $[\alpha]$ of $\mathcal{S}_0^n$ is defined as the ratio:

$$\varrho_\varepsilon(\mathcal{S}_0^n[\alpha]) \doteq \frac{|\varepsilon\text{-}rare(\mathcal{S}_0^n[\alpha])|}{|distinct(\mathcal{S}_0^n[\alpha])|} \quad (9)$$

1-rarity ϱ_1 is the fraction of knowledge appearing uniquely in $[\alpha]$. ε-rarity measures knowledge repeating ε times within $[\alpha]$. Evaluating (9) is in worst case polynomial time cf. (7).

Example 8. ***(ε-Rarity of Knowledge in Ontology Stream)***
Suppose $(\mathcal{Q} \cup \mathcal{R})_0^9$ over $[6, 7]$ in Figure 2. $\exists travel.Stop(r_2)$ is 1-rare while $\exists with.CongestedBus(r_2)$ is 2-rare as they are respectively entailed in snapshots $\{7\}$ and $\{6, 7\}$. Given the number of distinct, 1-rare and 2-rare knowledge being respectively 12, 4, 8 in $[6, 7]$, 1-rarity and 2-rarity of knowledge in $[6.7]$ of $(\mathcal{Q} \cup \mathcal{R})_0^9$ are respectively 1/3, 2/3 cf. Figure 3.

We will apply (8-9) with atomsets in $\mathcal{S}_0^n$ rather than its ABox assertions, as they are representative of DL rules.

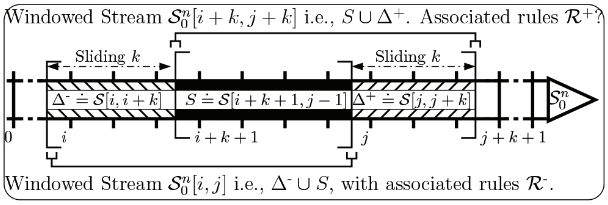

Figure 4: k-Sliding of Windowed Stream $\Delta^- \cup S$ to $S \cup \Delta^+$.

4 Incremental Maintenance of Knowledge

Scalable maintenance of knowledge in streams aims at keeping (discovered) association $\mathcal{EL}^{++}$ rules over a sliding window up-to-date rather than re-elaborating associations from scratch. Let $\mathcal{S}_0^n$ in Figure 4 be a set of streams that could be associated e.g., streams in Figure 2. Δ^-, Δ^+ are respectively knowledge removed and added by sliding from $\Delta^- \cup S$ to $S \cup \Delta^+$ while S captures what remains unchanged. The complexity of this problem is exponential in the number of atoms (to be composed by rules) in streams [Lécué and Pan, 2013]. We present some heuristics, illustrated in Algorithms 1, 2 and 3, that drastically reduce computation time in practice. Our approach is: (i) updating the interestingness of rules $\mathcal{R}^-$ discovered in $\Delta^- \cup S$, (ii) expanding rules $\mathcal{R}^-$ to cover incoming knowledge Δ^+, (iii) actioning maintenance only when accumulated knowledge update is significant. The approach benefits from $\mathcal{EL}^{++}$ as consistency checking of atomsets, together with atomset binding are polynomial [Bienvenu *et al.*, 2012].

4.1 Updating Interestingness of $\mathcal{EL}^{++}$ Rules

The interestingness of association $\mathcal{EL}^{++}$ rules is measured by the support of its atomsets and its confidence.

Definition 6. ***(Atomset Support)*** *Given axioms $\mathcal{O} \doteq \langle \mathcal{T}, \mathcal{A} \rangle$, stream $\mathcal{S}_0^n$, atomset $\mathbb{B}$. The support of $\mathbb{B}$, noted $\sigma(\mathbb{B})$, is defined by $|\{i \in [0, n] \mid \exists \vec{a} \in \mathcal{N}_I : \mathcal{O}, \mathcal{S}_0^n(i) \models \mathbb{B}[\vec{a}]\}|$ as the number of snapshots where $\mathbb{B}$ has at least a binding $\vec{a}$ in $\mathcal{A} \cup \mathcal{S}_0^n$ with respect to $\mathcal{T}$.*

We say that $\mathbb{B}$ is supported (or true) over $\sigma(\mathbb{B})$ snapshots in $[0, n]$ of $\mathcal{S}_0^n$. The support of any atomset in a stream can be computed as the sum of its supports on windowed streams.

Definition 7. ***(Confidence of an Association $\mathcal{EL}^{++}$ Rule)*** *Let $\rho : \mathbb{B} \twoheadrightarrow \mathbb{H}$ be an association $\mathcal{EL}^{++}$ rule in $\mathcal{S}_0^n$. The confidence γ of ρ, noted $\gamma(\rho)$, in $(0, 1]$ is defined by $\sigma(\mathbb{B} \cup \mathbb{H})/\sigma(\mathbb{B})$ i.e., the proportion of snapshots in $\mathcal{S}_0^n$ where $\mathbb{B} \cup \mathbb{H}$ has at least one binding with regard to those where $\mathbb{B}$ has a binding.*

The confidence is defined as the conditional probability of inferring $\mathbb{H}(\vec{a})$ given that we know $\mathbb{B}(\vec{a})$. We denote by $\sigma(\mathbb{B})|_{\mathcal{S}_0^n[\alpha]}$ and $\gamma(\rho)|_{\mathcal{S}_0^n[\alpha]}$ the support of atomset $\mathbb{B}$ and confidence of rule ρ in windowed stream $\mathcal{S}_0^n[\alpha]$.

Example 9. ***(Confidence of an Association $\mathcal{EL}^{++}$ Rule)*** *Let $\mathcal{S}_0^9$ be $(\mathcal{P} \cup \mathcal{Q} \cup \mathcal{R})_0^9$ in Figure 2. The confidence of rule $\gamma(\mathbb{B} \twoheadrightarrow \mathbb{H})$ with $\mathbb{B} : \{(4), (5)\}$, $\mathbb{H} : \{(6)\}$ in $[5, 6]$ of $\mathcal{S}_0^9$ is:*

$$\frac{\sigma(\mathbb{B} \cup \mathbb{H})|_{\mathcal{S}_0^9[5,6]}}{\sigma(\mathbb{B})|_{\mathcal{S}_0^9[5,6]}} = \frac{\sigma(\{(4), (5), (6)\})}{\sigma(\{(4), (5)\})} \text{ i.e., } \frac{1}{2}$$

$\mathbb{B} \twoheadrightarrow \mathbb{H}$ *is bound in* $1/2$ *of snapshots of* $S_0^9[5, 6]$.

Algorithm 1 (A1) updates interestingness of rules (learnt) in $\Delta^- \cup S$ to be suitable in S. All (line 5) are evaluated against Δ^- (line 7), the part removed by sliding from $\Delta^- \cup S$ to $S \cup \Delta^+$. If rules are applicable (lines 9-10), their support, confidence are updated. Otherwise Δ^- does not impact the interestingness of rules, which remain the same (line 11).

Remark 1. ***(Rule Interestingness when Adding Snapshots)*** *If snapshots Δ^+ are added to S, the interestingness is updated by applying* A1 *with parameters $\langle \mathcal{O}, \mathcal{S}_0^n, \Delta^+, S \cup \Delta^+ \rangle$, and upgrading (i) $-$ to $+$ in lines 9-10, (ii) $\Delta^- \cup S$ to S.*

Algorithm 1: `[A1]InterestUpdate`$\langle \mathcal{O}, \mathcal{S}_0^n, \Delta^-, S \rangle$

1 **Input**: Axioms $\mathcal{O} \doteq \langle \mathcal{T}, \mathcal{A} \rangle$, Streams $\Delta^- \cup S$ and S in $\mathcal{S}_0^n$.
2 **Result**: $\mathcal{R}$: Rules in $\Delta^- \cup S$ with support and confidence for S.
3 **begin**
4 $\mathcal{R} \leftarrow \emptyset$; *% Initialization of rules $\mathcal{R}$ in S.*
5 **foreach** $\rho : \mathbb{B} \twoheadrightarrow \mathbb{H}$ *in* $\Delta^- \cup S$ **do** *% Rules learnt in $\Delta^- \cup S$.*
6 *% Identification of rules in $\Delta^- \cup S$ valid in Δ^-.*
7 **if** $\exists \vec{a} \in \mathcal{N}_I \mid \mathcal{O}, \Delta^- \models (\mathbb{B} \cup \mathbb{H})[\vec{a}]$ **then**
8 *% Updated support σ, confidence γ of ρ in S*
9 $\sigma(\mathbb{B} \cup \mathbb{H}) \leftarrow \sigma(\mathbb{B} \cup \mathbb{H})|_{\Delta^- \cup S} - \sigma(\mathbb{B} \cup \mathbb{H})|_{\Delta^-}$;
10 $\gamma(\rho) \leftarrow \sigma(\mathbb{B} \cup \mathbb{H}) / (\sigma(\mathbb{B})|_{\Delta^- \cup S} - \sigma(\mathbb{B})|_{\Delta^-})$;
11 $\mathcal{R} \leftarrow \mathcal{R} \cup \{(\rho, \sigma(\mathbb{B} \cup \mathbb{H}), \gamma(\rho))\}$;
12 **return** $\mathcal{R}$;

Example 10. ***(Updating Interestingness of $\mathcal{EL}^{++}$ Rules)*** *By sliding from $[5, 6]$ to $[6, 7]$, the confidence of $\mathbb{B} \twoheadrightarrow \mathbb{H}$ in Example 9 requires an update. $\mathbb{B} \cup \mathbb{H}$ and $\mathbb{B}$ are supported at time $\{6, 7\}$, $\{5, 6, 7\}$. Thus, removing $\Delta^- : \mathcal{S}_0^9(5)$ from $\Delta^- \cup S : \mathcal{S}_0^9[5, 6]$ (lines 9-10) and adding $\Delta^+ : \mathcal{S}_0^9(7)$ to $S : \mathcal{S}_0^9(6)$ (lines 9-10 by applying Remark 1) update its confidence to* 1.

4.2 Approximating Expansion of $\mathcal{EL}^{++}$ Rules

Algorithm 2 (A2) expands rules in S (output of A1) to $S \cup \Delta^+$ (lines 5-10). Atomsets in S (line 5) are expanded with atomsets in Δ^+ if bindable (line 7). Such rules cover knowledge uniquely present in Δ^+, not in S. Only ε-rare atomsets, with ε parameterizable in $E \subseteq \{1, ..., |S|\}$, are considered (line 5) to limit the search space. This benefits scalability while not sacrificing significant rules (rules with minimum support, confidence) cf. Section 5. Finally, all potential rules in $S \cup \Delta^+$ are evaluated against the minimum confidence (lines 8-10).

Algorithm 2: `[A2]RuleExpansion`$\langle \mathcal{O}, \mathcal{S}_0^n, S, \Delta^+, E, \gamma_{\min} \rangle$

1 **Input**: (i) Axioms $\mathcal{O} \doteq \langle \mathcal{T}, \mathcal{A} \rangle$, (ii) Streams S, $S \cup \Delta^+$ in $\mathcal{S}_0^n$, (iii) Integer $\varepsilon \in E \subseteq \{1, ..., |S|\}$ with $|S|$: number of snapshots in S, (iv) Min. threshold of confidence $\gamma_{\min}$.
2 **Result**: $\mathcal{R}$: Rules expanding S with Δ^+ with min. thres. $\gamma_{\min}$.
3 **begin**
4 $\mathcal{R} \leftarrow \emptyset$; *% Initialization of expanded rules $\mathcal{R}$ in $S \cup \Delta^+$.*
5 **foreach** $\mathbb{B} \in \varepsilon$*-rare*$(S)$ **do** *% ε-rare Atomsets in S.*
6 *% Identification of an expansion of $\mathbb{B}$ in S with $\mathbb{H}$ in Δ^+.*
7 **if** $\exists \mathbb{H} \in \Delta^+, \mathbb{H} \notin S \mid \exists \vec{a} \in \mathcal{N}_I : \mathcal{O}, \Delta^+ \models (\mathbb{B} \cup \mathbb{H})[\vec{a}]$ **then** *% Rule $\mathbb{B} \twoheadrightarrow \mathbb{H}$ with min. confidence in $S \cup \Delta^+$.*
8 **if** $\sigma(\mathbb{B} \cup \mathbb{H})/\sigma(\mathbb{B}) > \gamma_{\min}$ **then** $\mathcal{R} \leftarrow \mathcal{R} \cup \{\mathbb{B} \twoheadrightarrow \mathbb{H}\}$;
9 *% Rule $\mathbb{H} \twoheadrightarrow \mathbb{B}$ with min. confidence in $S \cup \Delta^+$.*
10 **if** $\sigma(\mathbb{B} \cup \mathbb{H})/\sigma(\mathbb{H}) > \gamma_{\min}$ **then** $\mathcal{R} \leftarrow \mathcal{R} \cup \{\mathbb{H} \twoheadrightarrow \mathbb{B}\}$;
11 **return** $\mathcal{R}$; *% Rules expanding S with Δ^+ with conf. $\gamma_{\min}$.*

Lemma 1. ***(Non-Expandable Association $\mathcal{EL}^{++}$ Rules)***
No rule in S can be expanded with Δ^+ (with respect to `A2`*) if S and Δ^+ have no similarity i.e., $\Phi(S, \Delta^+) = 0$.*

Proof. Since $\Phi(S, \Delta^+) = 0$, there is no invariant entailments between S and Δ^+ i.e., no instance of S (resp. Δ^+) is in an ABox assertion of Δ^+ (resp. S). Thus, none of atom(sets) in S (resp. Δ^+) is bind-able in Δ^+ (resp. S). Condition line 7 is never satisfied: no rule can be expanded from S with Δ^+. ☐

Example 11. ***(Approximate Rule Expansion)***
Let $\mathbb{C} : \{\exists travel.Stop(x_2)\}$. By applying `A2` *($E = \{1\}$) to extend $S : \mathcal{S}_0^9(6)$ with $\Delta^+ : \mathcal{S}_0^9(7)$, $\{(4), (5)\}$ is expanded (lines 5-10) with $\mathbb{C}$ to capture $\{(4), (5)\} \twoheadrightarrow \mathbb{C}$ with confidence $1/2$. Rules such as $occur(x_2, x_1) \twoheadrightarrow Event(x_1)$, including 2-rare atomsets in S cannot be discovered due to E.*

4.3 Incremental Knowledge Discovery (InKD)

Algorithm 3 (`A3`) ensures incremental maintenance of association $\mathcal{EL}^{++}$ rules from $\Delta^- \cup S$ to $S \cup \Delta^+$. `A3` is actioned only if knowledge removed Δ^- and added Δ^+ from S have significative dissimilarities (line 6). Any update Δ^+, which is primarily 1-rare, is not considered for maintenance since the more rare knowledge the less significant associations, generalization and rules. Alternatively all rules in $\Delta^- \cup S$ are used as approximation for $S \cup \Delta^+$ (line 14). For significative update, the interestingness of rules, valid in $\Delta^- \cup S$ and $S \cup \Delta^+$, are revised (line 8). Rules from S are expanded with Δ^+ (line 11) when S, Δ^+ have similarities (line 9 - Lemma 1). `A3` completes the process (line 13) by mining rules in $\Delta^+ \times \Delta^+$ using `ap-genrules` [Lécué and Pan, 2015], noted `a-gr`.

Algorithm 3: `[A3] InKD`$\langle \mathcal{O}, \mathcal{S}_0^n, \Delta^-, S, \Delta^+, E, \varrho_{\min}, \gamma_{\min}, \Phi_{\min} \rangle$

1 **Input**: (i) Axioms $\mathcal{O} : \langle \mathcal{T}, \mathcal{A} \rangle$, (ii) Streams S, $\Delta^- \cup S$, $S \cup \Delta^+$ in $\mathcal{S}_0^n$, (iii) a set of integers $E \subseteq \{1, ..., |S|\}$, (iv) Min. rarity $\varrho_{\min}$, confidence $\gamma_{\min}$, similarity $\Phi_{\min}$.
2 **Result**: $\mathcal{R}$: Rules with min. threshold $\gamma_{\min}$ covering $S \cup \Delta^+$.
3 **begin**
4 $\mathcal{R} \leftarrow \emptyset$; *% Initialization of rules $\mathcal{R}$ in $S \cup \Delta^+$.*
5 *% Min. dissimilarity of Δ^- and Δ^+. Min. non-rarity of Δ^+.*
6 **if** $\Phi(\Delta^-, \Delta^+) < \Phi_{\min} \wedge \varrho_1(\Delta^+) > \varrho_{\min}$ **then**
7 *% Update of rules when adding Δ^+ and removing Δ^-.*
8 $\mathcal{R} \leftarrow$ `A1`$\langle \mathcal{O}, \mathcal{S}_0^n, \Delta^-, S \rangle \cup$ `A1`$\langle \mathcal{O}, \mathcal{S}_0^n, \Delta^+, S \cup \Delta^+ \rangle$;
9 **if** $\Phi(S, \Delta^+) \neq 0$ **then** *% Some similarities of S and Δ^+*
10 *% Expansion of rules from S to $S \cup \Delta^+$.*
11 $\mathcal{R} \leftarrow \mathcal{R} \cup$ `A2`$\langle \mathcal{O}, \mathcal{S}_0^n, S, \Delta^+, E, \gamma_{\min} \rangle$;
12 *% Mining of rules through associations uniquely in Δ^+.*
13 $\mathcal{R} \leftarrow \mathcal{R} \cup \{$`a-gr`$(\Delta^+ \times \Delta^+, \gamma_{\min})\}$
14 **else** $\mathcal{R} \leftarrow \{\rho : \mathbb{B} \twoheadrightarrow \mathbb{H}$ in $S \cup \Delta^-\}$; *% All rules in $S \cup \Delta^-$.*
15 **return** $\mathcal{R}$; *% Rules with confidence $\gamma_{\min}$ covering $S \cup \Delta^+$.*

Example 12. ***(Incremental Knowledge Discovery in Action)***
Knowledge discovery is incrementally maintained by sliding from $[5, 6]$ to $[6, 7]$ in $\mathcal{S}_0^9$ using `A3` *with S:$\mathcal{S}_0^9(6)$, Δ^-:$\mathcal{S}_0^9(5)$, Δ^+:$\mathcal{S}_0^9(7)$.* `A3` *is actioned as Δ^-, Δ^+ are different (line 6, Example 7). Rules are updated (line 8, Example 10), expanded from S (line 11, Example 11) as S, Δ^+ have similarities (line 9). Remaining rules e.g., $\{\exists impact.Serious(x_1), (5)\} \twoheadrightarrow \{\exists travel.Stop(x_2)\}$ are discovered from Δ^+ (line 13).*

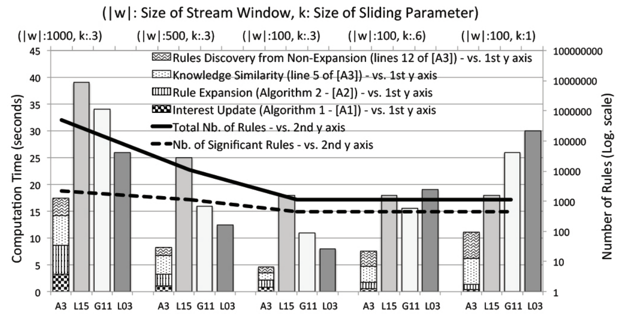

Figure 5: Scalability. 1^{st} x axis: Approaches on 5 Cases. 1^{st} y axis: Computation Time in Seconds. 2^{nd} x axis: 5 Types of Stream Windows and Sliding Configurations. 2^{nd} y axis: Search Space of $\mathcal{EL}^{++}$ Rules.

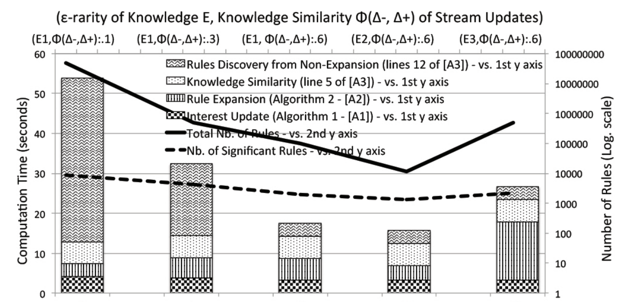

Figure 6: Scalability. 1^{st} x axis: Our Approach on 5 Cases. 1^{st} y axis: Computation Time in Seconds. 2^{nd} x axis: 5 Types of Knowledge Rarity and Similarity Configurations. 2^{nd} y axis: Search Space of $\mathcal{EL}^{++}$ Rules.

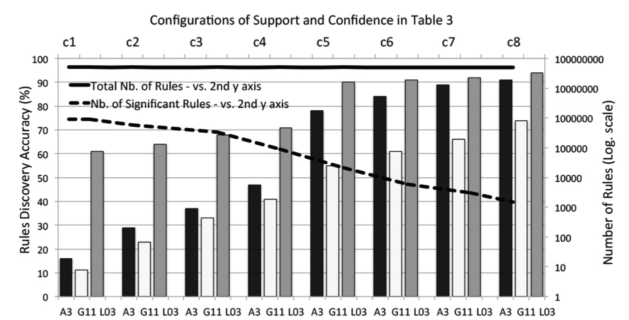

Figure 7: Accuracy. 1^{st} x axis: Approaches on 8 Cases. 1^{st} y axis: Accuracy of Discovered Rules. 2^{nd} x axis: 8 Types of Support / Confidence Configurations (Table 3).

5 Experimental Results

We report scalability, accuracy results by studying the impact of knowledge similarity, ε-rarity on `A3`. The system is tested on: 4 Intel(R) Xeon(R) X5650, 2.67GHz cores, 6GB RAM.

5.1 Context

• **Data:** Data streams (Table 2) related to road weather, travel time, incident, event, bus location in Dublin are transformed in $\mathcal{EL}^{++}$ ontology streams using mapping techniques [Lécué

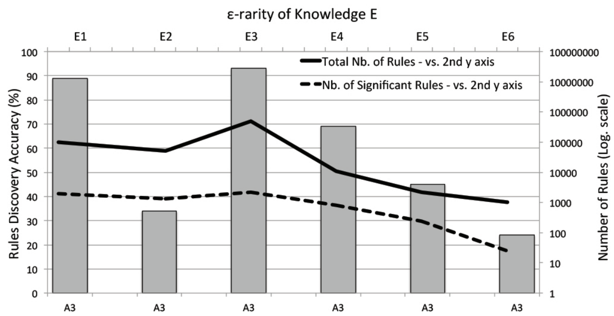

Figure 8: Accuracy. 1^{st} x axis: Our Approach on 5 Cases. 1^{st} y axis: Accuracy of Discovered Rules. 2^{nd} x axis: 5 Types of Knowledge Rarity Configurations.

DataSet	Size (Mb) per day	Frequency of Update (seconds)	#Axioms per Update	#RDF Triples per Update
Weather	3	300	53	318
Travel Time	43	60	270	810
Incident	0.1	600	81	324
Event	9.5	$6,000$	480	$1,150$
Bus	120	40	$3,000$	$12,000$

Table 2: Data Streams Details (average figures).

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
$\sigma_{\min}$	.4	.4	.4	.4	.8	.8	.8	.8
$\gamma_{\min}$	.2	.4	.6	.8	.2	.4	.6	.8

Table 3: Support $\sigma_{\min}$, Confidence $\gamma_{\min}$ Configuration.

et al., 2014]. An ontology with 55 concepts, 19 roles and $25,456$ ABox axioms is considered for semantic enrichment.
• **Settings:** The evaluation is achieved using a variable (i) size of stream window (i.e., snapshots) $|w| \in \{100, 500, 1000\}$, (ii) sliding k in $\{1/3, 2/3, 1\}$ of $|w|$; a variation of min. thresholds of (iii) knowledge similarity in $\{1/3, 2/3, 1\}$, ε-rarity with $\varepsilon \in E \subseteq \{1, ..., |w|\}$, (iv) support, confidence as in Table 3.
• **Baseline Methods:** `L15` [Lécué and Pan, 2015] for mining $\mathcal{EL}^{++}$ rules from scratch, `G11` [Gama and Kosina, 2011] for learning decision rules i.e., subset of association rules, `L03` [Lee *et al.*, 2003] for incremental syntactic update, are compared. We restricted $\mathcal{EL}^{++}$ rules from `A3`, `L15` to have a single concept atom in head similarly to `G11`. All rules from `L03`, `G11` are uplifted using $\mathcal{EL}^{++}$ descriptions in ontologies. Its computation time is not reported for fair comparison.
• **Objective:** The objective is to maintain association ($\mathcal{EL}^{++}$ rules) of knowledge across streams up-to-date while minimizing computation time and maximizing accuracy. Such rules empower reasoning with learning e.g., for prediction.

5.2 Scalability

• **Dynamics:** Figure 5 reports scalability with (i) variable size $|w|$, sliding k, (ii) fixed E i.e., $E_1 : \{1, ..., \lceil |w|/2 \rceil\}$, $\gamma_{\min} : .8$, $\Phi_{\min} : .6$. The scalability decreases with the size of windows and axioms. `A3` is the most scalable in both contexts of increasing windows (i.e., more axioms) and sliding (i.e., less common knowledge) size. This remains valid even when windows do not share snapshots ($k : 1$), which shows the benefits of knowledge similarity and rarity. `L15` is the least scalable for large windows since all rules are re-elaborated from scratch in each update. Its performance remains unchanged for any variation of k since the interleaving of windows is not exploited. On contrary `L03` and `G11` benefit from k i.e., the more interleaving snapshots the more scalable. `G11` reaches comparable performances with `A3` for large windows but fail to scale otherwise since all rules are systematically retrieved.
• **Knowledge:** Figure 6 reports scalability of `A3` with different knowledge similarity Φ and ε-rarity: E_1, $E_2 : \{\lceil |w|/2 \rceil + 1, |w|\}$, $E_3 : \{1, ..., |w|\}$, with fixed $|w| : 1000$ and $k : 1/3$. The scalability decreases with the dissimilarity of updates, limiting the benefits of rules expansion and favoring a time consuming fresh discovery (line 13 of `A3`). E_2 has a slightly better impact on `A3` than E_1 while E_3 does not impact it dramatically. These are caused by the number of ε-rare atomsets derived from E which dropped from E_1 to E_2. Indeed the number of ε-rare atomsets statistically decreases when ε is approaching $|w|$. `A3` scales the most when stream updates have similarities whatever the rarity of its knowledge.

5.3 Accuracy

Figures 7 and 8 reports accuracy by (syntactically) comparing the rules discovered in `A3`, `G11`, `L03` with `L15` (as baseline).

• **Interestingness:** Figure 7 considers (i) variable support, confidence (Table 3), and (ii) fixed $|w| : 1000$, $k : 1/3$. $E : \{1, ..., \lceil |w|/2 \rceil\}$ and $\Phi : .6$ are fixed for `A3`. In all cases the accuracy is more (positively) impacted by support than confidence. This is especially valid for `A3` since the support directly limits insignificant knowledge association, which reduces the number of rules and their approximation. `G11` is the least accurate since the rules expansion is probabilistic. Interestingly the (negative) effect of approximation, through similarity, rarity, is absorbed by the quality of support, confidence. Thus, accuracy can reach 91% since over-generalization / specialization, which drive insignificant updates, are pruned.
• **Knowledge** (Figure 8): Same configuration as Figure 6 except Φ fixed to .6, ε-rarity extended with $E_4 : \{1, ..., \lceil |w|/3 \rceil\}$, $E_5 : \{\lceil |w|/3 \rceil + 1, ..., \lceil 2 \times |w|/3 \rceil\}$, $E_6 : \{\lceil 2 \times |w|/3 \rceil + 1, ..., |w|\}$. The accuracy with E_1 (.89) and E_3 (.93) are very close, which demonstrates that most of ε-rare knowledge, with $\varepsilon > \lceil |w|/2 \rceil$, could not be significantly associated with knowledge from the stream update. The results with E_4, E_5, E_6 also emphasize the more rarity the more accurate is `A3`. This is even more significative in a context of "*concept drift*" i.e., when knowledge drastically changes over windows ($\Phi \approx 0$).

5.4 Lessons Learnt

Knowledge similarity, ε-rarity benefit scalability since they prune the time-consuming steps of (i) over-generalization, (ii) over-specialization, (iii) reaction to insignificant changes. Our approach scales the most when rare knowledge is discarded in line 7 of `A2` (cf. E_2 in Figure 6). However it is preferred to perform `A3` with rare knowledge (cf. E_1 in Figure 8) since the accuracy is not sacrificed with rarity. Indeed the gain of scalability is only of 11.4% for a loss of 61.7% in accuracy by discarding rare knowledge (from E_1 to E_2). In addition the gain of accuracy is only of 2.1% for a loss of 34.4% in scalability by considering all knowledge (from

E_1 to E_3). Considering more expressive DLs would have decreased scalability due to binding, consistency checking. All approaches have been over-performed by A3 in large-scale contexts even when knowledge drastically changed over time.

6 Conclusion and Future Work

Our approach, exploiting the semantics of data streams, discovers knowledge by incremental learning of association $\mathcal{EL}^{++}$ rules across DL-augmented stream data. The incremental maintenance of knowledge discovery, using systematic rules update and expansion, ensures scalability while efficiently maintaining accuracy. Semantics was essential for (i) capturing knowledge association as reusable $\mathcal{EL}^{++}$ rules across streams, (ii) identifying the properties of similarity and rarity in ontology streams which favor approximation and scalability of knowledge discovery. Experiments have shown highly scalable and accurate knowledge discovery in Dublin.

In future work we will investigate compact representations of ontology streams to support highly changing dynamic data.

Acknowledgments

The research leading to these results has received funding from the European Unions Seventh Framework Programme (FP7/2007-2013) under grant agreement ID 318201 (SIMPLI-CITY).

References

[Agrawal *et al.*, 1993] R. Agrawal, T. Imielinski, and A. N. Swami. Mining association rules between sets of items in large databases. In *SIGMOD*, pages 207–216, 1993.

[Agrawal *et al.*, 1996] Rakesh Agrawal, Heikki Mannila, Ramakrishnan Srikant, Hannu Toivonen, and A. Inkeri Verkamo. Fast discovery of association rules. In *Advances in Knowledge Discovery and Data Mining*, pages 307–328. AAAI/MIT Press, 1996.

[Baader *et al.*, 2003] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*, 2003.

[Baader *et al.*, 2005] Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the el envelope. In *IJCAI*, pages 364–369, 2005.

[Barbieri *et al.*, 2010] Davide Francesco Barbieri, Daniele Braga, Stefano Ceri, Emanuele Della Valle, and Michael Grossniklaus. Incremental reasoning on streams and rich background knowledge. In *ESWC (1)*, pages 1–15, 2010.

[Bienvenu *et al.*, 2012] Meghyn Bienvenu, Carsten Lutz, and Frank Wolter. Query containment in description logics reconsidered. In *KR*, 2012.

[Cheung *et al.*, 1996] David W Cheung, Jiawei Han, Vincent T Ng, and CY Wong. Maintenance of discovered association rules in large databases: An incremental updating technique. In *Data Engineering, 1996. Proceedings of the Twelfth International Conference on*, pages 106–114. IEEE, 1996.

[Galárraga *et al.*, 2013] Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. AMIE: association rule mining under incomplete evidence in ontological knowledge bases. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013*, pages 413–422, 2013.

[Gama and Kosina, 2011] João Gama and Petr Kosina. Learning decision rules from data streams. In *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 1255–1260, 2011.

[Glimm *et al.*, 2007] Birte Glimm, Ian Horrocks, Carsten Lutz, and Ulrike Sattler. Conjunctive query answering for the description logic shiq. In *IJCAI*, pages 399–404, 2007.

[Huang and Stuckenschmidt, 2005] Zhisheng Huang and Heiner Stuckenschmidt. Reasoning with multi-version ontologies: A temporal logic approach. In *International Semantic Web Conference*, pages 398–412, 2005.

[Krötzsch *et al.*, 2008] Markus Krötzsch, Sebastian Rudolph, and Pascal Hitzler. Description logic rules. In *ECAI*, pages 80–84, 2008.

[Lécué and Pan, 2013] Freddy Lécué and Jeff Z. Pan. Predicting knowledge in an ontology stream. In *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, 2013.

[Lécué and Pan, 2015] Freddy Lécué and Jeff Z. Pan. Consistent knowledge discovery from evolving ontologies. In *AAAI*, 2015.

[Lécué *et al.*, 2014] Freddy Lécué, Robert Tucker, Veli Bicer, Pierpaolo Tommasi, Simone Tallevi-Diotallevi, and Marco Luca Sbodio. Predicting severity of road traffic congestion using semantic web technologies. In *The Semantic Web: Trends and Challenges - 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings*, pages 611–627, 2014.

[Lee *et al.*, 2003] Chang-Hung Lee, Ming-Syan Chen, and Cheng-Ru Lin. Progressive partition miner: An efficient algorithm for mining general temporal association rules. *IEEE Trans. Knowl. Data Eng.*, 15(4):1004–1017, 2003.

[Völker and Niepert, 2011] Johanna Völker and Mathias Niepert. Statistical schema induction. In *The Semantic Web: Research and Applications - 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29-June 2, 2011, Proceedings, Part I*, pages 124–138, 2011.