Explaining and Predicting Abnormal Expenses at Large Scale using Knowledge Graph based Reasoning

Freddy Lecue, Jiewen Wu

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2017) 1–20

Procedia Computer Science

Explaining and Predicting Abnormal Expenses at Large Scale using Knowledge Graph based Reasoning

Freddy Lecue[a,b], Jiewen Wu[b]

[a]*INRIA, Sophia Antipolis, France*
(firstname.lastname)@inria.fr
[b]*Accenture Labs, Dublin, Ireland*
(firstname.lastname)@accenture.com

Abstract

Global business travel spend topped record-breaking \$1.2 Trillion USD in 2015, and will reach \$1.6 Trillion by 2020 according to the Global Business Travel Association, the world's premier business travel and meetings trade organization. Existing expenses systems are designed for reporting expenses, their type and amount over pre-defined views such as time period, service or employee group. However such systems do not aim at systematically detecting abnormal expenses, and more importantly explaining their causes. Therefore deriving any actionable insight for optimising spending and saving from their analysis is time-consuming, cumbersome and often impossible. Towards this challenge we present AIFS, a system designed for expenses business owner and auditors. Our system is manipulating and combining semantic web and machine learning technologies for (i) identifying, (ii) explaining and (iii) predicting abnormal expenses claim by employees of large organisations. Our prototype of semantics-aware employee expenses analytics and reasoning, experimented with 191, 346 unique Accenture employees in 2015, has demonstrated scalability and accuracy for the tasks of explaining and predicting abnormal expenses.

Keywords: Semantic web; Reasoning system; Intelligent system; Smart expenses; Intelligent Operation System; Knowledge graph

1. Introduction

Time and expense is the term used by the finance department of most large organisations to capture the "*process of recording and tracking hours worked and expenses as they relate to projects*". \$546 Billion worldwide has been estimated to be lost in 2015 because of the lack of spend optimisation in the process of managing expenses at organisation level [1].

Such non optimization is characterized by (i) absence of context capture, which strongly restricts the interpretation of high-value expenses, (ii) rudimentary approaches to catch any level of employees' frauds, which result in over \$1 billion lost each year to fraudulent expense reimbursement[1], (iii) ad-hoc auditing of employee's expenses, which results in unflagged abnormal spending, (iv) rigid and static expenses policy which does not fit all expenses and limits it efficiency to well-know expense types, (v) inappropriate tools for expensing and capturing causes of abnormal expenses. All limitations are due to the manual design of predefined rules and policies for patterns that are supposed to command all expenses data.

AIFS[2,3] (**A**rtificial **I**ntelligence **F**inance **S**ystem), as a system which integrates exogenous data from hetero-

[1]https://www.accountingtoday.com/opinion/expense-report-fraud-will-cost-companies-1-billion
[2]Video (.mp4 format) available: https://goo.gl/K8UcI2
[3]Live system: http://54.194.213.178:8111/ExplanatoryReasoning/demo.jsp

geneous structured and unstructured data, aims at (i) identifying, (ii) explaining and (iii) predicting abnormal expenses. We illustrate our approach using accommodation expenses but our approach also covers transportation (i.e., taxi), flight and entertainment (i.e., meals) expenses. All previous abnormal expense types are denoted as abnormalities. Most existing modern expenses management systems such as Concur Technologies[4] or Chromeriver[5] provide tools for basic control and reporting using various internal systems such as employees' expenses, time or credit card status. Others such as Expensify[6] expose views by geography, career level, service group for manually pinpointing abnormal expenses but do not deliver insight to identify and interpret abnormal expenses i.e., expenses which are unexpected in a given context. Basic in-depth but semantics-less state-of-the-art analytics are employed, limiting any interpretation, explanation or prediction of abnormal patterns. Therefore, context-aware computing together with reusability of the underlying data is quite limited.

AIFS, designed as a semantic web-based application for (i) interpreting expenses and their context, and (ii) deriving innovative and easy-to-explore insights, tackles these limitations by seamlessly and smoothly integrating and augmenting the following in an intelligent user interface:

(i) in-depth *analysis* of any-time employee's expenses, which supports semantic comparison of expenses,
(ii) *explanation* [2], which aims at connecting abnormal expense to its causes through explanations,
(iii) tentative expenses *forecasting* using recent theoretical research work in contextual predictive reasoning [3].

The system has been matured and is now used on a daily basis by business owner together with auditors to understand abnormal expenses, their context, their causes, and take appropriate actions depending on the insight derived from our system. From a business perspective, the explanation of abnormal expenses is used to prevent them in a near future and identify new expenses policy on-the-fly, mainly by interpreting their underlying context. The case of Accenture has demonstrated a potential reduction of 7.8% of the overall travel expenses amount by enforcing learnt contextualized rules for future travel related expense items. Although expenses management still requires classic mechanisms of delegation, authorisation and reviews, our system has shown to cut the overhead of some very manual tasks from the auditors such as "*requesting abnormal expenses justification from employees*" or "*requesting expenses contexts from employees*", which is very difficult for an individual as context might not be completely known.

The novelty of AIFS lies in the ability of the system to ingest highly heterogeneous data (cf. Table 1) and perform various types of inferences i.e., analysis, explanation and prediction. These inferences are all elaborated through a combination of various types of reasoning i.e., (i) semantic Description Logic (DL) $\mathcal{EL}^{++}$-based reasoning i.e., distributed ontology classification-based subsumption [4], (ii) rules-based i.e., semantic pattern association [3] and (iii) machine learning based i.e, consistent knowledge discovery from temporal and semantic representation [5].

AIFS completely relies on the W3C semantic Web stack e.g., OWL 2 (Web Ontology Language) and RDF (Resource Description Framework[7]) for representing semantics of information and delivering inference outcomes. Currently applied using data from Accenture employees' expenses, AIFS can scale to any other company, which exposes expense data.

This paper is organized as follows: The next section comments on related work and position the existing systems, technologies and research work with respect to identified challenges. Section 3 presents the travel expenses context in Accenture. Section 4 sketches in-use scenarios for AIFS together with associated research challenges and our approach. Section 5 describes the main technologies behind AIFS. Section 6 reports experimental results regarding its scalability and accuracy. Finally we draw some conclusions and talks about future directions in Section 7.

2. Related Work

We appraise the existing approaches to explain any (past or future) temporal abnormalities and classify them according to the following three dimensions: (1) the extent to which an approach support *interpretation and explanation*; (2) the *accuracy* of the approach; (3) the level of *industry maturity* of the approach. The analysis is driven towards our finance application.

Sample of the system: (1) prediction functionality not exposed, (2) only subset of data exposed due to Accenture restriction.

[4]https://www.concur.com/

[5]https://www.chromeriver.com/

[6]https://www.expensify.com/

[7]Semantic web standard: http://www.w3.org/RDF/

The chart in Figure 1 positions the reviewed approaches in relation to these three dimensions. These dimensions will be used to structure the remainder of this section.

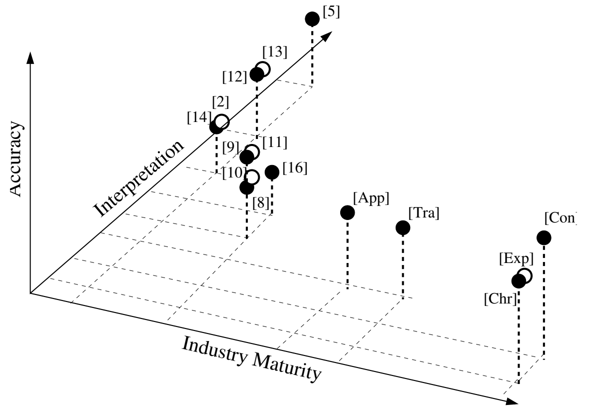

Figure 1: Classification of Approaches. 3 Dimensional Metric: (i) their support of interpretation and explanation, (ii) their accuracy, and (3) level of industry maturity.

2.1. Interpretation / Explanation Dimension

In most of data mining [6] and machine learning [7] applications, patterns are learnt from data for detecting anomalies or deriving prediction. In particular the following tasks are performed: (i) correlating current and past data (e.g., levels of abnormal expenses), (ii) identifying patterns using different distance metrics [8], and (iii) selecting "rules" that are used for identifying abnormalities predicting future conditions. More sophisticated approaches e.g., [9] investigated sequential pattern mining for capturing the time-based evolution of patterns, which is relevant to our application, specially for comparing expenses across time and context.

These approaches are designed for isolated data where context has no impact. Therefore they rarely utilize exogenous sources of information for adjusting and explaining estimated anomaly detection or prediction. Towards this issue some approaches such as [10, 11] go further by augmenting existing models with cross-correlation across multiple data sources by learning contextual rules. They largely improve the accuracy of the model but do not support any interpretation of the results. Recent works in the area of explanatory reasoning [2, 5, 12], combining machine learning and reasoning, have emerged to address the old problem of interpretation [13], also known as abductive reasoning [14]. Promising results have been released but limitations related to their scalability, ease-of-use, and maturity made existing approaches not mature and robust enough for end-user driven explanation [15] and real-world industrial application deployment.

From a more applied perspective the existing and most mature industry-ready systems such as Concur[3] (noted [Con] in Figure 1), Expensify[5] (noted [Exp] in Figure 1) or Chromeriver[4] (noted [Chr] in Figure 1) are the most scalable system but do not capture insight from the integration of heterogenous and exogenous open data sources. Therefore context-aware anomaly detection, explanation or prediction cannot be supported in their current release.

2.2. Accuracy Dimension

Industrial systems in the area of expenses management such as Concur[3], Chromeriver[4], Expensify[5] are the most accurate but the least innovative, specially when evaluating functionalities of interpretation and explanation. Data mining and machine learning models [8, 16] have various degrees of accuracy. The approach of [9] has shown to have better accuracy but more limited scalability. In addition most of the models have demonstrated limited robustness to sudden changes over time [17], also known as concept drift in temporal data (which is crucial in our context of expenses and temporal evolution of their context).

From an interpretation and explanation perspective, existing logics-driven frameworks have clearly shown limitations specially when combining and correlating large amount of semantics-driven data. The authors of [2], extending models of [13, 14], analyses the semantic correlation of data using Description Logic to derive explanation of abnormal patterns. Although the approach tackles the challenge of identifying potential causation among semantic representation of data, accuracy remains an open problem. Recent work [5, 12] extended the framework by automatically determining dynamic rules. Our work follows the same principles for supporting compilation of any-time abnormal expenses and its explanations.

2.3. Industry Maturity Dimension

There are expenses management systems, as standalone applications, which support advanced control and reporting of expenses. They usually provide elaborated tools for (i) basic data management e.g., Concur[3], which is storing, organizing or joining large amount of data from various internal systems e.g., employees' expenses, time, credit card status or travel booking system, (ii) summarising expenses at various levels e.g., Expensify[5] which is providing views by geography, career level, service group, (iii) detecting top spenders e.g., Chromeriver[4], which is identifying the top spenders on various domains e.g., expense type, timespan.

However they do not expose systems for smoothly analyzing, easily exploring and interpreting global expenses on various combinations of contexts e.g., spatial, temporal, event-based. Indeed they also all fail in using and interpreting underlying semantics of data, making anomaly detection and prediction not as accurate and consistent as it could be, specially when underlying data is characterized by texts. Recent research works driven by AppZen[8] (noted [App] in Figure 1) have integrated external data, but mainly for (i) fraud detection and (ii) disambiguation of text-based expenses such as name of amenities where expenses occur. ExpenseTrack[9] (noted [Tra] in Figure 1) is bringing new innovation in the area of expenses management by handling better data ingestion pipelines. However they still limit the capture of insight from internal company data, which is reducing the complexity of the problem and limiting the interpretation of abnormal expenses.

Other approaches, capturing the semantics of data to interpret abnormalities such as [2, 5, 12, 13, 14], have either not been applied in our finance domain, or not been considered for large scale deployment in industry.

Most of data-intensive approaches [8, 16, 9, 10], which learn patterns and models from raw data, have demonstrated excellent research contributions. However their transfer to industrial applications have shown some limitations due to the (i) robustness of the solutions, (ii) open challenges such as concept drift in temporal data [17].

2.4. The Need for a Holistic Approach

Review of existing approaches to explain any (past or future) temporal abnormalities reveals that no approach has specifically addressed the problem of *accurate* abnormal expenses detection, *interpretation / explanation* and prediction using semantic representation of information in an *industrial context*. Indeed main approaches focus on either (i) large scale systems for basic expenses control and reporting or (ii) advanced techniques for learning and interpreting models for very targeted application domains, either with limited accuracy or constrained scalability. This motivates our innovative model that addresses the problem of *predicting and explaining abnormal expenses* by fusing machine learning and reasoning techniques.

[8]https://www.appzen.com/
[9]http://www.expensetracks.com/

Regarding this issue, we follow [5] and suggest the use of semantic representation of data to automatically determining dynamic rules using exogenous expenses context. Such rules provide insight for explanation or more accurate prediction. Our approach also extends recent works in diagnosis reasoning [2] by supporting compilation of any-time abnormal expenses and its explanations (Figures 5-8). Last but not least we present a unique system that combines and unifies technologies from machine learning and reasoning communities to address real-world challenges from business owners and auditors of the finance department of a top 500 fortune company, comprising more than 375, 000 employees.

3. Context: Travel Expenses

Table 1 reports all data sources processed by AIFS in the Travel Expenses in Accenture scenario with respect to their format, and size. They report various types of information coming from static or dynamic sources, exposed as open or private data and described along heterogeneous formats.

The *expenses* data is used for capturing travel expenses of 191, 346 unique Accenture employees. Expenses types range from accommodation, flight, taxi, public transportation meals and entertainment. They are captured on biweekly basis with a minimum, mean and maximum number of respectively 0, 43.38 and 201 items per employee per period. Overall 1, 335, 691, 105 travel related items have been expensed in 2015. The *expenses* data set is considered as our source of anomalies (cf. Anomaly in Table 1). We aim at extracting "abnormal" expenses among all expenses.

Example 1. (Expense Item)
A (simplified and partially anonymized) expense item is given below as a tabular entry.

ExpenseId	*FromDate*	*ToDate*	*ExpenseType*	*TotalAmt*
2323423	*2015-08-03*	*2015-08-05*	*Hotel*	*550*
Country	*City*	*CareerLvl*	*Currency*	*PeopleKey*
United States	*Austin*	*6*	*dollar*	*216532*

This expense item, identified with ID 2323423 *captures an accommodation expense of USD* 550 *in Austin, USA between August* 3^{rd} *and* 5^{th}*, 2015. The expense has been occurred by an employee with identifier:* 216532.

Analyzing, explaining and predicting abnormal expenses consists in interpreting, contextualizing and correlating its content with the following two exogenous data sources (cf. Potential Explanation in Table 1): (1) *social media events* which characterize events of various type e.g., music, sport, politics, family, with an average of 187 events per day and city, all updated on a daily

Source Type	Data Source	Description	Format	Historic (Year)	Size per day (GBytes)	Data Provider
Anomaly	190,000+ unique travellers in 500+ cities recorded for 2015	Min., max. number of respectively 2,521 and 24,800 expenses per city[a]	CSV	2015	.93 (complete) .41 (aggregated)	Private
Potential Explanation	Social media events e.g., music event, political event	Planned events with small attendance	JSON format Accessed through Eventbrite APIs[a]	2011	Approx. 94 events per day (.49 GBytes)	Eventbrite
		Planned events with large attendance	XML format Accessed through Eventful APIs[b]	2011	Approx. 198 events per day (.39 GBytes)	Eventful
	Media news event	Events reported in the media worldwide	JSON format Accessed through EventRegistry APIs[c]	2015	Approx. 198 events per day (.76 GBytes)	EventRegistry
Semantics	DBpedia	Structured facts extracted from wikipedia	RDF[d]	-	Approx. 33,000+ resources in use (.23 GBytes)	Wikipedia
	Wikidata	Structured data from Wikimedia projects	RDF[e]	-	Approx. 189,000+ resources in use (.63 GBytes)	Freebase Google inc.
	Accenture Categories	Structured is-A taxonomy of event categories	RDF[f]	-	25 resources in use (.001 GBytes)	Accenture inc.
Spatial	World Map (listing of type, GPS coordinate)		OSM XML	-	666 GBytes	Open StreetMap[g]

[a] https://www.eventbrite.com/api
[b] http://api.eventful.com
[c] http://eventregistry.org/ (restricted access with limited calls) - http://beta.eventregistry.org (less restricted)
[d] http://wiki.dbpedia.org/Datasets
[e] https://www.wikidata.org/wiki/Wikidata:Database_download
[f] http://54.194.213.178:8111/ExplanatoryReasoning/ontology/categories.n3
[g] ftp://ftp.spline.de/pub/openstreetmap/

Table 1: (Raw) Data Sources for Travel Expenses in Accenture Scenario.

basis. (2) *Media news events* which captures real-time news articles published by over 100,000 news publishers globally, with an average of 981 news articles per day and city (with population higher than 1,000,000).

Figure 2 captures events (cf. sources of potential explanations in Table 1) on September 19th, 2016 in Europe and east coast of north America. The heat-map type representation, exploiting the ESRI SHAPE file of the world, gives an overall view on the intensity of number of events during this day.

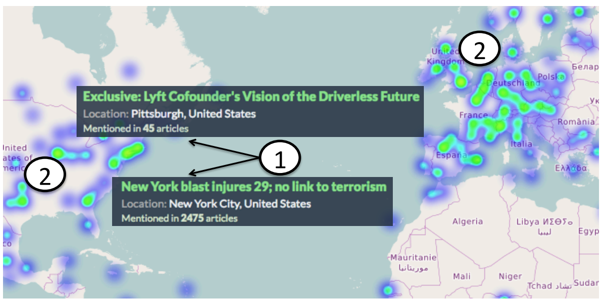

Figure 2: Events and their Occurrence on September 19th, 2016. ①: Particular events occurring in Pittsburgh and New-York. ②: Heat-map on intensity of number of events in Europe and east coast of North America. (color print).

4. AIFS Scenario, Challenges and Approach

The AIFS system is illustrated through a list of scenarios, where each highlights actions that any expenses business owner or auditor is required to perform on a daily basis. Such scenarios and their underlying actions, have been identified by our users groups as non supported by state-of-the-art systems in place. In-use industrial solution from Concur Technologies, which exposes travel and expense management services to businesses, is limited to the use of internal employee expenses for its reporting and analysis. Therefore existing systems fails in supporting these scenarios because of the unsupported: (i) not-so-easy tasks of data integration, (ii) automated abnormal pattern detection, and (iii) underlying contextual semantic reasoning across exogenous and heterogenous data.

The use of semantic web technologies in all our scenarios is transparent to end-users. However such technologies are strongly required to (i) encode semantic information through query-able knowledge bases, (ii) compile and deliver contextual analysis, explanation and prediction form such heterogenous data sets. All user interactions (UI) are achieved through simple UI paradigms e.g., spatial and temporal selection for initialization (respectively ①, ② and ③ in Figure 4) where dates interval together with region, country and city can be directly selected from the control panel e.g., 01/01/2015 - 31/12/2015 and Austin city in ②. All results, delivered by analysis, explanation and prediction, are dynamically exported as parallel, spider, pie, graph-based and time-series charts.

For each scenario, we sketch its (i) description, (ii) motivation, (iii) state-of-the-art approach, (iv) challenge - emphasised by our expenses business owner and auditors group, with (v) the AIFS approach from a technical and UI perspective, its (vi) scalability, (vii) limitation.

Algorithm 1: Context-aware Anomaly Detection

1 **Input**: (i) Terminology $\mathcal{T}$ for representing eXpenses, (ii) all eXpenses $\mathbf{X}$ in a d representation space of $\mathcal{T}$, (iii) an employee eXpense x, (iv) Min. threshold of semantic similarity m_t between expenses, (v) Min. threshold of decision rule support m_s, confidence m_c.
2 **Result**: Boolean if x is **Abnormal** w.r.t. $\mathbf{X}$ and $\mathcal{T}$.
3 **begin**
4 $\tilde{\mathbf{X}} \leftarrow x$; *% Init. of items semantically similar to x.*
5 *% Expenses with semantic similarity with x.* ①
6 **foreach** $x_i \in \mathbf{X}$ *of the form* $(x_i^1, \cdots, x_i^d)$ **do**
7 **if** $Sim_{\mathcal{T}}(x, x_i) > m_t$ **then** $\tilde{\mathbf{X}} \leftarrow \tilde{\mathbf{X}} \cup x_i$;
8 $\mathcal{R} \leftarrow \emptyset$; *% Initialization of relevant rules in* $\tilde{\mathbf{X}}$.
9 *% Rules of dimension* $d' < d$ *in* $\mathcal{T}$ *w.r.t* $\tilde{\mathbf{X}}$. ②
10 **foreach** *rule* $\rho \in \mathcal{T}|_{\tilde{\mathbf{X}}}^{d'}$ *with* $\forall d' < d$ **do**
11 **if** $support(\rho) > m_s \wedge confidence(\rho) > m_c$ **then** $\mathcal{R} \leftarrow \mathcal{R} \cup \{\rho\}$;
12 $\mathcal{T}_x \leftarrow \mathcal{T}$; *% Init. of semantic context of x w.r.t* $\mathcal{T}$.
13 *% Semantic context for expense x, noted* $\mathcal{T}_x$. ③
14 **foreach** $\rho \in \mathcal{R}$ *of the form* $\mathcal{G} \twoheadrightarrow h$ **do**
15 **if** $h(x)$ *in semantically consistent with* $\mathcal{T} \cup \{x\}$ **then** $\mathcal{T}_x \leftarrow \mathcal{T} \cup \{x\} \cup \{h(x)\}$;
16 *% Abnormal expense x w.r.t* $\mathcal{T}_x$. ④
17 **if** $\mathcal{T}_x \models$ **Abnormal**(x) **then return true**;
18 **return false**;

4.1. Spatio-Temporal Analysis of Expenses

• ***Description and Motivation***: Expense business owner and auditors are interested in any-time expense types, ideally grouped by level of abnormality in order to visually extract problematic expenses together with its context at any time and space.

• ***State-of-the-art Approaches***: Existing systems are mainly focusing on reporting (i..e, compiling information using pre-defined views) and do not expose any context for dynamic analysis of expenses over time. Therefore the detection of abnormal expenses is rudimentary e.g., via basic selection of highest expenses.

• ***Challenge***: Any expense is incurred in a given context e.g., for different profiles of employee, with various levels of business justification, or in a specific city profile. Thus it is crucial to capture the context of all expenses to determine the groups of abnormality in expenses. Rules for detecting abnormal expenses cannot be static, and need to be learnt and updated over context and time.

• ***Approach (Technical)***: AIFS relies on the semantic integration of heterogenous and exogenous raw data such as events, employee expenses and their profile (among others). This is particularly important to detect the context of expenses together with their degree of abnormality. All content of data sources is semantically described by aligning their metadata with vocabularies in Table 1 cf. *Semantics* source type. Algorithm 1, illustrated with Figure 3, aims at detecting whether an expense x is abnormal given all expenses $\mathbf{X}$ and their associated context with dimension d e.g., travel trip, business justification, city destination or travel duration.

We denote by $\mathcal{T}$ the terminology used for semantic representation. The approach mainly consists of: (i) retrieving all similar expenses to x (lines 5 - 7 - ① in Figure 3) where $Sim_{\mathcal{T}}(x, x_i)$ is a matching function in $[0, 1]$ capturing subsumption-based semantic similarity between x and x_i w.r.t $\mathcal{T}$ [18], (ii) learning rules from similar expenses through association mining of semantic descriptions [3] (lines 9-11 - ② in Figure 3) for inferring correlation and rules between expenses and city context, (iii) retrieving relevant semantic context for comparison with x (lines 13-15 - ③ in Figure 3), and (iv) firing all rules which derive abnormality (lines 16 - 17 - ④ in Figure 3). 50 features describing expense, city, employee profile, social and media news events have been captured and are considered to learn rules (lines 9-11 - ② in Figure 3). Rules are selected depending on minimal threshold of support and confidence. An initial sample set of 106,809 North America and Europe travel related expense items (spreading across 167 different days of 2015) have been labeled regarding the abnormality level. This phase has been conducted by audit experts over a period of 89 days.

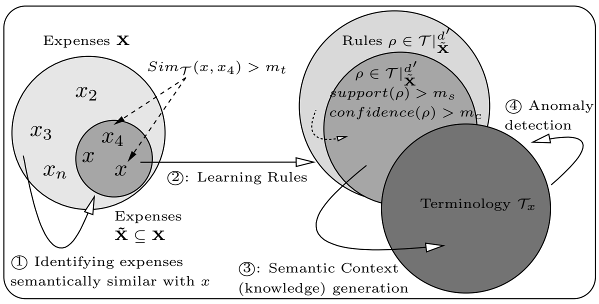

Figure 3: Schematic Flow Diagram of the 4-Steps Context-Aware Anomaly Detection Approach in Algorithm 1.

All rules are filtered based on their occurrence (i.e., support) and confidence in line 11. Rules are encoded using the implementation of DL $\mathcal{EL}^{++}$ rules [19] in [5].

Our system could, for example, learn, adapt and trigger Description Logic rules [3] for anomaly detection e.g., (1-7): "*the accommodation related expense x is abnormally high if its amount is higher than* 90% (4) *of expenses from manager e* (5) *traveling to cities c with a population higher than* 1 *million* (6)*, with events ev of type music and movie* (7)".

$$HighAbnormalExpense(x) \leftarrow \tag{1}$$
$$expensed(e,x) \;\wedge\; inCity(x,c) \;\wedge\; events(c,\{d\},ev) \;\wedge \tag{2}$$
$$(Expense \;\sqcap\; \exists type.Accommodation \;\sqcap\; date.\{d\} \tag{3}$$
$$\exists amount.(\exists\; higherThan.90\%\_Context\_Expense))(x) \;\wedge \tag{4}$$
$$(Employee \;\sqcap\; \exists career.Manager)(e) \;\wedge \tag{5}$$
$$(City \;\sqcap\; (\exists p\_size.(\exists moreThan.1Million)))(c) \;\wedge \tag{6}$$
$$(Event \;\sqcap\; \exists category.(\{Music, Movie\}))(ev) \tag{7}$$

This rule connects *employee profile*, *expense description*, *social events* and *city characteristics*, which demonstrates how context is "*glued*" to internal data. Contrary to state-of-the-art approaches, rules are based on contextual information.

Example 2. (Context-aware Anomaly Detection) *Suppose expense x in Example 1 to be compared with all other expenses (lines 5 - 7 - ① in Figure 3), and in particular expense x′ defined as:*

ExpenseId	FromDate	ToDate	ExpenseType	TotalAmt
6124356	2015-08-06	2015-08-08	Hotel	630
Country	City	CareerLvl	Currency	PeopleKey
United States	San Jose	5	dollar	653211

Both Austin and San Jose are cities where the number of conferences is higher than 6 *during week* 32 *of year* 2015*. Suppose x′ is representative of abnormal weekly expenses data for employees with a career level less than* 6 *in cities with a number of conferences higher than* 6*. Therefore the example expense item x in Example 1, will be classified by Algorithm 1 as "highly abnormal". Indeed the previous context of x′ ensure to capture salient Description Logics rule* (8) *with high confidence and support (lines 9-11 - ② in Figure 3).*

$$\begin{aligned}
&HighAbnormalExpense(x) \leftarrow\\
&\quad expensed(e,x) \;\wedge\; inCity(x,c) \;\wedge\\
&\quad (Expense \;\sqcap\; \exists type.Accommodation \;\sqcap \exists inWeek.\{w\}\\
&\qquad \sqcap \exists amount.(\exists\; moreThan.90\%_Cont._Exp.))(x) \;\wedge\\
&\quad (Employee \;\sqcap\; \exists careerLvl. \leq_6)(e) \;\wedge\\
&\quad events(c,\{w\},nbConf) \wedge nbConf \geq 6
\end{aligned} \tag{8}$$

Suppose that (8) *is the unique learnt rule in the system. Since ABox assertion HighAbnormalExpense*(x) *does not contradict with the initial background knowledge and no any other rules is conflicting with* (8) *(through inconsistency checking) then the derived knowledge from* (8) *is consistent and validated (lines 13-15 - ③ in Figure 3).*

A particular note regarding this rule is that it considers the number of conferences during a week to be impactful on abnormal expenses in that period.

• ***Approach (UI)*** (Figure 4): The component ⑤ in Figure 4 embeds the results of the exploration phase in a parallel chart, where the status (cf. colored abnormality levels using green, yellow, orange, red, and lack coloring) of each employee expense together with its amount, duration, date, day of year, type, people key (or id), and career level are reported. All types of expenses at city level ④ e.g., Austin in Figure 4, can be also displayed to understand the context of the analysis i.e., number of employees travelling to Austin in 2015, their average number of business trips in this city, duration, amount and abnormality level. The table, reported as ⑥, gives a detailed view, where a selection of a row highlights the corresponding elements in the parallel chart. The pie chart ⑦ establishes the proportion of expenses status in the boundary box ③.

• ***Scalability***: The models are computed off-line with 1 year of data, which ensures scalability. However it is more problematic for a model with more than 3 years of data due to the exponential complexity of computing semantic rules.

• ***Limitation***: Rules are continuously learnt given the temporal nature of expenses and their context. 22% of rules have changed between a model with 6 and 12 months of expense data. By considering a model with 2 years of data, the proportion of false positive (normal expenses flagged as abnormal) is 8.6% and false negative is 9.7%. This could be improved by identifying the best values of the hyper-parameters of the algorithm (i.e., support, confidence of rules, threshold of semantic similarity) through cross-validation. Another complimentary approach would consist in considering a larger set with daily aggregates of expenses.

4.2. Explanation of Abnormal Employee Expenses

• ***Description***: How to identify the nature and cause of abnormal employee expense? How to capture explanations on a spatial and (historical) temporal basis?

• ***Motivation***: Above questions cannot be answered by existing state-of-the-art business travel and expense management software solution, but are really important for (i) business owner to establish new policy and (ii) auditors to not only better target expenses but also to

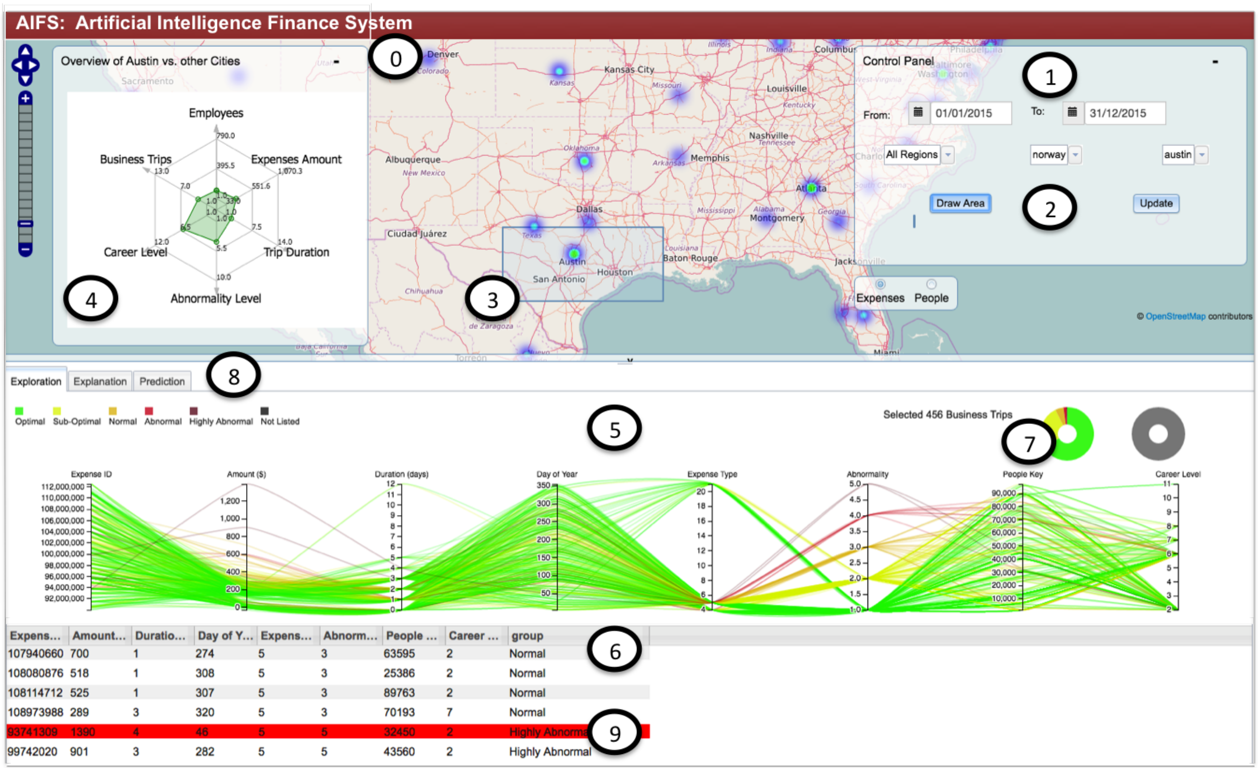

Figure 4: Interface of AIFS. ⓪: Name of the system, ①. ②: Temporal context (subject to user selection). ③: Spatial (Map Area) context (subject to user selection). ④: Expenses overview context for the spatio-temporal analysis. ⑤: Individual expense status and profile i.e., type, amount, duration, date, employee id, and career level (segments are select-able). ⑥: Detailed version of expenses (records are select-able with automated update on the parallel chart). ⑦: Spatio-temporal proportion of expenses status. ⑧: Tab-based selection of analytics and reasoning: explanation and prediction. ⑨: Selection of an abnormal expense to be explained in further step in Figure 5. (color print).

minimize communication with employees by pinpointing causes of abnormal expenses.

• ***State-of-the-art Approaches***: No system provides explanation of abnormal expenses. In the best case scenario, anomalies are triggered by (manually defined) rules, and such rules are reported back to decisions makers as "*explanation*". Rules are encoded using in-house data from employees' expenses, which limit any evidence of exogenous explanation.

• ***Challenge***: Such questions remain open because (i) historical expenses, (ii) their contexts through relevant data sets (e.g., minor / major city events, temporal: seasons, spatial: city data), and (ii) their correlation (e.g., large music festival events correlated with high price of accommodation in large cities) and potential causation with / to abnormal expenses have not be fully integrated and interpreted, even considered.

• ***Approach (Technical)*** (Figure 6): AIFS exploits the semantics of both historical abnormal expenses, their context and sources of potential explanations.

It compiles off-line a sample of abnormal expenses and their explanation (i.e., a 12% sample has been annotated by domain experts - it covers $8,324$ abnormal expense items of the initial sample set of $106,809$ North America and Europe travel related expense items) into a deterministic finite state machine. The state machine is augmented with respect to the semantic-augmented context (e.g., events, temporal: seasons, spatial: city data) where each abnormal expense x is connected to (i) its semantic context, c (ii) explanation e. The explanation of new abnormal expenses is performed by analyzing the state machine to retrieve similar contexts i.e., expenses type, spatial such as cities and events for travel related expenses, temporal. The similarity is estimated by comparing semantic descriptions of the context through matchmaking functions introduced by [18] and [20]. Explanations of a new abnormal expense are then retrieved from similar expenses and contexts. Therefore, the closer are the contexts the more common explanations are shared.

The semantic representation of explanations is exploited to return two levels of description. Suppose an

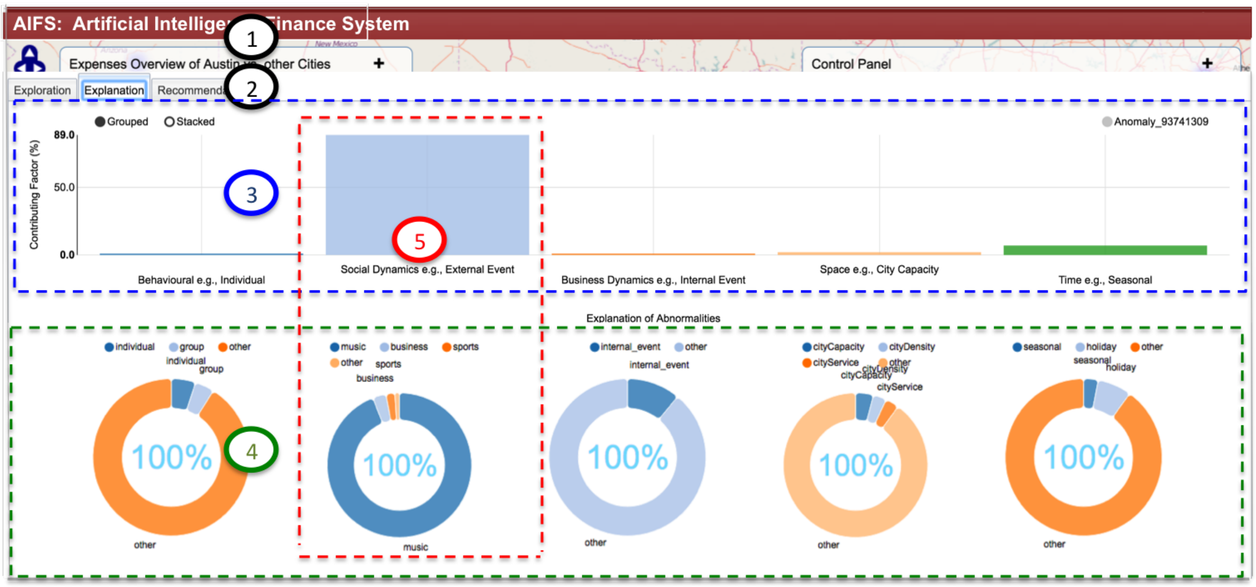

Figure 5: AIFS - (Business) Explanation Component. ①: Spatial (Map Area) context (subject to user selection) for spatial interpretation of abnormal expenses. ②: Tab-based selection of reasoning: explanation and prediction. ③: Level-1 Explanation structured in 5 classes of explanation with contributing factors in [0, 100]. ④: Level-2 Explanation, identifying specification of Level-1 explanation. ⑤: Illustration of the top contributing explanation and its specification for anomaly 93741309. (color print).

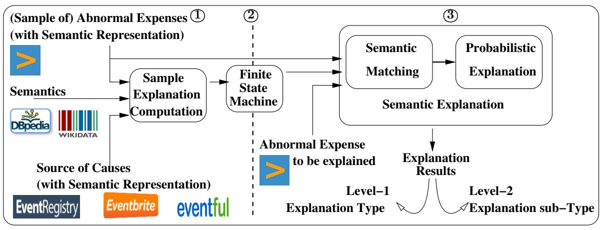

Figure 6: Semantic Explanation Approach Overview. ①: Compilation of (sample of) abnormal expenses and their explanation. ②: Finite State Machine based representation of sample expenses and their context. ③: Semantics-based comparison, evaluation of explanation.

abnormal expense related to accommodation in Austin. Level-1 is capturing the main type e.g., external event while Level-2 representing a more specific explanation e.g., festival music event. Each level has a predetermined number of class, capturing different potential types of causes. All classes, validated by our user group, represent various facets of explanation, extracted from the semantic analysis. Level-1 type is predetermined e.g., all social media and news events have been considered as either social dynamics (for nb. of employee attending #employee $\leq$ 5) or business dynamics (for #employee $>$ 5). Social dynamics events capture external and general events occurring in a city e.g., sport or music events while business dynamics are related to internal events of the employee's company. Such groups delimit social events from business events, which are then used for categorising explanation of targeted abnormal expenses. Finally the finite state machine is augmented with special edges connecting abnormal expenses to any similar context and potential explanation. Such edges embed a conditional probability of explaining x with e given context c.

Example 3. (Explanation of Abnormal Expenses) *Events captured from external data sources have shown that, during the week of August 5th, 2015, there have been many events in the city of Austin, for example, there were* 30 *conferences,* 2 *concerts, and many minor events. Not all events are explanation of the abnormal expense in Example 2. Following approach illustrated in Figure 6 only events, with (semantically) similarities of past events causing abnormal expenses in (semantically) similar cities, are shortlisted and evaluated as contributing events. Level-1 and Level-2 explanations are extracted from the semantic types (e.g., social dynamics) and sub-types (e.g., music) of the events shortlisted.*

• ***Approach (Business UI)***: Figure 5 captures the explanations component of AIFS. This functionality is activated when an abnormal expense has been selected from the data analysis part of AIFS cf. red-highlighted row /

abnormal expense ⑨ in Figure 4. In addition to a spatial representation of expenses (zoomed-in version of ③ in Figure 4 cf. ① in Figure 5), AIFS exposes two levels of explanation, noted respectively Level-1 and Level-2.

Level-1 (③ in blue), represented as bar charts, breaks down explanations in 5 categories: (i) behavioural which captures individual fraud by analysing recurring abnormal expenses patterns (cf. anomaly row in Table 1), (ii) social dynamics which identifies external events as main causes by analyzing the impact of events on expenses (cf. potential explanation row in Table 1), (iii) business dynamics which catches internal event to the employee's company by capturing recurrence of employees' expenses in a given city and date (cf. anomaly row in Table 1), (iv) spatial which emphasizes city capacity as core contributing factor by analysing the semantic representation of city e.g., number of accommodation types, rooms, population density (cf. semantics row in Table 1), (v) temporal which defines seasonal causes. Each of these five categories is scored in $[0, 100]$ (with min: 0 and max: 100), as a potential contributing factor of the abnormality. The sum of all contributing factors is 100. In the example of Figure 5, behavioural, business dynamics and city capacity are all 2 while time is 5 and social dynamics is 89.

Level-2 (④ in green), represented as pie charts, provides a more specific and detailed version of each of the Level-1 type e.g., type of events for external events, or city services and density for city capacity. ⑤ explicits that travel expense 93741309 is abnormal and caused by external events in the city. In particular 89% of similar abnormal expenses (and context) has been caused by similar type of external events. Level-2 details that the contributing factor of the festive music event in Austin on October 2-11, 2015 is 94%. All results in Figure 5 can be interpreted by business owners and auditors for root cause analysis and better understand how expenses are impacted by individual behaviour or external sources such as city events for accommodation prices.

• ***Approach (Technical UI)*** (Figure 7): In addition to a business UI, designed for business owner and auditors, we augmented the system with a technical UI. This additional interface has been requested to be available for exploitation by Accenture analysts, who are reporting to business owner and who are responsible for preparing reports and recommended actions to be undertaken. This interface is particularly important as it provides them rational and evidence of the reasoning. The analysts aim at extracting the rational by selecting abnormal expenses in the main UI, which automatically flagged nodes of the knowledge graph ⑦ in red in Figure 7. By clicking on those nodes, they obtained new nodes that are blinking. Such nodes are derived by the system to be (i) explanations of the anomalies, (ii) their context, and (iii) similar historical contexts - where the explanations have been retrieved from[1,2,10].

• ***Scalability***: The evaluation and computation of explanation strongly relies on classification of data augmented with semantics in Table 1. Classification, or the computation of subsumption hierarchies for classes and properties of the semantic representations, is required to quickly determine semantic matching (or similarity) of contexts (e.g., events, expense or city profile). Its scalability is ensured through a distributed classification [21] of individual contexts. All rules, which are required for classification, are distributed across various nodes based on their types. Fast processing and search is also ensures through temporal indexes of all anomalies (i.e., abnormal expenses and contexts).

• ***Limitation***: The current implementation is limited to OWL EL[11] as semantic encoding of expenses and their context for the computation of semantic similarity. The OWL 2 EL profile is designed as a subset of OWL 2 that is suitable in our context since ontology classification can be decided in polynomial time, hence scalable. The computation performance would have been strongly altered when considering much more expressive semantics such as OWL 2 Full or DL (cf. Section 6 for experimentation). The five classes of explanation remain static but semantic clustering could be investigated further to present dynamic classes, especially if new sources of potential explanation are to be considered. The on-the-fly integration of new data sources with our semantic model in Table 1 might be challenging due to non-trivial alignment of vocabularies.

4.3. Abnormal Expenses Prediction

• ***Description***: Prediction, or the problem of estimating future observations given some historical information such as spending and expenses, is an important inference task required by the spend optimization, finance process and operation teams. It is crucial for planning expenses budget ahead, but also for avoiding unnecessary spending.

• ***Motivation***: This process determines the state of future expenses, which will support the spend optimization team during the process of expenses approval e.g., by approving, raising alerts, recommending alternatives or dismissing.

[10] http://54.194.213.178:8111/IF-KnowledgeGraph-Mid-Moderate (optimal version in Firefox, Safari - visual representation not encoded in RDF/OWL for optimal memory management in web browser).

[11] http://www.w3.org/TR/owl2-profiles/

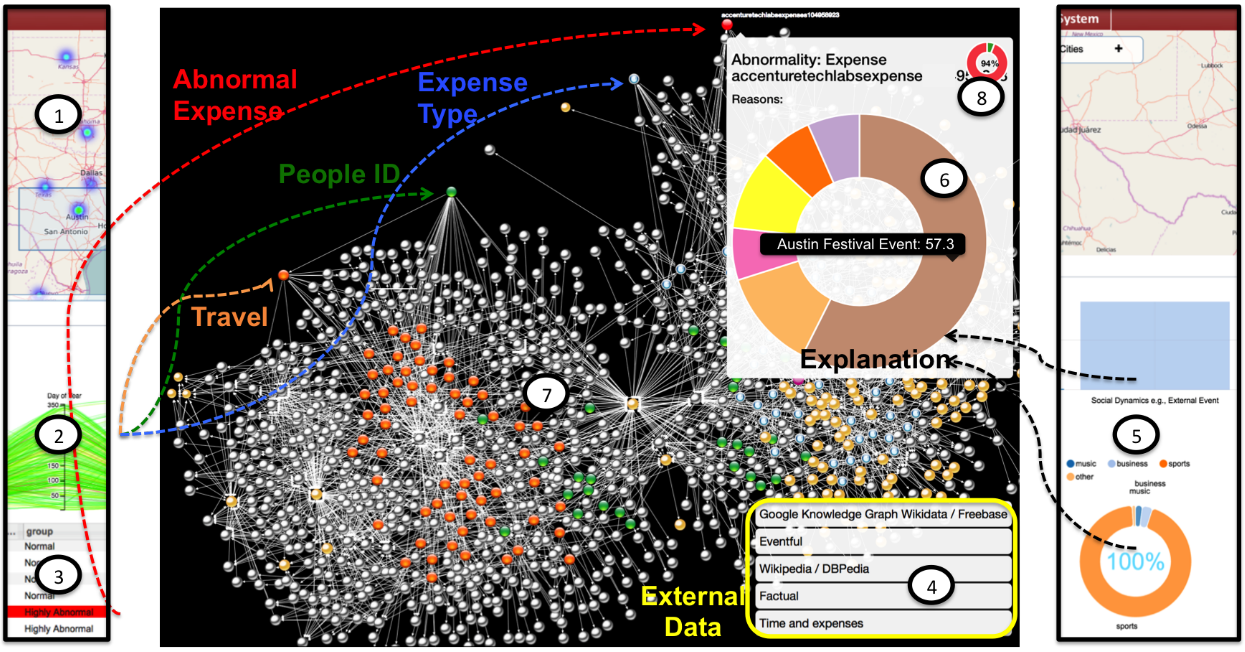

Figure 7: AIFS - (Technical) Explanation Component. ①: Spatial (Map Area) context of abnormal expenses (extract of Figure 4). ②: Individual expense status and profile i.e., type, amount, duration, date, employee id, and career level (extract of Figure 4). ③: Abnormal expense to be explained (extract of Figure 4). ④: Source of data considered for abnormality detection and explanation. ⑤: Level-1 and -2 explanation for anomaly in ③ (extract of Figure 5). ⑥: Instance-based representation of Level-1 and -2 explanation (name of the instance that explains the abnormality is given). ⑦: Graph-based representation of expenses, their type and associated semantics (rational of explanation, through blinking nodes is given by selecting abnormal expense in the graph, denoted in red). ⑧: Confidence score of Level-1 and -2 explanation.(color print).

• ***State-of-the-art Approaches***: Most of the existing systems provide limited forecast, usually based on previous spending from last quarter / year, break down per service group. They are simple rules for estimation. Mainly internal data is used for establishing some indicators, which has been shown to be inaccurate in commercial settings cf. *cost of forecasting versus cost of inaccuracy for a medium-range forecast* in Harvard Business Review [22].

• ***Challenge***: Predictive analytics spans many research fields, from Statistics, Signal Processing to Database and Artificial Intelligence. All existing predictive analytics approaches e.g., [23, 24] have been mainly designed for discovering correlation from data which is (i) numeric and (ii) very siloed i.e., context-free. Thus they rarely utilize exogenous sources of information for adjusting estimated prediction in a systemic way. City size, season, or number of events, and employee profile are examples of external factors that strongly impact travel expenses. All approaches do not aim at using extensive contextual data, specially when data is characterized by texts or various types of semantics (e.g., event categories). Therefore existing approaches and models do not fully exploit semantics of data, and reach to sub-optimal accuracy of prediction as demonstrated in our experimental results (Section 6).

• ***Approach (Technical)*** (Algorithm 2): AIFS shows that the integration of heterogenous and exogenous data is a way forward to improve accuracy and consistency of abnormal expenses prediction. Algorithm 2 sketches the approach of predicting (temporal) abn$\mathcal{O}$rmal expenses $\mathcal{O}_m^n$ at point of time $j \in [m, n]$ using (temporal) $\mathcal{P}$otential contextual data such as events $\mathcal{P}_m^n$, all using semantic representations. The approach mainly consists of:

(i) auto-correlation of data on a time basis (lines 5, 6) for retrieving all similar past events;

(ii) semantic association rules mining (lines 7-9) for inferring correlation and rules between past events and abnormal expenses over time e.g., rule (1-7) extended with temporal dimension t such that x, t and ev, t are substitute variables of x and ev;

(iii) validation of prediction results in $\mathcal{O}_m^n(j)$ by analyzing its semantic consistency i.e., checking no conflict of knowledge. An average accommodation price for more than 20% of nights in a large multiple-day event such as a music festival is an example of conflicting knowledge.

All prediction rules are extracted through association

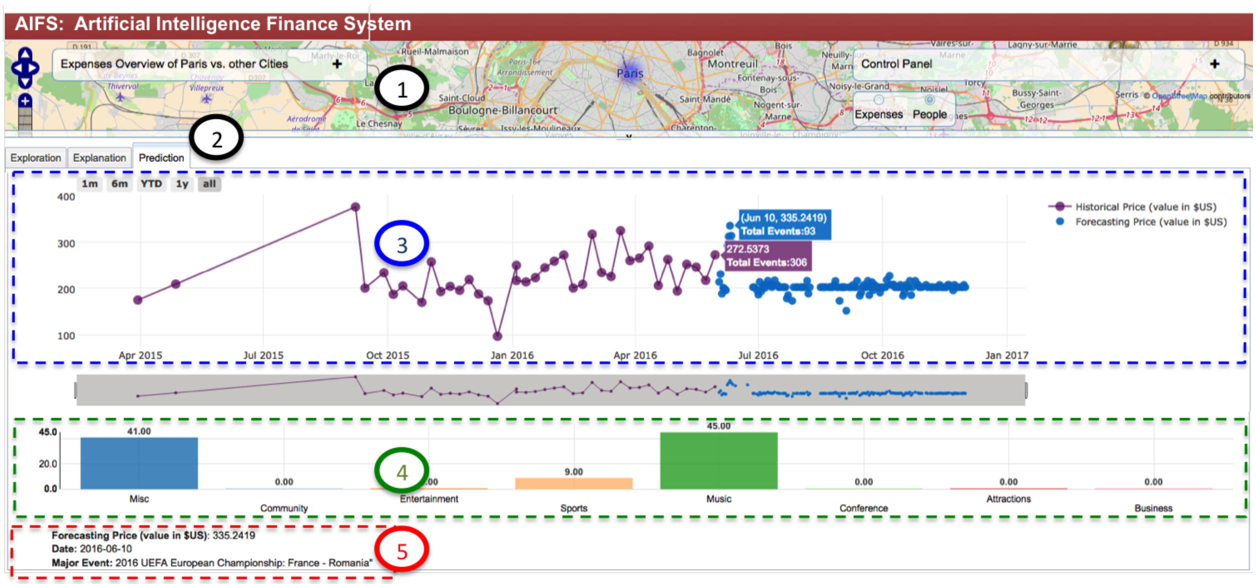

Figure 8: AIFS - Prediction Component. ①: Spatial (Map Area) context (subject to user selection) for spatial interpretation of abnormal expenses. ②: Tab-based selection of reasoning: explanation and prediction. ③: Historical (purple color) and forecast (blue color) average expenses amount in USD per employee / day. ④: Category and type of events occurring during the day selected in part ③. ⑤: Explanation (also defined as context in Algorithm 2) of abnormal forecast expense. (color print).

mining of semantic descriptions (across temporal data). Then they are filtered based on their occurrence (i.e., support) and confidence in line 9. The rule (1), extended with temporal dimension, is one example of such semantic rules. Similarly to rules learnt in Algorithm 1 their inference and consistency validation is achieved at semantic level. Their semantics is then associated with knowledge from other sources e.g., city context, type of events to infer weighted recurring rules over time. Contrary to Algorithm 1, all consistent rules persist in the background knowledge (line 12).

• ***Approach (UI)*** (Figure 8): Figure 8 illustrates how predictions are handled in AIFS. The future status of expenses and their respective amount (③ in blue) are captured. Historical expenses are captured from April 2015 to June 3rd 2016 (③ in purple) and forecast (③ in blue) up to 6 months ahead. June 10th captures an average expense amount of USD 335.34 in Paris while the average is close to USD 191. ④ emphasizes the context, and in particular the types of event that occur on any past of future day selected in ③. In our example respectively 45% and 41% are music and miscellaneous events. The information is then interpreted following Algorithm 2, in order to derive explanation of such prediction (⑤ in red). While music is the most representative type of events occurring in Paris on June 10th, our system manages to capture: *2016 UEFA European Championship opening game* as main contributing factor of high price of accommodation.

• ***Scalability***: Similarly to the explanation component, the scalability of predictive reasoning is highly coupled with the polynomial-time characteristics of subsumption-based reasoning in OWL 2 EL. Subsumption and classification are achieved to derive the consistency of prediction in line 12 of Algorithm 2. Scalability was not an issue in our context, but could be with more employee and external data. The latter can be the case with (i) larger companies and (ii) more contextual information such as detailed description of expense type.

Example 4. (Context-aware Anomaly Prediction) *We show how Algorithm 2 applies in the context of the aforementioned example expense and rules. First, from the rules mined for the context of the example, e.g., in Austin during some week with 30 conferences, the algorithm compares and collects all similar rules from different context, e.g., in Austin during a different week, illustrated below.*

$$
\begin{aligned}
&AbnormalExpense(x) \leftarrow \\
&\quad inCity(x, c) \wedge \\
&\quad (Expense \sqcap \exists type.Accommodation \sqcap \exists inWeek.\{w\} \\
&\quad \sqcap \exists amount.(\exists\ moreThan.90\%\_Cont.\_Exp.))(x) \wedge \\
&\quad events(c, \{w\}, nbConcerts) \wedge nbConcerts \geq 5 \qquad (9)
\end{aligned}
$$

By Algorithm 2, the rule (9)*, together with rules pre-*

Algorithm 2: Context-aware Anomaly Prediction

Input: (i) Temporal an*O*maly data O_m^n (from time m to n) to be predicted (e.g., abnormal expense), (ii) Temporal $\mathcal{P}$otential contextual data $\mathcal{P}_m^n$ (e.g., event), (iii) Point of prediction time $j \in [m, n]$, (iv) Min. threshold of prediction rule support m_s, confidence m_c.
Result: $O_m^n(j)$: Anomaly predicted at point of time j.
begin
  $\mathcal{R} \leftarrow \emptyset$; *% Initialization of prediction rules set.*
  % Auto-correlation of contextual information.
  $\tilde{\mathcal{P}}_m^n \leftarrow$ all similar contexts of $\mathcal{P}_m^n(j)$ in $[m, n]$;
  % Semantic association rules between $\tilde{\mathcal{P}}_m^n$, O_m^n.
  foreach *rule* $\rho \in \tilde{\mathcal{P}}_m^n(k) \times O_m^n(k)$, $\forall k \in [m, n]$ **do**
    if $support(\rho) > m_s \wedge confidence(\rho) > m_c$ **then** $\mathcal{R} \leftarrow \mathcal{R} \cup \{\rho\}$;
  % Semantic evaluation of rule $\rho \in \mathcal{R}$ *at time* j.
  foreach $\rho \in \mathcal{R}$ *of the form* $\mathcal{G} \twoheadrightarrow h$ **do**
    if h *in semantically consistent with* $O_m^n(j)$ **then** Apply rule ρ at point of time j of O_m^n;
  return $O_m^n(j)$;

viously derived, will be used for predicting abnormal expenses in some new temporal dataset.

• ***Limitation***: The integration of new data, as context, needs a careful analysis of historical data in order to identify the most appropriate knowledge to be correlated with. The automated integration of relevant data is an open problem, and it has been addressed by semi-automatically injecting pre-determined data sources for correlation and prediction. Our approach has strong dependencies with the selection of (i) context, (ii) configuration of Algorithm 2 e.g., min. threshold, support, confidence, which all impact precision.

5. AIFS Technologies

This section sketches the main technologies behind AIFS and focuses on its innovative reasoning parts, and the web-based application.

5.1. Semantic Representation

The model we consider to represent semantics of data (abnormal expenses and potential explanation in Table 1) is provided by an ontology, encoded in OWL 2 EL. The selection of the W3C standard OWL 2 EL profile has been guided by (i) the expressivity which was required to model semantics of data in Table 1, (ii) the scalability of the underlying basic reasoning mechanisms we needed (cf. scalability challenges of explanation and prediction) e.g., subsumption in OWL 2 EL is in PTIME [25]. Semantic technologies were used to compare and evaluate different context e.g., events (and their properties: venue, category, size, types / subtypes), city information (high / moderate / low population, density; good / moderate / bad level of accommodation). More importantly they were required for (automatically) learning, applying rules at reasoning time for analysis, explanation and prediction components. All interfaces of AIFS produce and consume semantic representation of data. All interactions of AIFS are possible because of the semantic engine, which runs behind the scene. For instance, explanation and prediction are only possible if the underlying data is described with semantics.

5.2. Semantic Enrichment

All raw data in Table 1 is served as temporal OWL 2 EL ontologies (i.e., temporal representation of semantic-encoded data) [3] by extending Jena TDB[12] with temporal representation of entities. The DBpedia and wikidata vocabularies have been used for cross-referencing entities, and established a complete knowledge graph to cover our domain. Different mapping strategies are used depending on the data format. For instance XSLT for XML, custom OWL 2 EL mapping for CSV and JSON have been used. All the temporal EL ontologies have the same static background knowledge to capture time (W3C Time Ontology[13]) and space (W3C Geo Ontology[14]) but differ only in some domain-related vocabularies e.g., abnormality level, employees' characteristics, event type and city information. We used an OWL2 EL compatible variant of the OWL Time (W3C Time Ontology[15]) as our approach does not need the full $\mathcal{SHOIN}(\mathcal{D})$ expressivity of the W3C version. For instance, we do not need the *ProperInterval* in OWL Time, and temporal reasoning is mainly achieved to detect anteriority and posteriority of temporal statements and axioms. These ontologies have been mainly used for enriching raw data (and its context), facilitating its integration, comparison and matching over time.

5.3. Knowledge Extraction

The main challenges were (i) heterogenous data format, (ii) difficulty to re-use existing vocabularies for data description. We addressed them by using semantic representation and defining our own vocabularies for integration (example[16]). They have been linked with core

[12] http://jena.apache.org/documentation/tdb/index.html
[13] http://www.w3.org/TR/owl-time/
[14] http://www.w3.org/2003/01/geo/
[15] http://www.w3.org/TR/owl-time/
[16] http://54.194.213.178:8111/ExplanatoryReasoning/ontology/categories.n3

entities of DBpedia, wikidata for extracting knowledge i.e., context of abnormalities.

5.4. Distributed Semantic Reasoning

The matching-based computation of context similarity, which is crucial in explanation and prediction components of AIFS, is ensured by semantic classification of temporal ontologies. Such classification is achieved by distributing all the standard completion OWL 2 EL rules [26] across various nodes based on their types. Each node is dedicated to at most one type of axioms and runs its appropriate rules. OWL2 EL compatible variants of all ontologies which are beyond $\mathcal{EL}^{++}$, for instance W3C Time ontology, have been elaborated to ensure that semantic classification can be achieved following [21].

5.5. Lightweight Temporal Semantic Reasoning

Temporal evolution of knowledge is represented as ontology OWL2 EL streams [3] i.e., dynamic and evolutive version of ontologies [27]. Data (ABox), its inferred statements (entailments) are evolving over time while its schema (TBox) remains unchanged. We use DL $\mathcal{EL}^{++}$ formalization in the following for illustration purpose.

Example 5. ***(DL $\mathcal{EL}^{++}$ Ontology Stream)***
Figure 9 illustrates 3 *partial $\mathcal{EL}^{++}$ streams $\mathcal{P}_0^n$, $\mathcal{Q}_0^n$ and $\mathcal{R}_0^n$, related to events, expenses and their location through snapshots at point of time $i \in \{0, 1, 2\}$ (i.e., a view on window $[0, 3]$). In our example n is any integer greater than* 2 *and time $i \in \{0, 1, 2\}$ can be seen as* 3 *consecutive days of a year. Their dynamic knowledge is captured by evolutive ABox axioms e.g.,* (10) *captures e_1 as "a social poetry event occurring in r_2" at time* 0 *of $\mathcal{P}_0^n$.*

$$\mathcal{P}_0^n(0) : (SocialEvent \sqcap \exists type.Poetry)(e_1) \quad (10)$$
$$\mathcal{Q}_0^n(0) : (Expense \sqcap \exists type.Accommodation \sqcap \exists cost.Low)(a) \quad (11)$$
$$\mathcal{R}_0^n(0) : expenseInCity(a, Dublin) \quad (12)$$
$$\mathcal{P}_0^n(1) : (SocialEvent \sqcap \exists type.Movie)(e_2) \quad (13)$$
$$\mathcal{Q}_0^n(1) : (Expense \sqcap \exists type.Accommodation \sqcap \exists cost.Low)(a) \quad (14)$$
$$\mathcal{R}_0^n(1) : expenseInCity(a, Dublin) \quad (15)$$
$$\mathcal{P}_0^n(2) : (SocialEvent \sqcap \exists type.Concert)(e_3) \quad (16)$$
$$\mathcal{Q}_0^n(2) : (Expense \sqcap \exists type.Accommodation \sqcap \exists cost.High)(a) \quad (17)$$
$$\mathcal{R}_0^n(2) : expenseInCity(a, Dublin) \quad (18)$$

Figure 9: Ontology Streams $\mathcal{P}_0^n(i)$, $\mathcal{Q}_0^n(i)$ and $\mathcal{R}_0^n(i)_{i\in\{0,1,2\}}$.

Semantic comparison and matching of temporal ontology at one point of time are operated through lightweight temporal reasoning. Such computing is required by predictive reasoning, and explanation for elaborating semantic context (events, employees' profile) similarity and correlation over time. In more details the temporal ontology correlation is established by comparing the number of changes i.e., *new*, *obsolete*, *invariant* ABox axioms and entailments between various evolution e.g., number and types of events that change among two different temporal instances. Definition 1 provides basics, through ABox entailments, for understanding how knowledge is evolving over time. This is unique to AIFS for establishing context-aware explanation and prediction.

Definition 1. ***(ABox Entailment-based 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 (19)*, obsolete* (20)*, invariant* (21).

$$\mathcal{G}_{new}^{[\alpha],[\beta]} \doteq \{g \in \mathcal{G} \mid \mathcal{T} \cup \mathcal{S}_0^n[\beta] \models g \ \wedge \mathcal{T} \cup \mathcal{S}_0^n[\alpha] \not\models g\} \quad (19)$$
$$\mathcal{G}_{obs}^{[\alpha],[\beta]} \doteq \{g \in \mathcal{G} \mid \mathcal{T} \cup \mathcal{S}_0^n[\beta] \not\models g \ \wedge \mathcal{T} \cup \mathcal{S}_0^n[\alpha] \models g\} \quad (20)$$
$$\mathcal{G}_{inv}^{[\alpha],[\beta]} \doteq \{g \in \mathcal{G} \mid \mathcal{T} \cup \mathcal{S}_0^n[\beta] \models g \ \wedge \mathcal{T} \cup \mathcal{S}_0^n[\alpha] \models g\} \quad (21)$$

The symbol $\models$ captures logical entailment with respect to terminological axioms $\mathcal{T}$ and some assertional axioms from $\mathcal{S}_0^n$. (19) reflects knowledge we gain by sliding window from $[\alpha]$ to $[\beta]$ while (20) and (21) denote respectively lost and stable knowledge. All duplicates are supposed removed.

Example 6. ***(ABox Entailment-based Changes)***
Table 2 illustrates changes occurring from $(\mathcal{Q}\cup\mathcal{R})_0^n[0, 1]$ to $(\mathcal{Q} \cup \mathcal{R})_0^n[2, 2]$ through ABox entailements. For instance "a as a high cost accommodation in window $[2, 2]$ of $(\mathcal{Q} \cup \mathcal{R})_0^n$ is new with respect to knowledge in $[0, 1]$.

Windowed Stream Changes	$(\mathcal{Q}\cup\mathcal{R})_0^n[2,2]\ \nabla\ (\mathcal{Q}\cup\mathcal{R})_0^n[0,1]$		
	obsolete	*invariant*	*new*
$expenseInCity(a, Dublin)$		✓	
$(Expense \sqcap \exists type.Accommodation \sqcap \exists cost.Low)(a)$	✓		
$(Expense \sqcap \exists type.Accommodation \sqcap \exists cost.High)(a)$			✓

Table 2: ABox Entailment-based Changes.

5.6. Semantic Rule Association and Mining

Predictive reasoning is achieved following state-of-the-art principles i.e., association rules mining to learn rules for classification (in the context of abnormal expenses identification and prediction). The generation of association rules between temporal ontology is based on a semantic extension [3] of Apriori [28], aiming at supporting subsumption for determining association rules. Contrary to the initial version of Apriori which infers associations between data instances, the association in AIFS is achieved between their descriptions e.g., type of events, expenses, context. Rules are encoded using DL $\mathcal{EL}^{++}$ rules, and all consequents of each rule are validated though consistency checking (cf. line 12 in Algorithm 2). This ensures to obtain consistent and accurate prediction results (cf. Experimental Results). Data related to (a) expense (8 features e.g., type, season, amount, country, city), (b) city (15 features e.g., population, country, density, hotel rooms), (c) employee profile (9 features e.g., career level, industry, years of experience), (d) social event (12 features e.g., country, city, type, season, attendance), and (d) news event (6 features e.g., type, country, city) are considered to capture rules.

5.7. Web-based Application

• *REST Interface*: All functionalities of AIFS are exposed through REST services, providing highly component-ization, evolve-ability via loose coupling and hypertext.
• *Web User Interface*: AIFS strongly relies on HTML, CSS, Javascript (Dojo toolkit, D3, JQuery libraries) to produce an appealing user interface. Time-series, spider, bar, pie charts together with parallel charts are examples where Dojo and D3 components were combined with HTML and CSS.
• *Deployment*: Our technology stack is based on well-established open source and commercial components such as Apache Tomcat as the HTTP/Application Server and (ii) state-of-the-art components such as OpenLayers[17] as an open source JavaScript library for displaying dynamic map data, pssh for parallel distribution of reasoning, Jena TDB as RDF store (extended with temporal indexes). A B+ Trees TDB indexing structures has been used, which turns to scale better in our context of temporal / (minimal) dynamic updates.

6. Experimental Results

We focus on the scalability and accuracy of the results that AIFS delivers, which have been raised as the most important metrics by our business users. In particular we highlight the explanation and predictive reasoning, as the most critical and resource consuming components of AIFS. Experiments were run on a server of 16 Intel(R) Xeon(R) CPU E5-2680, 2.80GHz cores and 32GB RAM.

[17]http://openlayers.org/

6.1. Open Data Context

Data in Table 1, transformed in OWL/RDF (Table 3), is used to experiment AIFS. A sample version of the OWL/RDF graph is maintained for demo purpose[9]. Explanation and predictive functionalities are experimented on a basis of 429 days (from April 1^{st} 2015 to June 3*rd* 2016) with temporal datasets [a], [d] and [e]. The semantic representation is OWL EL in all experimentation unless specified differently cf. experimentation related to expressivity in Figures 12-11.

Data	Temporal Update	Raw Update Size (KB)	Semantic Update		Semantic Conversion (ms)
			Size (KB)	#RDF Triples	
[a] Expenses	fortnightly	$\approx 10^6$	$\approx 10^8$	$\approx 10^9$	$\approx 10^7$
[b] City (average)	-	-	634	1,189	-
[c] Employee Profile	-	$\approx 10^7$	$\approx 10^9$	$\approx 10^{10}$	$\approx 10^8$
[d] Social Events	daily	240.7	297	612	0.681
[e] News Events	daily	8356.9	1295	7177	0.891

Table 3: Datasets Details (average figures). [b,c] are not temporal. [c] is pre-encoded in RDF using DBpedia, wikidata.

All data sets [a,b,c,d,e] in Table 3 is considered for explanation computation (Section 6.2) as they all equally contribute for elaborating explanations. The importance of data set [e] in Table 3 for the prediction task (Section 6.3) is not as significant as for the explanation task. Prediction accuracy only gains 4% by adding data set [e]. Therefore we only consider the data set [a,b,c,d] for prediction experimentation.

6.2. Explanation Experimentation (with semantic representation of data [a,b,c,d,e] in Table 3)

• ***Scalability***: Figure 10 presents the computation time for explaining an abnormal expense by varying the context (indirectly the underlying model) i.e., number of historical $|H|$ and real-time $|R|$ atomic abnormal expense occurring per day (approx. $1,135$). H and R are respectively expenses explained, and to be explained.

The historical modelling of the semantic explanation model (cf. finite state machine in Figure 6) is strongly impacted by the number of historical expenses while the real-time explanation remains constant. The modelling part compiles all explanation results and evaluates semantic similarity between H and R together with their context (after distributed semantic classification) while real-time explanation consists in retrieving the most relevant ones. $10,539,891$ abnormal expenses have been detected out of $1,335,691,105$ items for our time period of 429 days. The computation of the off-line explanation model required approximately 286 minutes. The

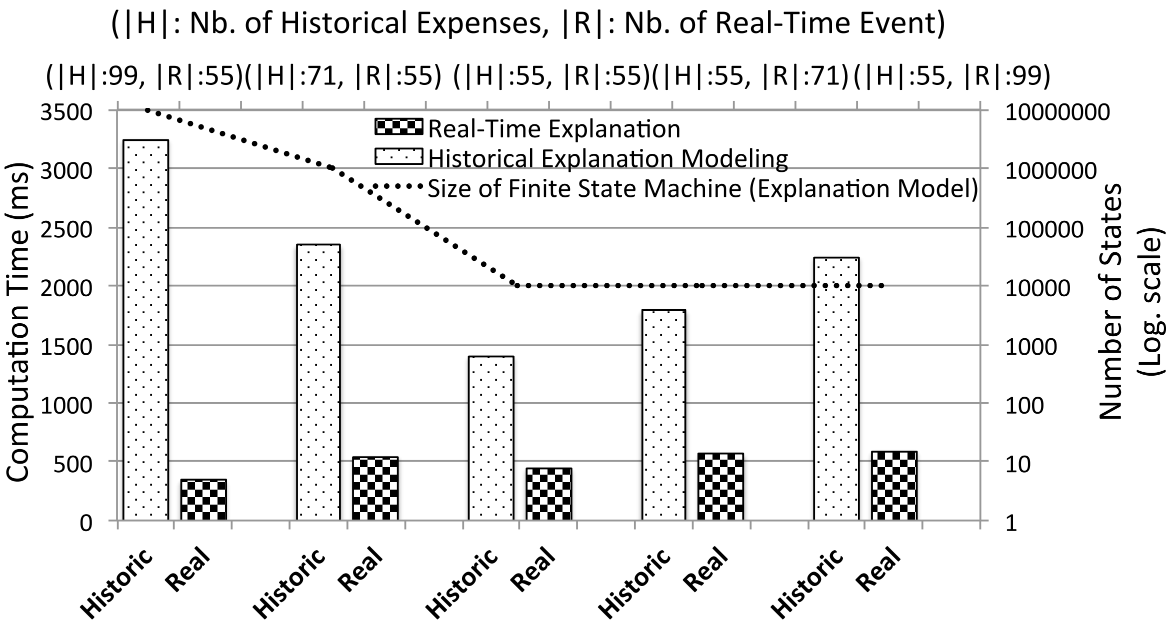

Figure 10: Scalability of Explanation Computation using 3 Different Partial Explanation Models (cf. $|H|$ as indicator).

average computation time for retrieving a unique explanation is 0.982 seconds using the complete model.

• ***Accuracy***: Table 4 depicts the precision and recall of returned explantion results by varying the size of the model ($|H|$ at 100% of the model).

Ratio of Explanation Model $|H|$ (%)	10	20	30	40	50	60	70	80	90	100
Precision (%)	22	29	32	41	54	61	68	73	83	89
Recall (%)	14	23	29	31	41	48	53	62	71	78

Table 4: Accuracy of Explanation.

We report the ratio of up to 10 explanation results identified and compared against those estimated by business experts (used as ground truth). The more historical data the better accuracy. Such result confirms the importance of capturing historical context of past abnormal expenses.

• ***Expressivity***: Figure 12 reports the scalability of the approach by varying the expressivity of the underlying background ontology (which is a strong indicator for evaluating subsumption-based matching of expenses, contexts and explanations). We consider our baseline model OWL EL against other with more expressivity OWL RL, OWL $\mathcal{SROIQ}(\mathcal{D})$ and less expressivity RDF/S. We used a distributed version of (i) CEL [26] for OWL EL ontologies, and (ii) TrOWL [29] for contexts with more expressive representation.

We break down the experiment per region in the world to evaluate the impact of the size of data. North America NA and Europe E are the regions where occur most of the expenses, while the expenses in South America SA and Asia A are less significant. In other words $|NA| > |E| > |SA| > |A|$ with $|X|$ expenses in region X. All elements of semantic computation are tracked: data transformation, OWL / RDF loading in Jena, anomaly detection and explanation reasoning. Unsurprisingly the system is the most scalable with RDF/S and the least with OWL $\mathcal{SROIQ}(\mathcal{D})$. Scalability is impacted by the pair: data size and expressivity. More interestingly the difference between OWL EL and RDF/S is not significant in this context, and the number of learnt rules do not change across the various models. In all cases, the explanation part is the most time consuming while loading is the least.

Figure 11 reports the accuracy of the approach by varying the above levels of expressivity. Unsurprisingly the system is the most accurate with OWL $\mathcal{SROIQ}(\mathcal{D})$ and the least with RDF/S. Similarly to scalability, accuracy is impacted by the pair: data size and expressivity. Very interestingly, although OWL EL is significantly less expressive than OWL $\mathcal{SROIQ}(\mathcal{D})$, it only lost 4.7% of accuracy (on average) against the OWL $\mathcal{SROIQ}(\mathcal{D})$.

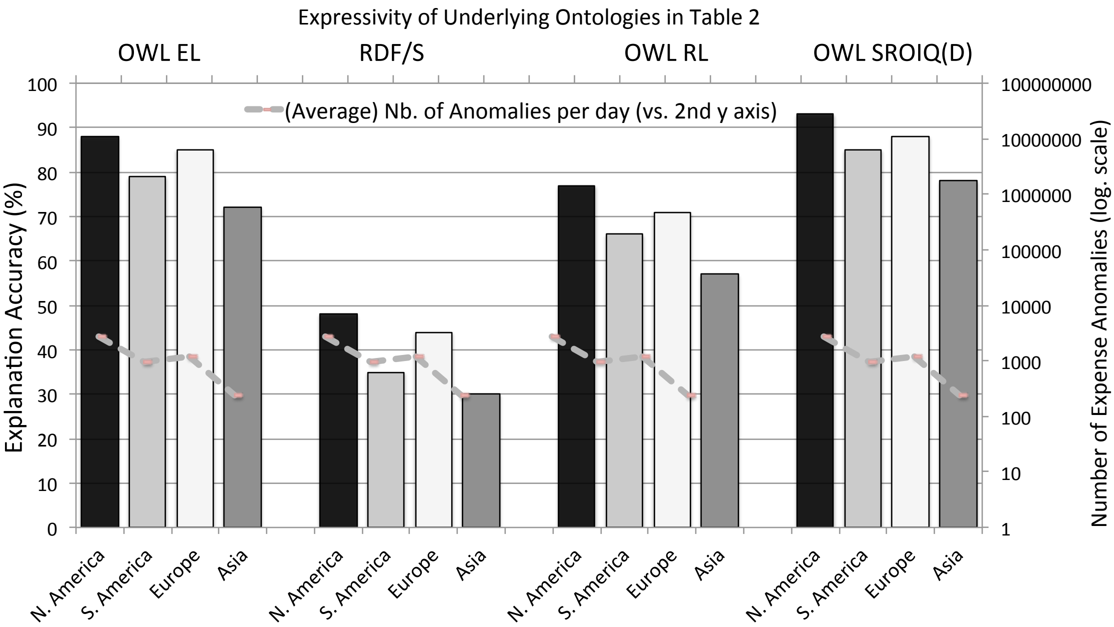

Figure 11: Impact of Semantic Expressivity on Accuracy.

• ***Lessons Learned***: Reducing the number of historical expenses (as baseline for building the explanation model) decreases the computation time, but also decreases accuracy. The more similar historical expenses and associated context the higher the probability to catch accurate explanation. The computational performance of our approach is mainly impacted by the expressivity of the semantics as it impacts both semantic classification and similarity computation. OWL EL has been demonstrated to have the best tradeoff expressivity, scalability and accuracy. The experimentation strongly confirms our choice of using OWL EL as representation model.

6.3. Prediction Experimentation (with semantic representation of data [a,b,c,d] in Table 3)

The objective is to predict abnormal travel related expenses in the next 6 months using contextual information. The evaluation is achieved using different contextual information i.e., [a], [a,b], [a,c], [a,d], [a,b,c], [a,b,d], [a,b,c,d] in Table 3, to evaluate their impacts on scalability and accuracy.

• ***Scalability*** : Figure 13 reports the scalability of our approach, noted AA, and compares its computation time with a state-of-the-art approach [11] in predictive analytics.

Contrary to our approach, similarity is detected at raw data level using (i) statistics-based data analysis and (ii) mathematical properties of the temporal data. [11] scales better than our approach in all contextual configurations. Our approach requires some non-negligible computation time for reasoning on top of the semantics-enriched data. The identification of significant rules is strongly impacted by the number of potential rules, which grows exponentially with the number of semantic representations of raw data (secondary vertical axis). Once all rules are identified, consistent prediction is delivered from $1.8s$ to $2.9s$.

• ***Accuracy***: Figure 14 reports the prediction accuracy of both approaches [11] and AA. The accuracy is measured by comparing predictions (level of abnormality) with real-time expenses in respective cities (when expenses are available). All results can be easily extracted and compared from the raw and semantic data in respectively [11] and our approach. The more contexts the better the accuracy of prediction for both approaches. However our approach reaches a better accuracy when text-related context [b,c,d] are jointly interpreted. On contrary [11] cannot take much benefit of their semantics. Overall, our approach obtains a better accuracy, mainly because all the rules are pruned based on the consistency of their consequent. By enforcing their consistency, we ensure that rules are selected based on the surrounding context i.e., exogenous data. The semantic enrichment of data is then for correlating, cross-associating and then predicting abnormal expenses on a common basis.

• ***Expressivity***: The results related to the impact of ex-

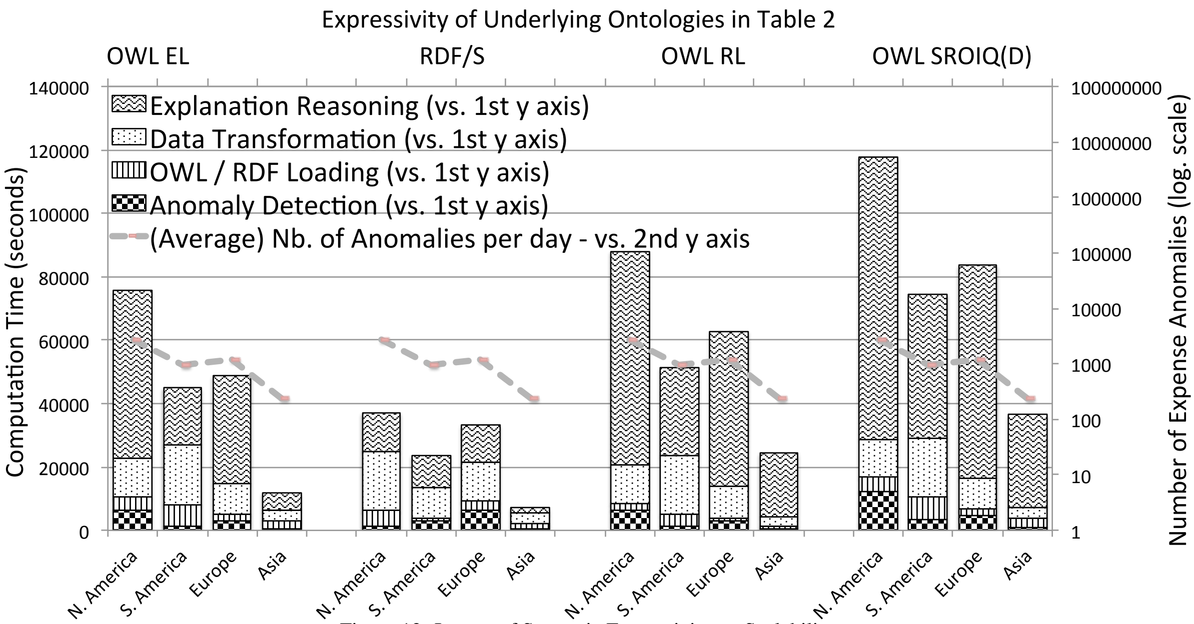

Figure 12: Impact of Semantic Expressivity on Scalability.

pressivity on predictive reasoning are aligned with the ones for Figures 12-11 i.e., the more expressive the less (more) scalable (accurate). Experimentation has shown that size of data has a larger impact than expressivity on accuracy of the system. The OWL EL model lost (on average) 6.9% of accuracy against the OWL $\mathcal{SROIQ}(\mathcal{D})$.

• ***Lessons Learned***: Our experimental results emphasize the advantage of using semantic Web technologies for abnormal expenses prediction i.e., accuracy, but also point out the scalability limitation, specially compared to pure statistical approaches. In particular the more contexts the more rules which positively (resp. negatively) impacts accuracy (resp. scalability). Since state-of-the-art approaches fail to encode text-based data, they simply fail to interpret their semantics. On contrary, our approach interprets their semantics to enrich the prediction model, ensuring more accurate prediction. OWL EL has shown to have the best tradeoff of expressivity, accuracy and scalability.

7. Conclusion

This paper presented AIFS, an innovative and integrated system which has been designed for (i) seamlessly aggregating heterogeneous and exogenous data and more importantly (ii) delivering integrated contextual analysis, explanation and prediction of abnormal expenses from $191,346$ unique Accenture employees, while (iii) being scalable to any large organisations using various types contexts through semantic web technologies. AIFS delivers insight to interpret any-time abnormal expenses, making expenses easier to be managed and supporting spend optimization. Thus AIFS supports both business owner to establish new policy and auditors to (i) better pinpoint causes of abnormal expenses, and (ii) minimize communication with employees. The experiments have shown scalable, accurate, consistent explanation and prediction of abnormal expenses, which are the main benefits of the semantic encoding and underlying reasoning.

Handling automated parameters configuration (Algorithms 1, 2), data summarization, flexible data integration (cf. *Limitation* sections) are future domains of investigation. In addition the integration of reinforcement learning in our explanation and prediction component would benefit accuracy.

References

[1] M. Enoch, Sustainable transport, mobility management and travel plans, Routledge, 2012.

[2] F. Lécué, Diagnosing changes in an ontology stream: A DL reasoning approach, in: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Ontario, Canada., 2012.
URL `http://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/4988`

[3] F. Lécué, J. 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, pp. 2662–2669.
URL `http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6608`

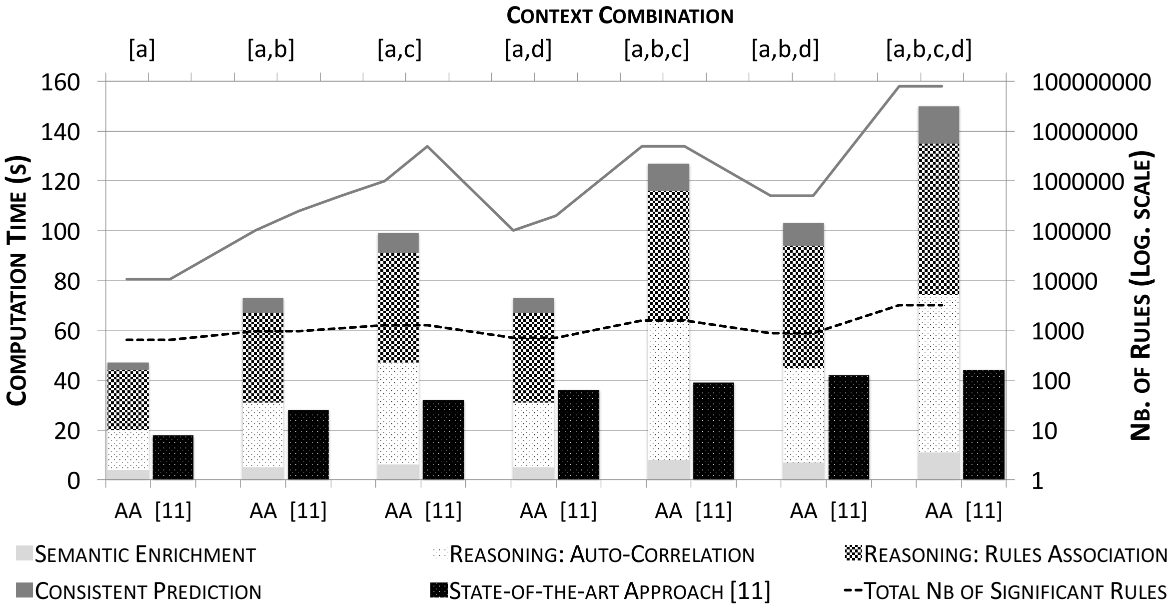

Figure 13: Scalability of Prediction (with different contexts).

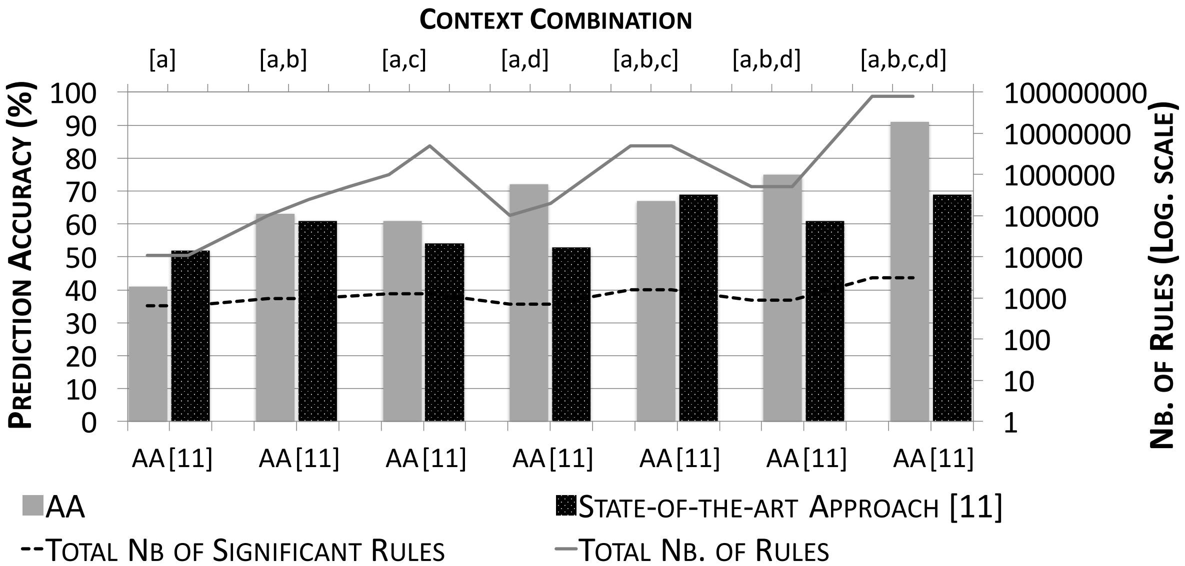

Figure 14: Accuracy of Prediction (with different contexts).

[4] R. Mutharaju, Very large scale owl reasoning through distributed computation, in: International Semantic Web Conference (2), 2012, pp. 407–414.

[5] F. Lécué, J. Z. Pan, Consistent knowledge discovery from evolving ontologies, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA., 2015, pp. 189–195.
URL `http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9518`

[6] C. Luo, H. Thakkar, H. Wang, C. Zaniolo, A native extension of sql for mining data streams, in: SIGMOD Conference, 2005, pp. 873–875.

[7] S. Babu, J. Widom, Continuous queries over data streams, SIGMOD Record 30 (3) (2001) 109–120.

[8] J. Gehrke, F. Korn, D. Srivastava, On computing correlated aggregates over continual data streams, in: SIGMOD Conference, 2001, pp. 13–24.

[9] M. J. Zaki, Spade: An efficient algorithm for mining frequent sequences, Machine Learning 42 (1/2) (2001) 31–60.

[10] S. Papadimitriou, J. Sun, C. Faloutsos, Streaming pattern dis-

covery in multiple time-series, in: VLDB, 2005, pp. 697–708.
[11] M. Robnik-Šikonja, Improving random forests, in: European Conference on Machine Learning, Springer, 2004, pp. 359–370.
[12] T. Teijeiro, P. Félix, On the adoption of abductive reasoning for time series interpretation, arXiv preprint arXiv:1609.05632.
[13] D. L. McGuinness, A. Borgida, Explaining subsumption in description logics, in: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 95, Montréal Québec, Canada, August 20-25 1995, 2 Volumes, 1995, pp. 816–821.
URL `http://ijcai.org/Proceedings/95-1/Papers/105.pdf`
[14] P. Thagard, C. Shelley, Abductive reasoning: Logic, visual thinking, and coherence, in: Logic and scientific methods, Springer, 1997, pp. 413–427.
[15] M. T. Ribeiro, S. Singh, C. Guestrin, "why should I trust you?": Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016, 2016, pp. 1135–1144. doi:10.1145/2939672.2939778.
URL `http://doi.acm.org/10.1145/2939672.2939778`
[16] B. Taskar, V. Chatalbashev, D. Koller, C. Guestrin, Learning structured prediction models: A large margin approach, in: Proceedings of the 22nd international conference on Machine learning, ACM, 2005, pp. 896–903.
[17] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, A. Bouchachia, A survey on concept drift adaptation, ACM Comput. Surv. 46 (4) (2014) 44:1–44:37. doi:10.1145/2523813.
URL `http://doi.acm.org/10.1145/2523813`
[18] L. Li, I. Horrocks, A software framework for matchmaking based on semantic web technology, in: Proceedings of the Twelfth International World Wide Web Conference, WWW 2003, Budapest, Hungary, May 20-24, 2003, 2003, pp. 331–339. doi:10.1145/775152.775199.
URL `http://doi.acm.org/10.1145/775152.775199`
[19] M. Krötzsch, F. Maier, A. Krisnadhi, P. Hitzler, A better uncle for OWL: nominal schemas for integrating rules and ontologies, in: Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011, 2011, pp. 645–654. doi:10.1145/1963405.1963496.
URL `http://doi.acm.org/10.1145/1963405.1963496`
[20] M. Paolucci, T. Kawamura, T. Payne, K. Sycara, Semantic matching of web services capabilities, in: Proceedings of the First International Semantic Web Conference, LNCS 2342, Springer-Verlag, 2002, pp. 333–347.
[21] R. Mutharaju, P. Hitzler, P. Mateti, F. Lécué, Distributed and scalable OWL EL reasoning, in: The Semantic Web. Latest Advances and New Domains - 12th European Semantic Web Conference, ESWC 2015, Portoroz, Slovenia, May 31 - June 4, 2015. Proceedings, 2015, pp. 88–103.
[22] J. C. Chambers, S. K. Mullick, D. D. Smith, How to choose the right forecasting technique, Harvard Business Review.
[23] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, et al., Knowledge discovery and data mining: towards a unifying framework., in: KDD, Vol. 96, 1996, pp. 82–88.
[24] J. Han, Y. Sun, X. Yan, S. Y. Philip, Mining knowledge from data: An information network analysis approach, in: 2012 IEEE 28th International Conference on Data Engineering, IEEE, 2012, pp. 1214–1217.
[25] C. Haase, C. Lutz, Complexity of subsumption in the [escr][lscr] family of description logics: Acyclic and cyclic tboxes, in: ECAI, 2008, pp. 25–29.
[26] F. Baader, S. Brandt, C. Lutz, Pushing the el envelope, in: IJCAI, 2005, pp. 364–369.
[27] Z. Huang, H. Stuckenschmidt, Reasoning with multi-version ontologies: A temporal logic approach, in: The Semantic Web - ISWC 2005, 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings, 2005, pp. 398–412.
[28] R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, in: VLDB, 1994, pp. 487–499.
[29] J. Z. Pan, Y. Ren, Y. Zhao, Tractable approximate deduction for OWL, Artif. Intell. 235 (2016) 95–155. doi:10.1016/j.artint.2015.10.004.
URL `http://dx.doi.org/10.1016/j.artint.2015.10.004`