Using Semantic Web Technologies for Explaining and Predicting Abnormal Expenses

Freddy Lécué[1,2], John Vard[1], and Jiewen Wu[1]

[1] Accenture Technology Labs, Ireland
[2] INRIA, France

Abstract. Travel expenses represent up to 7% of organisations overall budget. Existing expenses systems are designed for reporting expenses types and amount, but not for understanding how to save and spend. We present a system, manipulating semantic web technologies, which aims at identifying, explaining, predicting abnormal expense claims by employees of large organisations in 500+ cities.

1 Introduction and Motivation

Time and expense in most companies is the process of recording and tracking hours worked and expenses as they relate to projects. $546 Billion worldwide has been estimated to be lost because of non spend optimisation in the process of managing expenses at organisation level. We address the problem of *spend optimisation* i.e., *determining how to save and spend*. Our system, AIFS (**A***ccenture* **I***ntelligent* **F***inance* **S***ystem*), aims at (i) identifying, (ii) explaining and (iii) predicting abnormal expenses e.g., flight, accommodation, entertainment (denoted as abnormalities) in 500+ cities. The system is mainly used by (i) travel and expenses business owner to better manage spend optimisation, (ii) expenses auditor for tracking abnormal expenses, and (iii) internal travel system administrator for defining expenses policy based on the reasoning results.

Our system is exposing a semantic description of causation between external events and expenses type. It leverages open data (e.g., events), LOD and reasoning functionalities to expose causation using semantic representation in RDF. DBpedia and Wikidata are used by our system for representing heterogeneous data together with explanation, and then by our back-end reasoning engine for semantic association.

2 Functionalities

Our system exposes three core functionalities: abnormality (i) detection, (ii) explanation, and (iii) prediction. They cover the particular case of accommodation price variance over time in 500+ cities. All functionalities consume JSON inputs and expose back JSON, RDF and JSON-LD with contexts mapped to the LOD cloud. JSON-LD has been considered for further easy integration in any application.

2.1 Abnormality Detection

• ***Description***: This service retrieves abnormal accommodation price in a given {*city*}, {*country*} for year 2015 worldwide.

• ***Output***: Abnormal accommodation prices are detected in 500+ cities. The JSON description captures the *type* of expenses, its *date*, its severity *level* (scaled from 1 to 5 as the most severe), its *average* and *observed* amount in US dollar. The RDF description exposes semantics of the search and abnormalities. For instance, cities and countries are contextualised using their DBpedia resources.

• ***Back-end Technology***: A standard statistical analysis of data has been conducted to detect abnormal expenses. An analysis of 300,000+ unique travellers in 500+ cities worldwide with a minimum and maximal number of respectively 2,521 and 24,800 expenses per city recorded for 2015 has been initially performed.

2.2 Abnormality Explanation

• ***Description***: We retrieve semantic explanation of abnormal price of accommodation in a given $\{city\}$ and $\{country\}$ between $\{date\_start\}$ and $\{date\_end\}$.

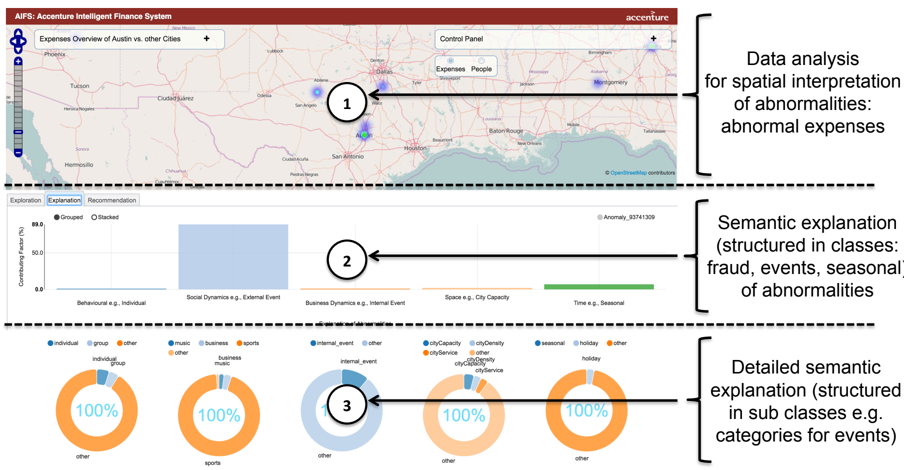

Fig. 1: Accenture Intelligent Finance System - Explanation Part. ① refers to a spatial representation of abnormalities. ② refers to an histogram representation of one level of diagnosis (not exposed in this resource). ③ refers to a pie chart representation of the level / type of explanation exposed by our system. (color).

• ***Output***: Semantic explanation, contextualised using DBpedia and Wikidata, of abnormal accommodation price is provided in 500+ cities. Contributing factors (scaled from 0 to 100), categorised as frauds, external events, city capacity, season together with their types (e.g., music category for social event) are provided c.f. Figure 1. Besides RDF schema, an internal ontology, characterising links across categories, is also used.

• ***Back-end Technology***: Abnormality explanation has been achieved by applying core principles of explanatory reasoning [1]. In a nutshell, the reasoning task of explaining anomalies (abnormal high price of accommodation) consists in interpreting the impact of external events (e.g., social events, seasonal effect). The interpretation is achieved by measuring the semantic similarity (through ontology matching) of any 4-uple of the form $< city, accommodation, event, impact >$ where city, accommodation, event and impact are defined by respectively 89, 41, 24 and 11 semantic properties using vocabularies of DBpedia and Wikidata.

2.3 Abnormality Prediction

• ***Description***: The system retrieves prediction of accommodation price and its explanation when the price is abnormally high.

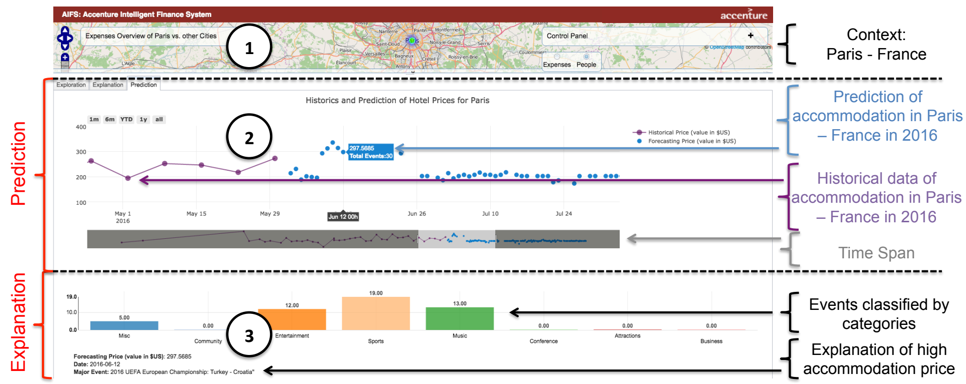

Fig. 2: Accenture Intelligent Finance Application - Prediction Part. ① refers to the spatial contextual representation: Paris, France. ② refers to a line chart capturing historical accommodation price and a scatter chart capturing predicted price. ③ refers to a histogram representation of events and their categories occurring on the selected date in ② in Paris. A detailed explanation is given in the bottom part i.e., 2016 UEFA European Championship. (color).

• ***Output***: Prediction of abnormal accommodation price is provided in 500+ cities up to two months in advance. Explanations, captured using DBpedia and Wikidata, are attached to highly abnormal accommodation price c.f. Figure 2.

• ***Back-end Technology***: Prediction of abnormality has been achieved by applying core principles of predictive reasoning [2] and revisiting the application domain in [4]. In a nutshell, the reasoning task of predicting anomalies (abnormal high price of accommodation) consists in capturing the context (using semantic representation of events and their categories with DBpedia and Wikidata vocabularies) and attaching it to accommodation price. Prediction is achieved by capturing semantic association rules, and reasoning over their content and the initial knowledge background (capturing the context of cities, population, density, events, among others).

3 Data and Vocabularies

Table 1 describes all data and vocabularies exploited by our system.

• ***Background Data in Use***: We define background data as data consumed silently by the system with no user specification. Only relevant background data (i.e., data which could be mapped to DBpedia, Wikidata for semantic comparison cf. [1]) is transformed in RDF for reasoning purpose. The size in column "*Size per day*" is the maximal amount of data collected i.e., for diagnosing / predicting abnormal price of accommodation in all cities.

• ***Output Data***: An average number of 61 (resp. $4,509$) RDF triples, with an average size of 3.1 (resp. 165.8) KB is retrieved for the explanation (resp. abnormality detection). All results are maintained in different internal RDF stores (through various size of graphs) for scalability purpose. The current size of the knowledge base is $31,897,325$ RDF triples.

Source Type	Data Source	Description	Format	Historic (Year)	Size per day (GBytes)	Data Provider
Anomaly	300,000+ unique travellers in 500+ cities recorded for 2015	Min. and max. number of respectively 2,521 and 24,800 expenses per city[a]	CSV	2015	.93 (complete) .41 (aggregated)	Private
Diagnosis	Social 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 (0.49 GB)	Eventbrite
		Planned events with large attendance	JSON format Accessed through Eventful APIs[b]	2011	Approx. 198 events per day (0.39 GB)	Eventful
Semantics	DBpedia	Structured facts extracted from wikipedia	RDF[c]	-	Approx. 33,000+ resources in use (0.23 GB)	Wikipedia
	Wikidata	Structured data from Wikimedia projects	RDF[d]	-	Approx. 189,000+ resources in use (0.63 GB)	Freebase Google inc.
	Accenture Categories	Structured is-A taxonomy of event categories	RDF[e]	-	25 resources in use (0.001 GB)	Accenture inc.

[a] https://www.eventbrite.com/api
[b] http://api.eventful.com
[c] http://wiki.dbpedia.org/Datasets
[d] https://www.wikidata.org/wiki/Wikidata:Database_download
[e] http://54.194.213.178:8111/ExplanatoryReasoning/ontology/categories.n3

Table 1: Data Sources in Use for REST Services.

4 Conclusion and Plan for Extensibility

We presented AIFS (**A***ccenture* **I***ntelligent* **F***inance* **S***ystem*) i.e., a system for interpreting, explaining and predicting abnormally high expense using accommodation price as abnormality in 500+ cities. The system, exposing semantic descriptions of causation between events and expenses type, is mainly used by (i) travel and expenses business owner to better manage spend optimisation, (ii) expenses auditor for tracking abnormal expenses, and (iii) internal travel system administrator for defining expenses policy based on the reasoning results. Semantics has been crucial to contextualise and interpret expenses and their amount.

References

1. Lécué, F.: 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)
2. Lécué, F., Pan, J.Z.: Predicting knowledge in an ontology stream. In: IJCAI. (2013)
3. Lécué, F., Schumann, A., Sbodio, M.L.: Applying semantic web technologies for diagnosing road traffic congestions. In: The Semantic Web - ISWC 2012 - 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part II. (2012) 114–130
4. Lecue, F., Tucker, R., Bicer, V., Tommasi, P., Tallevi-Diotallevi, S., Sbodio, M.L.: Predicting severity of road traffic congestion using semantic web technologies. In: ESWC. (2014) 611–627