

Proposta di Progetto di Dottorato

Giovanni Ciatto

May 19, 2017

1 Stato dell'arte

Oggigiorno assistiamo ad un rinnovato interesse, nell'ambito dei sistemi distribuiti e paralleli, per quanto riguarda i modelli, i linguaggi e le tecnologie legate allo *stream processing*, cioè quel termine-ombrello sotto al quale vengono accorpati, senza valorizzarne opportunamente le differenze, i paradigmi che si occupano di gestire flussi di **dati** e quelli che si occupano di flussi di **eventi**. La parola “evento”, in informatica, è tanto ricorrente quanto sovraccarica di significati: essa appare in maniera trasversale rispetto ai modelli ed alle architetture, onde per cui troppo spesso gli si associa una semantica vaga, che facilmente collassa su quella di “dato”.

Nei sistemi *event-based*, gli eventi rappresentano avvenimenti significativi per il sistema, come ad esempio il cambiamento di stato di un componente o la variazione di una proprietà del sistema o di un partecipante al sistema stesso, cui viene associato un contesto che ne permetta l'*interpretazione*. In questi sistemi, alla *percezione* dell'evento segue la produzione di una **notifica**, ovvero una *rappresentazione* sotto forma di dato strutturato della specifica occorrenza dell'evento stesso, e la sua *propagazione* agli altri partecipanti al sistema ad esso interessati. Spessissimo, come anche nel seguito, con la parola “evento” ci si riferisce in realtà alla sua notifica. Il vantaggio primario dei sistemi *event-based* risiede nell'elevato grado di disaccoppiamento tra i partecipanti al sistema che essi permettono. L'espressione *event-stream processing* si riferisce invece alla capacità di gestire grandi quantità di eventi, prodotti con una frequenza prossima al *real-time*, e di elaborarli in maniera continuativa. Infine, si parla di *Complex Event Processing* (CEP) laddove si ponga la necessità di aggregare, trasformare, correlare semanticamente, casualmente, temporalmente o spazialmente gli eventi che si susseguono in un sistema allo scopo di individuare l'occorrenza di eventi più complessi, identificati da pattern di eventi più semplici. Gli eventi possono rappresentare il meccanismo con il quale ogni sistema raggiunge i propri requisiti funzionali e, al tempo stesso, una importante sorgente di informazione che può essere usata dal sistema stesso per apprendere ed adattarsi dinamicamente o dall'organizzazione utilizzatrice del sistema per meglio comprendere le proprie attività ed i propri processi [cit. process mining].

Viceversa, nei sistemi che si occupano di gestire flussi di **dati**, si impiegano architetture distribuite per elaborarne moli imponenti, al fine di produrre un modello dei dati che permetta di analizzare l'informazione ivi contenuta, visualizzarla o sfruttarla per fini previsionali. In questo caso, i dati possono provenire anche da estrazioni *batch* da basi di dati, il che implicherebbe una frequenza di immissione minore rispetto ai flussi di eventi, ed il *focus* è sulla produzione, sulla calibrazione e sul mantenimento di modelli predittivi finalizzati all'acquisizione di informazioni rivendibili riguardanti i gusti e le abitudini di classi di utenti [cit. data mining].

Flussi di dati e flussi di eventi presentano spesso molti tratti in comune, prova ne sia l'elevato numero di tecnologie che dichiarano di supportare entrambi indistintamente, e.g. Apache Storm, NiFi, Flink, Ignite, etc., ma si distinguono per la differenza intrinseca che sussiste tra eventi e dati.

Il concetto di evento è talmente generale da far sì che i sistemi event-based presetino intersezioni e similitudini con molte altre macro-aree attive dal punto di vista accademico e industriale.

Nelle varianti più semplici di reti di sensori *wireless* (WSN), ogni nodo della rete è un produttore di eventi **situati**, cioè relativi ad un preciso punto dello spazio, i.e. la posizione del sensore. Le funzionalità di aggregazione e correlazione dei sistemi CEP ben si conciliano con gli scenari applicativi tipici delle WSN, come il monitoraggio di ambienti sia artificiali (e.g. *data-centers*) che naturali (e.g. inquinamento dell'aria, dei mari, situazione idrogeologica, etc), in cui l'identificazione di pattern di eventi aggregati in maniera decentralizzata risulta assai utile.

L'*Internet of Things* (IoT) è un altro ambito che presenta una naturale affinità con i sistemi ad eventi: in questa visione si immagina che gli oggetti di uso comune (e.g. luci, elettrodomestici, porte, finestre, automobili, telefoni, etc) siano tutti dotati di un'interfaccia di rete che permetta loro di interagire e cooperare per venire incontro (o addirittura prevenire) le esigenze dei loro utilizzatori. Non è difficile immaginare una chiave di lettura in cui ogni "cosa" diventa al tempo stesso produttore e consumatore (in inglese *prosumer*, da **producer** + **consumer**) di eventi, relativi alle attività compiute dagli esseri umani. Queste attività devono essere opportunamente *interpretate* affinché il sistema possa mettere in campo una strategia per supportare la quotidianità degli utenti, attuata per mezzo delle "cose" stesse. Risulta chiara l'importanza di riuscire a dare significato agli eventi prodotti dall'interazione tra gli utenti, l'ambiente e le "cose": questi vanno interpretati sia singolarmente che previa aggregazione temporale, spaziale e causale. Dalla pressoché infinita varietà di situazioni possibili derivano anche la necessità di un'interpretazione **semantica** (e non meramente sintattica) degli eventi unita alla possibilità per il sistema di imparare a riconoscere, dalle azioni degli utenti, le loro **attività** ed i legami di consequenzialità, mutua esclusione, etc. tra attività diverse di una o più persone.

La capacità di interpretare gli eventi come semanticamente correlati a partire dalla loro rappresentazione (che si suppone contenga il nome dell'evento in linguaggio naturale), presuppone a sua volta la facoltà per i partecipanti al sistema — che quindi possono essere considerati a buon titolo agenti intelligenti [cit.] — di avere accesso ad una rappresentazione formale delle relazioni semantiche tra termini (e.g. "allarme!" è sinonimo "pericolo!") e di poter quindi mettere in campo il ragionamento simbolico per

interpretare gli eventi. Basi di conoscenza che codifichino in maniera formale le relazioni tra i termini di un particolare dominio del discorso, sono dette ontologie [cit.?] ed esistono degli strumenti per i) esprimerle (e.g. OWL [cit.]), ii) produrle (e.g. Protégé [cit.]) iii) apprenderle tramite Text-Mining.