Sistema di allarme per la rete SEPA

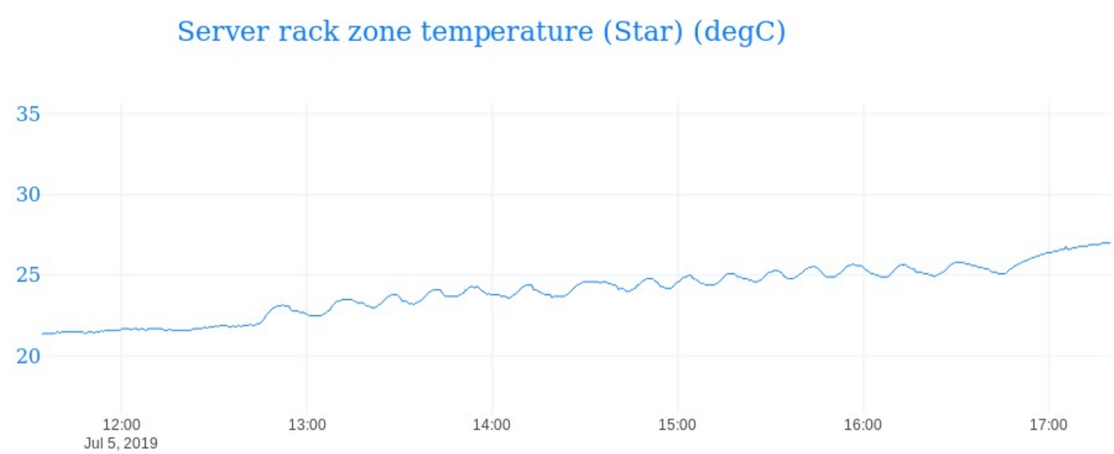
Introduzione

*Il SEPA(**S**PARQL **E**vent **P**rocessing **A**rchitecture) è un'architettura software che ha come obiettivo quello di supportare l'interoperabilità tra sistemi diversi di informazione.*
L'esponenziale crescita di dispositivi "intelligenti" connessi alla rete Internet presenta un problema di carattere pratico dovuto ai differenti protocolli e tecnologie utilizzati. L'architettura, che nasce appunto da questo bisogno, attraverso un meccanismo di publish-subscribe consente di rilevare, descrivere e notificare cambiamenti all'interno di base dati semantiche. Trova il suo impiego in tutti gli scenari caratterizzati da una forte dinamicità ed eterogeneità dei dati, come ad esempio quello del Web of Things.
Il sistema è composto da un ampio numero di sensori in Brasile e in Italia che monitorano alcune proprietà dell'ambiente come la temperatura, la qualità dell'aria e l'umidità del terreno. I dati dei senori sono accessibili attraverso l'interfaccia grafica web sul sito del SEPA dal quale è possibile anche accedere allo storico delle misurazioni di ogni singolo sensore.
Questo progetto si inserisce nello scenario appena descritto, dove il controllo di tutti i sensori non può essere sempre effettuato. Per esempio alcuni sensori misurano grandezze che non possono mai eccedere un certo limite, perciò è necessario un sistema che possa prevedere una situazione di instabilità in modo da intervenire per tempo.
È necessario quindi sviluppare un sistema d'allarme, che nello spirito dell'interoperabilità tra sistemi diversi, preveda situazioni di instabilità.

Obiettivo

L'obiettivo è quello di sviluppare un sistema di allarme capace di rilevare una situazione di instabilità, della quale condizioni sono definite arbitrariamente in una fase iniziale.
In questo documento non si affronterà lo sviluppo pratico del sistema di allarme, bensì le considerazioni teoriche che hanno permesso di realizzare un modello su cui basarsi durante la produzione del lavoro.
I dati a cui si farà riferimento sono alcune misurazioni effettuate da un sensore di temperatura dell'aria, posizionato in una stanza condizionata finalizzata al mantenimento di un server rack. La temperatura della stanza non può superare la soglia dei 27° , altrimenti l'efficienza dei server calerebbe fino a raggiungere il deterioramento dell'intero sistema. In particolare si farà riferimento a tutte le misurazioni effettuate nella giornata del 5/07/2019 dalle 11:34 fino alle 17:20 per un totale di 800 misurazioni(circa 1 misurazione ogni 26 secondi). In questo lasso di tempo il sistema di condizionamento ha subito un guasto portando la temperatura della stanza da 21.4° fino 27.7°.
La situazione, descritta dal Grafico 1, rappresenta una situazione che il sistema d'allarme dovrà prevedere ed allertare. L'obiettivo è quindi quello di creare un modello capace di studiare il trend delle misure e di ipotizzare i valori futuri assunti dalla temperatura nella stanza, così da poter allertare con anticipo il superamento della soglia.

Grafico 1



Funzione Interpolatrice

Per descrivere il trend delle misurazioni è possibile procedere calcolando la funzione interpolante che descriva con minor scarto possibile l'andamento nel tempo delle misurazioni.

Per comodità si prende come unità di misura dell'asse delle ascisse(Tempi):

$$1 \text{ u} = 26 \text{ s}$$

Perciò ogni misurazione sarà descritta come una coppia di numeri del tipo:

$$M_i(X_i, Y_i) \qquad n°1$$

Questa notazione si leggerà come: "*La misura M_i è stata fatta all'istante 11:34+X_i*26 s e il suo valore è pari a Y_i°.*"

Perciò dato l'insieme di *n* dati sarà possibile definire una funzione che descriva il trend assunto dai dati al passare del tempo.

Il *teorema di unicità del polinomio interpolatore* enuncia che:

"*Esiste uno ed un sol polinomio di grado n+1 che assume valori in corrispondenza di n punti distinti* $M_i(X_i, Y_i)$"

Perciò date 20 misurazioni, la funzione interpolante sarà di 19esimo grado, 40 misurazioni 39esimo grado, 800 misurazioni 799esimo grado e così via… Tale metodo risulta essere computazionalmente oneroso e inutile dal punto di vista dell'informazione sul trend delle misurazioni.

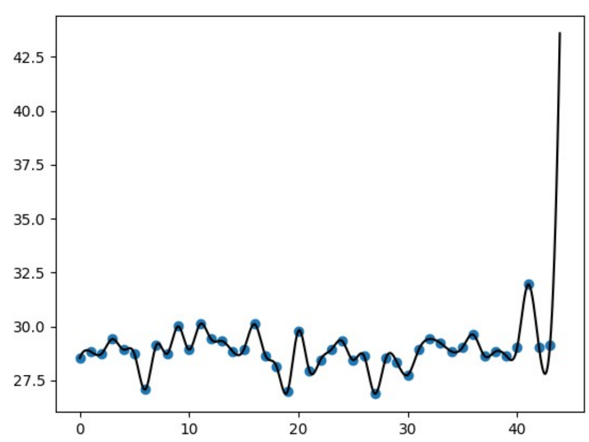


Grafico 2

Nel grafico 2 sono rappresentate 44 misure di temperatura interpolate da un funzione di 43-esimo grado. Come si può notare la funzione passa per tutte e 44 le misure(pullini blu) dando un'infomazione molto precisa sulla temperatura nell'intervallo di tempo [0t;44t] ma un'infomazione falsata invece in tutto il resto del dominio. Con tale metodo quindi non è permesso fare né previsioni affidabili né correggere eventuali errori di misurazione perché la funzione sarebbe costretta a passarci.

Il problema della funzione interpolatrice è il grado troppo alto. Per visualizzare il trend non è necessario descrivere con precisione tutte le misurazioni bensì trovare una funzione di grado discretamente basso che permetta di isolare l'errore e di descrivere al meglio l'incremento nel tempo delle misurazioni(con eventuale accelerazione dell'incremento).

Minimi quadrati

Una tecnica comune di interpolazione è quella dei minimi quadrati. Essa consiste nel trovare una funzione(in questo caso un polinomio) che abbia il minor scarto con i punti sperimentali. La funzione avrà uno scarto definito come la somma dei quadrati di tutti gli scarti:

$$\sum_{i=0}^{n} (Y_i - f(x_i))^2 \quad n°2$$

Scegliendo arbitrariamente come grado il 4°, la funzione $f(X_i)$ sarà del tipo:

$$f(X_i) = a \cdot x_i^4 + b \cdot x_i^3 + c \cdot x_i^2 + d \cdot \dot{x}_i + e$$

Dalla quale si ottiene successivamente sostituendo $f(X_i)$ alla n°2 :

$$\sum_{i=0}^{n} (Y_i - a \cdot x_i^4 - b \cdot x_i^3 - c \cdot x_i^2 - d \cdot \dot{x}_i - e)^2 \quad n°3$$

L'espressione ottenuta esprime lo scarto totale tra la funzione $f(X_i)$ e le misurazioni. Chimando φ la funzione che descrive lo scarto totale al variare di *a,b,c,d* ed *e* si ottiene una funzione in 5 incognite che permette di affrontare il quesito come un problema di minimo.

$$\varphi(a, b, c, d, e) = \sum_{i=0}^{n} (Y_i - a \cdot x_i^4 - b \cdot x_i^3 - c \cdot x_i^2 - d \cdot \dot{x}_i - e)^2$$

A che valore di *a,b...e* la funzione φ raggiunge il valore minimo, cioè lo scarto totale tra funzione $f(X_i)$ e dati sperimentali è minore? È possibile rispondere studiando gli zeri delle derivate parziali prime, cioè dove φ assume un valore di minimo o di massimo.

È opportuno precisare che φ non ha massimi. Infatti sarà sempre possibile individuare un valore di a,b...e tale che φ abbia un numero maggiore di qualunque numero prefissato, cioè lo scarto sia infinitamente grande.

Perciò in corrispondenza del punto dove tutte le derivate parziali prime di φ valgono 0 f sarà, tra le funzioni della famiglia dei polinomi di 4° grado, la funzione con meno scarto dai dati sperimentali.

Perciò i valori di a,b...e sono dati dalla soluzione del sistema:

$$\begin{cases} \frac{D\varphi}{Da}[\sum_{i=0}^{n}(Y_i - a\cdot x_i^4 - b\cdot x_i^3 - c\cdot x_i^2 - d\cdot \dot{x}_i - e)^2]=0 \\ \frac{D\varphi}{Db}[\sum_{i=0}^{n}(Y_i - a\cdot x_i^4 - b\cdot x_i^3 - c\cdot x_i^2 - d\cdot \dot{x}_i - e)^2]=0 \\ \frac{D\varphi}{Dc}[\sum_{i=0}^{n}(Y_i - a\cdot x_i^4 - b\cdot x_i^3 - c\cdot x_i^2 - d\cdot \dot{x}_i - e)^2]=0 \\ \frac{D\varphi}{Dd}[\sum_{i=0}^{n}(Y_i - a\cdot x_i^4 - b\cdot x_i^3 - c\cdot x_i^2 - d\cdot \dot{x}_i - e)^2]=0 \\ \frac{D\varphi}{De}[\sum_{i=0}^{n}(Y_i - a\cdot x_i^4 - b\cdot x_i^3 - c\cdot x_i^2 - d\cdot \dot{x}_i - e)^2]=0 \end{cases}$$

Il sistema dopo una serie di semplificazioni ha la seguente scrittura matriciale:

$$\begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} \begin{bmatrix} \sum_{i=0}^{n}(x_i^8) & \sum_{i=0}^{n}(x_i^7) & \sum_{i=0}^{n}(x_i^6) & \sum_{i=0}^{n}(x_i^5) & \sum_{i=0}^{n}(x_i^4) \\ \sum_{i=0}^{n}(x_i^7) & \sum_{i=0}^{n}(x_i^6) & \sum_{i=0}^{n}(x_i^5) & \sum_{i=0}^{n}(x_i^4) & \sum_{i=0}^{n}(x_i^3) \\ \sum_{i=0}^{n}(x_i^6) & \sum_{i=0}^{n}(x_i^5) & \sum_{i=0}^{n}(x_i^4) & \sum_{i=0}^{n}(x_i^3) & \sum_{i=0}^{n}(x_i^2) \\ \sum_{i=0}^{n}(x_i^5) & \sum_{i=0}^{n}(x_i^4) & \sum_{i=0}^{n}(x_i^3) & \sum_{i=0}^{n}(x_i^2) & \sum_{i=0}^{n}(x_i) \\ \sum_{i=0}^{n}(x_i^4) & \sum_{i=0}^{n}(x_i^3) & \sum_{i=0}^{n}(x_i^2) & \sum_{i=0}^{n}(x_i) & n \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n}(Y_i\cdot x_i^4) \\ \sum_{i=0}^{n}(Y_i\cdot x_i^3) \\ \sum_{i=0}^{n}(Y_i\cdot x_i^2) \\ \sum_{i=0}^{n}(Y_i\cdot x_i) \\ n\cdot Y_i \end{bmatrix}$$

Utilizzando la regola di Cramer sappiamo che:

$$a=\frac{Det(a)}{Det} \quad b=\frac{Det(b)}{Det} \quad c=\frac{Det(c)}{Det}$$

$$d=\frac{Det(d)}{Det} \quad e=\frac{Det(e)}{Det}$$

(Si risparmieranno i passaggi per calcolare il determinante delle 6 matrici).

Una volta ottenuta la funzione approsimata f, si può procedere effettuando alcune considerazioni:

-La soluzione del sistema è determinata solo per *n>=5* per il *teorema di unicità del polinomio interpolatore*;

-La funzione f dà informazioni sempre meno precise all'avvicinarsi di *n* a 5 per le stesse considerazioni effettuate sulla retta interpolante;

-La funzione è meno soggetta alle variazioni delle misure, di conseguenza anche a quelle falsate, all'aumentare di n;

-In questo caso è stato utilizzato un polinomio di 4° grado, ma si deduce che è possibile utilizzare una qualsiasi funzione al posto di f procedendo in maniera analoga a quella appena mostrata.