

Relazione Finale Stage
Fondazione Bruno Kessler - Unità SHELL
Referente aziendale Chiara Ghidini e Chiara Di Francescomarino
Tutor Universitario Fausto Giunchiglia

Williams Rizzi

July 1, 2015

1 Estratto

I processi di business sono spesso supportati da sistemi informativi che registrano dati riguardanti le esecuzioni. Questi dati possono essere estratti sotto forma di log di eventi. Mediante l'utilizzo di tali log il monitoring predittivo di processi si propone di predire come casi ancora incompleti evolveranno fino al completamento. All'interno di questo progetto è proposto un framework di monitor predittivo di processi per stimare la probabilità che un dato predicato sia verificato o meno entro il completamento dell'istanza in esecuzione e il tempo necessario perché il predicato sia soddisfatto o violato. Il predicato può essere, per esempio, un vincolo descritto in logica temporale o in qualsiasi linguaggio che possa essere valutato su di una traccia completata. Il framework tiene conto sia delle sequenze di eventi osservati nella corrente traccia sia degli attributi associati a questi eventi. L'approccio alla predizione si compone di due fasi, inanzitutto i prefissi delle tracce passate sono suddivisi in clusters sulla base dell'informazione contenuta nel flusso di controllo, successivamente viene costruito un classificatore per ogni cluster per discriminare le verifiche dalle violazioni. Durante l'esecuzione viene espressa una predizione sull'istanza corrente mappandola sul cluster ad essa più simile ed applicandovi quindi il corrispondente classificatore. Il framework è stato implementato nel toolset ProM e validato su dati relativi al trattamento di pazienti di un grande ospedale olandese.

2 Obiettivi

I fini dello stage possono essere divisi in obiettivi di carattere tecnico e di carattere personali. I primi riguardano aspetti di natura prettamente concettuale e pratica quali la comprensione di algoritmi e il loro utilizzo in un contesto di lavoro reale, i secondi invece riguardano prettamente l'aspetto umano del lavoro ad un software in un gruppo di ricerca.

Obiettivi Tecnici:

- Acquisizione di tecniche di machine learning quali decision tree classifiers come l'algoritmo C4.5 e la relativa implementazione Java denominata J48.

- Comprensione e padroneggiamento del linguaggio XML XES.
- Consolidamento delle conoscenze acquisite a lezione per quel che riguarda il linguaggio di logica matematica LTL.
- Presa confidenza con la piattaforma ProM.
- Utilizzo di classificatori nell'ambito del machine learning tratti dal toolset Weka e padroneggiamento relativi log ARFF.
- Esperienze di modellazione concettuale consistenti nel categorizzare un'entità continua come il tempo tramite categorie discrete.
- Implementazione di moduli Java in grado di utilizzare le conoscenze sopra riportate per predire soddisfacimento o violazione di vincoli su istanze di processo.

Obiettivi Personali:

- Capacità di lavorare e fare team.
- Inserimento all'interno di un team di lavoro consolidato.
- Capacità di seguire delle direttive, quali task, rispettare impegni e scadenze, quali meeting e consegne.
- Relazionamento e organizzazione con i propri colleghi di gruppo.

3 Conoscenza e strumenti acquisiti

Per il corretto svolgimento dei compiti legati al tirocinio è stato necessario prendere dimestichezza con alcuni strumenti quali una IDE per il supporto alla modifica del codice, i concetti di clustering, classificazione e apprendimento supervisionato. Questa fase di preparazione è stata per me particolarmente stimolante in quanto mi ha dato la possibilità di approfondire concetti appresi a lezione, ma soprattutto ampliare i miei orizzonti in un campo che trovo oltremodo affascinante.

4 Svolgimento

Ho avuto l'occasione grazie alla Dott.ssa Chiara Ghidini e alla Dott.ssa Chiara Di Francescomarino di prendere parte attivamente all'implementazione di alcuni moduli Java facenti parte del software "Predictive Monitoring Time", per la predizione del tempo necessario per la verifica/violazione di predicati. L'attività proposta è volta a formare e dare un input su quello che sarà il mondo del lavoro nel campo dell'analisi di possibili approcci per il monitoring predittivo della realtà. Tecniche e strumenti basati su questi approcci possono essere utilizzati per catalogare la conoscenza in modo da poterla sfruttare per fare predizioni che possano aiutare l'uomo nella vita di tutti i giorni. Il

contesto di studio è un caso reale in ambito ospedaliero anche se l'approccio potrebbe essere adattato anche per altri modelli di business. Il mio primo task è stato l'utilizzo delle API di WeKa per la scrittura dei file ARFF, i file presi in input dagli algoritmi di classificazione di WeKa. Il modulo per l'utilizzo delle API per la scrittura del file ARFF è andato poi a sostituire la scrittura manuale di tale file. Una volta verificata la correttezza dell'implementazione, ho isolato la parte che riguardava la stampa dell'oggetto ARFF e la ho sostituita con il mantenimento del relativo oggetto in memoria in modo da rendere più veloce il software, lasciando comunque all'interno delle configurazioni la possibilità di scegliere se stampare o meno il report. Ultimato questo compito, il task successivo prevedeva l'integrazione di una funzionalità alternativa all'interno del software per la verifica di una formula LTL sulla singola traccia a runtime. Per fare questo è stato necessario aggiungere altre librerie di ProM che si occupano di tale verifica, integrarle con il Predictive Monitoring time e verificare la correttezza di quanto fatto tramite appositi test. Successivamente abbiamo fatto dell'altro testing e ricercato valori ottimali da assegnare ai parametri utilizzati per definire configurazioni del software. Questa fase è stata molto utile per comprendere e prendere dimestichezza con il ruolo dei singoli parametri e il loro significato.

5 Contesto di lavoro

L'attività lavorativa si è svolta perlopiù negli uffici della sede del gruppo SHELL in FBK e in minima parte da casa e dalla biblioteca universitaria. L'orario di lavoro è stato infatti concordato come un orario flessibile in modo da darmi la possibilità di conciliare le lezioni universitarie e lo stage, senza dover rinunciare né all'una né all'altra attività. Il team di lavoro era composto da me e un mio compagno di corso, anch'esso stagista, con il quale ho dovuto coordinarmi per lo svolgimento dei task ricevuti. Abbiamo concordato dei meeting settimanali con la Dott.ssa Chiara Di Francescomarino per definire i task settimanali, parlare di eventuali difficoltà incontrate nell'implementazione e nella comprensione del codice necessario a portare a termine il task e mantenere il team al corrente dello stato di avanzamento dei vari task in modo da adattare il dispendio di energie in funzione della situazione in cui ci trovavamo.

6 Conclusione

Quest'attività mi ha dato la possibilità di mettere in pratica molte delle cose viste a lezione in qualcosa di fruttuoso. Mi ha insegnato molto sia per quel che riguarda l'aspetto meramente pratico, ho editato codice con un livello di complessità che non avevo mai raggiunto prima d'ora, sia per quel che riguarda l'aspetto teorico dandomi degli ottimi spunti per una tesi futura, piuttosto che per un piano di studi in vista di un impiego. Ho avuto inoltre la possibilità di stare spalla a spalla con professionisti del campo dai quali ho cercato di imparare il più possibile sia per quel che riguarda i tool utilizzati sia per quel che riguarda il metodo e il ritmo di lavoro. Ritengo che il tirocinio formativo durante il percorso di studi costituisca una grande opportunità, non solo perché consente

di ampliare le proprie conoscenze, ma soprattutto perché permette un'interazione con persone prettamente del settore dotate di grande esperienza e professionalità.