



Mining Probabilistic Declarative Process Models

Autori

Elena Bellodi

Fabrizio Riguzzi

Evelina Lamma



Process Mining

- **Business process (BP)**

Processo composto di un insieme di attività interdipendenti, che specifica l'uso delle risorse e l'instradamento del lavoro in un'organizzazione.

- I **sistemi informativi** registrano a run-time gli eventi di un *business process*, corredati di timestamps, in un file di **log**. Sequenza di eventi correlati = **traccia di processo**.

- **Process Mining**

Estrazione di una descrizione strutturata (modello di processo) a partire dal log di eventi.

Process Mining da carriere universitarie

Log di eventi: database con 813 studenti triennali (con carriera terminata) in Ing. dell'Informazione, iscritti dal 2004 al 2008

ATTIVITA' (evento)	ATTRIBUTI
Dati matricola	<ul style="list-style-type: none"> ■ Identificativo numerico ■ Prov. e Stato di residenza ■ Tipo e voto di maturità
Immatricolazione	Anno
Iscrizione_n (n=2..9)	Anno d'iscrizione, anno di corso (1,2,3), ripetente (IC,FC)
Esame	Materia, categoria voto, lode (s/n)
Fine carriera	<p>Tracce Positive (327): laurea</p> <p>Tracce Negative (486): passaggio di corso, iscrizione non rinnovata, trasferimento in uscita, rinuncia, sospensione</p>

Linguaggi probabilistici per PM

■ Logica del primo ordine

se un mondo (interpretazione) viola una formula è ritenuto impossibile

■ Logica di Markov (ML)

- Se un mondo viola una formula è meno probabile ma non impossibile
- Estende le formule logiche del primo ordine con pesi (valori reali):

peso più alto \longrightarrow vincolo più forte \longrightarrow più probabile il mondo che soddisfa il vincolo

peso infinito \rightarrow logica standard

- Modello di BP = set di formule in logica di Markov

Algoritmi di PM

■ DPML (Declarative Process Model Learner)

Il modello di processo è una teoria di vincoli logici **classificatori** nella forma *Body*→*Head* che:

- coprono (sono veri) tutte o il maggior numero possibile di tracce positive
- eliminano (sono falsi) il maggior numero di tracce negative

```
hap(dati_matricola, [tipo_maturita(scientifica), G, voto_mat(basso)]) →  
EN(hap(esami, [materia(11828), A, B, categoria_voto(basso)]), C, D).
```

■ Alchemy (sistema open-source): apprendimento

- Pesi per i vincoli logici, dato un db di training
- Probabilità di uno studente di appartenere alla classe + (laureato) o – (interruzione carriera), dato un db di test

Esperimenti

1. 10-fold cross validation: DPML su 10 db di train → 10 teorie logiche con accuratezza media = 67,5%
2. Apprendimento pesi sulle teorie tradotte in ML
3. Inferenza delle probabilità su 10 db di test a partire da:
 1. Teorie con pesi appresi → classificazione logico-probabilistica
 2. Teorie con pesi settati a 1e+10 → classificazione logica
4. Valutazione performance classificatori
 - Metodo AUC (Area Under the receiver operating characteristic Curve)
Class. Logica: AUC = **0.710** VS Class. Logico-prob.: AUC = **0.722**
 - One-tailed paired t test: l'ipotesi nulla che i 2 algoritmi siano equivalenti può essere rifiutata con una probabilità del 90.58%.

Conclusioni

DPML+Alchemy → classificazione più accurata del solo DPML