

Introduzione al Data Mining e applicazioni al dominio del Contact Management

Parte VI: applicazioni ed esempi

Andrea Brunello

Università degli Studi di Udine



*In collaborazione con dott. Enrico Marzano, CIO Gap srl
progetto Active Contact System*

Parte VI:

- ▶ applicazioni ed esempi relativi al dominio del Contact Management.

- ▶ Grande mole di dati a disposizione, riguardante il lavoro degli agenti (*call center representatives*) e le sessioni (telefoniche) da essi gestite;
- ▶ diversi obiettivi:
 - ▶ previsione dell'esito di una sessione;
 - ▶ previsione della presa in carico o meno di una telefonata da parte di un'agente;
 - ▶ valutazione delle performance lavorative degli agenti;
- ▶ scenario ideale per l'applicazione delle tecniche di Data Mining:
 - ▶ attribute selection;
 - ▶ clusterizzazione;
 - ▶ classificazione.

Idea → strumento per la valutazione automatica degli agenti.

Funzione analitica che considera diversi parametri:

- ▶ **dati tecnici:** provenienti dai log dei sistemi operazionali;
- ▶ **dati di servizio:** raccolti direttamente dall'agente, e diversi per attività di *Inbound* e *Outbound*.

- ▶ La performance è un valore di *ranking* compreso fra 0 e 1;
- ▶ la valutazione finale di ciascun agente viene costruita a partire dai valori delle singole metriche considerate.

Estensioni:

- ▶ valutazione dei dati di servizio non strutturati raccolti dall'agente;
- ▶ stima del carico di lavoro dell'agente;
- ▶ rilevamento delle anomalie di lavorazione.

- ▶ Per essere valutati, i dati non strutturati necessitano di pre-processamento;
- ▶ a tal fine vengono definite e calcolate su di essi delle proprietà/metriche oggettive;
- ▶ sulle stesse vengono poi applicate delle analisi, secondo modelli di segmentazione e classificazione;
- ▶ i risultati così ottenuti vengono incorporati nel sistema di valutazione delle performance.

Si vogliono premiare gli agenti in grado di gestire più tipologie di servizio, e/o servizi complessi.

Parametri presi in considerazione, ad esempio:

- ▶ numero di tipologie di servizio distinte gestite dall'agente;
- ▶ impatto del servizio in esame;
- ▶ peso del servizio preso in esame;
- ▶ peso degli altri servizi gestiti dall'agente nel periodo.

Confronto fra i risultati ottenuti attraverso:

- ▶ funzione analitica di scoring;
- ▶ metodologie di analisi predittiva.

Alcuni operatori talvolta inseriscono esiti di sessione non corrispondenti alla realtà:

- ▶ inconsciamente, a causa di errori;
- ▶ consciamente, ad esempio simulando esiti positivi non reali.

Possibilità di rilevare tali operatori, valutando:


- ▶ distribuzione degli esiti;
- ▶ tempi di lavorazione.


Due fronti di analisi:

- ▶ valutazione dello scostamento della distribuzione degli esiti dell'agente rispetto ai valori del servizio (in termini di media e deviazione standard);
- ▶ generazione di un modello predittivo per la previsione di un esito di sessione sulla base dei tempi di lavorazione.

Le diverse applicazioni hanno permesso di affrontare praticamente questioni riguardanti:

- ▶ clustering: gerarchico, *k-means++*, *EM*, sia per l'esplorazione del dominio che come base per classificazione;
- ▶ classificazione con albero di decisione, e tuning dei parametri di costruzione dell'albero (es. per prevenire overfitting o per aumentare la leggibilità);
- ▶ trattamento di *class* ed *attribute noise* negli insiemi di training;
- ▶ definizione di funzioni analitiche.

 I. H. Witten, E. Frank, M. A. Hall: *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, 3rd edition, 2011.

 G. James, D. Witten, T. Hastie, R. Tibshirani: *An Introduction to Statistical Learning with Applications in R*. Springer, 2013.