



Se fino a oggi i dati hanno permesso al marketing di migliorare enormemente le sue performance, i prossimi anni saranno ricordati per i successi, anche commerciali, delle analisi dei testi e delle parole. Il data mining, ossia l'analisi di dati numerici strutturati e raccolti in appositi database, consente di comprendere molto bene "i fatti" accaduti. Ad esempio: chi sono i nostri migliori clienti, quali sono le loro abitudini di acquisto, il loro valore nel tempo... Ma se vogliamo capire anche "il perché" di certi comportamenti (ad esempio: un improvviso calo negli acquisti), spesso non siamo in grado di saperlo con i dati in nostro possesso e dobbiamo quindi ricorrere a ricerche di mercato o a interviste sul campo. Ma queste oltre a costare, comportano tempi di esecuzione non sempre compatibili con l'urgenza delle decisioni da prendere. Specialmente se è entrato improvvisamente nel nostro mercato un nuovo concorrente. È in situazioni come queste che si comprende bene il valore del text mining. Se questa metodologia fosse integrata nei processi di Business Intelligence e Knowledge Management, l'azienda saprebbe in anticipo della nascita di un nuovo concorrente, delle criticità dei propri prodotti, dell'insoddisfazione dei propri clienti e, probabilmente, saprebbe come prevenire in tempo la contrazione delle vendite. Non c'è niente di magico in tutto questo. Il text mining non è altro che l'applicazione delle consolidate tecniche di data mining, a dati poco o per niente strutturati. E per dati non strutturati non si intendono solo documenti, ma tutti i dati testuali in genere: e-mail dei clienti, newsgroup o chat, brevetti, ricerche di mercato, lamentele scritte (o trascritte, se telefoniche), pratiche legali e amministrative, articoli di giornale, ecc. I campi di applicazione non riguardano solo il marketing. Il text mining può interessare i settori più disparati: dall'ufficio del personale (analisi automatica dei curricula per ottimizzare le selezioni) alla pubblica amministrazione (analisi delle richieste degli utenti per migliorare i servizi); dalla salute (analisi di basi dati epidemiologiche e medico-farmaceutiche), alla gastronomia (evoluzione dei gusti e delle tendenze alimentari), e così via. La differenza con i motori di ricerca è sostanziale: anche se i motori sono molto utili, come tutti possono confermare, non consentono però una grande precisione né garantiscono una elevata pertinenza dei documenti che analizzano. Il motivo è che operano su piattaforme diverse, con differenti linguaggi e, soprattutto, senza un'associazione con 'metadati' che permettano la "comprensione" del contenuto. Tanto per fare un esempio: se io voglio fare un'indagine sull'Assicurazione Toro, i motori di ricerca mi sottoporrebbero anche tutto ciò che ha a che fare con le corride o con l'omonimo segno zodiacale. Con il Text Mining, no. Per chi volesse approfondire, segnalo che in Europa il Text Mining ha iniziato a svilupparsi nell'ambito del progetto NEMIS (Network of Excellence in Text Mining & its Applications in Statistics: <http://nemis.cti.gr/>) del cui comitato scientifico fa parte per

La parola è d'oro.

Di Mattia Camellini

Mercoledì 15 Marzo 2006 12:41

In Italia il Prof. Sergio Bolasco dell'Università la Sapienza di Roma. Per concludere, una curiosità. È stata appena pubblicata una recentissima applicazione politica di questa metodologia, curata dallo stesso Prof. Bolasco (uno statistico), da Luca Giuliano (un sociologo) e da Nora Galli de' Paratesi (una linguista). Si tratta del Libro Parole in Libertà (Manifestolibri, 2006), in cui sono analizzati nell'arco di 12 anni 111 discorsi di Silvio Berlusconi. Il libro è interessante perché, di fronte alle impressioni che ognuno può essersi fatto del personaggio politico più noto d'Italia, qui può trovare una puntuale analisi lessicale e retorica dei suoi discorsi e dei suoi scritti. Con significativi raffronti tra il linguaggio dei momenti ufficiali e quello delle convention di partito; tra ciò che ama associare a "nuovo" e ciò che considera decisamente "vecchio"; tra le parole che usa di più e quelle che evita di più.