

UNIVERSITÀ DEGLI STUDI DI PISA
Facoltà di Economia
Facoltà di Scienze, Matematiche, Fisiche e Naturali
Corso di Laurea specialistica in
Informatica per l'Economia e per l'Azienda



TESI DI LAUREA

BUSINESS INTELLIGENCE NELL'IMPRESA PROATTIVA

RELATORE:

Prof. Antonio Albano

RELATORE:

Prof. Luciano Marchi

Candidato:
Nicola Ciaramella

ANNO ACCADEMICO 2003-04

Introduzione.....	5
Capitolo 1 L'impresa proattiva.....	11
1. Costrutti teorici.....	11
2. Formula imprenditoriale.....	15
3. Customer Based View.....	22
4. Resource Based View.....	24
5. Competition Based View.....	26
6. Verso un quadro di riferimento per la business intelligence.....	28
Capitolo 2 Il capitale relazionale.....	31
1. Il concetto di capitale relazionale.....	31
2. Il ciclo di vita della relazione.....	32
3. Modelli della relazione e tecnologie di business intelligence.....	38
4. Il customer equity aggregato.....	40
5. Il customer equity individuale.....	42
6. Il customer equity potenziale.....	49
7. Customer equity e decisioni di marketing.....	58
Capitolo 3 Segmentazione e ottimizzazione del valore.....	64
1. Sul concetto di segmentazione.....	64
2. Segmentazione come euristica di ottimizzazione.....	70
3. Il problema della segmentazione dei cataloghi.....	72
4. Considerazioni metodologiche.....	74
5. Euristiche per il problema della segmentazione.....	76
Capitolo 4 Valore e business intelligence.....	81
1. Qualità di un sistema di data mining.....	81
2. Un modello microeconomico della classificazione.....	84
3. Determinazione dei parametri del modello microeconomico.....	89
4. Ottimizzazione nel modello microeconomico.....	91
5. Un esempio di applicazione del modello microeconomico.....	95
6. Valore di un classificatore.....	99
7. Estensioni dinamiche del modello.....	100
Capitolo 5 Riflessioni sulla customer intelligence.....	103
1. Azioni e risultati.....	103

2. Segmenti e stati del cliente	109
3. Modelli multidimensionali segmentati	112
4. Attributi non osservabili	115
Capitolo 6 Autonomic customer intelligence	118
1. Obiettivi e motivazioni del progetto	118
2. Modello computazionale	123
Bibliografia.....	127

Introduzione

Il presente lavoro nasce da un'esigenza maturata in una serie di progetti di ricerca applicata attuati da varie imprese e dipartimenti universitari. A nostro giudizio le due tecnologie che costituiscono il nucleo più caratterizzante dei sistemi di *business intelligence* sono il *data warehousing* e il *data mining*, che cooperano nel più generale processo di *knowledge discovery*. L'esigenza di cui dicevamo è di chiarire sul piano teorico e utilizzare su quello operativo alcune analogie e opportunità di sinergia che sussistono fra queste tecnologie e le teorie elaborate nelle discipline economiche e aziendali. I sistemi, i progetti e le applicazioni di business intelligence possono essere più efficaci sul piano decisionale, più fondate sul piano culturale, più convincenti sul piano psicologico, e non da ultimo più giustificate sul piano economico, se realizzano una consapevole sintesi di contributi intellettuali che vengono da aree disciplinari accomunate da molti interessi comuni, ma divise da differenze culturali che sono giustificabili ma che a volte pongono limitazioni in realtà superflue.

La domanda che viene dal mondo aziendale e l'offerta che viene dal mondo tecnologico soffrono entrambe di un *troppo* e di un *troppo poco*. I manager e gli analisti di impresa, nella positiva ansia di raggiungere risultati, pongono a volte alle tecnologie richieste troppo ambiziose, quasi a risolvere problemi decisionali e conoscitivi che sempre e comunque spettano a loro. Ma altrettanto se non più spesso chiedono soltanto soluzioni e supporti tradizionali, sperimentati, a volte anche banali, senza aprirsi a nuove possibilità applicative aperte dal progresso tecnologico. Dal lato dell'offerta, con logica opposta ma anche speculare, i produttori di tecnologie e gli esperti nel loro uso promettono spesso soluzioni irrealisticamente ambiziose, trascurando quegli aspetti del mondo "non perfetto" dell'impresa che purtroppo (o per fortuna) pongono problemi sempre nuovi e destrutturati. Oppure si lasciano sfuggire realistiche opportunità applicative perché non ne sono abbastanza consapevoli, e non li colgono nei loro interlocutori di impresa.

Gli attori di entrambe le parti possono operare in modo assai più incisivo e consapevole se riescono a sintetizzare in un quadro teorico unitario almeno alcuni temi delle due aree disciplinari.

Il presente lavoro non pretende certamente di risolvere, e neanche di affrontare, questi problemi nella loro globalità, compito evidentemente immane. Il contributo che intende dare è di mettere in chiaro (o almeno più in chiaro di quanto già siano) alcuni aspetti microeconomici del data warehousing e del data mining e, in senso opposto, alcuni aspetti dell'uso di queste tecnologie per il decisore che si pone obiettivi non tanto conoscitivi quanto, appunto, microeconomici, cioè che cerca prima di tutto l'utilità e solo strumentalmente la conoscenza. Per questo motivo lungo tutta la discussione, sia essa di momento in momento centrata sugli aspetti scientifici e tecnologici o su quelli applicativi e decisionali, il concetto di *valore* compare come idea unificante e sempre ricorrente. La business intelligence *crea valore* ed essa stessa *ha valore*: questo enunciato cerca nel presente lavoro un chiarimento e degli spunti per una fondazione teorica interdisciplinare.

Il lavoro, come detto, si inquadra in un programma di ricerca applicata assai più ampio e in corso di svolgimento. Il contesto teorico di questo programma, e il suo settore applicativo di elezione, è l'*impresa proattiva*, cioè l'impresa è basata sull'anticipazione, il governo e anche la creazione delle tendenze di mercato e sull'innovazione continua in un contesto ipercompetitivo. Questa è l'impresa che dà origine alle più interessanti esperienze di business intelligence, e quindi l'impresa più significativa per il programma di ricerca.

Nel capitolo 1 *L'impresa proattiva* si presenta quindi un sintetico quadro dei concetti teorici che caratterizzano l'impresa proattiva.

Essa può essere vista secondo tre ottiche, che si differenziano nella individuazione delle fonti primarie del valore:

- la *customer based view*: il valore è generato dai clienti, vince chi sa rapportarsi meglio ai clienti, ciò che conta è la soddisfazione e la fedeltà dei clienti;
- la *resource based view*: il valore è generato dalle risorse e dalle competenze, vince chi ha risorse più avanzate e non imitabili, ciò che conta è l'eccellenza nelle competenze;
- la *competition based view*: il valore è generato dalle mosse competitive, vince chi

innova e difende il vantaggio acquisito innovando, ciò che conta è fare la mossa giusta prima degli altri.

Benché il programma complessivo di ricerca indaghi su tutte le tre visioni dell'impresa, del suo valore, dei processi che lo generano e della conoscenza che quei processi abilita, nel presente lavoro ci concentriamo sulla customer based view, per delineare alcuni principi teorici e metodologici utili alla progressiva costruzione di una innovativa fondazione teorica per la *customer intelligence*, essenziale branca della business intelligence.

Nel capitolo 2 *Il capitale relazionale* si tratta del valore dei clienti dell'impresa, che nella customer based view è in ultima analisi il valore dell'impresa.

Dopo aver parlato di relazione fra cliente e impresa, di ciclo di vita della relazione, di soddisfazione e fedeltà e dei fattori che le determinano, si affronta il problema della stima del *customer equity*, il valore del cliente. Il concetto è visto nella forma aggregata, il valore dell'intera clientela, e nella forma individuale, il valore di ciascun singolo cliente; ed è anche visto nella forma attuale, il valore che i clienti hanno nello scenario esistente, e nella forma potenziale, il valore che i clienti avrebbero in scenari possibili (con una escursione sul metodo delle opzioni reali).

Si presenta quindi un modello di calcolo del customer equity discutendo queste distinzioni, indicando come le tecnologie di business intelligence trovino in questo tema delle stimolanti opportunità sia applicative che teoriche, e anche suggerendo che le stesse discipline manageriali potrebbero trovare nella business intelligence spunti per potenziare il loro quadro teorico.

Nel capitolo 3 *Segmentazione e ottimizzazione del valore* si delinea una *visione computazionale* della segmentazione.

Si definisce una classe di problemi di ottimizzazione matematica con evidente interesse applicativo, qui esemplificato nel marketing ma rilevante in molti tipi di decisioni economiche. Questi problemi hanno intuitive analogie con i problemi di segmentazione (e anche di marketing individuale inteso come applicazione estrema della segmentazione). Questi problemi sono in genere di complessità non trattabile nella loro forma pura. Per risolverli in modo approssimato è necessario ricorrere a euristiche capaci di ridurre lo

spazio effettivo di ricerca rispetto a quello di tutte le innumerevoli soluzioni possibili. Si arriva quindi a pensare la segmentazione, nel suo senso classico del marketing, come euristica per risolvere questi astratti problemi di ottimizzazione. Il concetto è esemplificato su una formulazione astratta del problema del marketing diretto basato su cataloghi. Si formulano anche delle considerazioni metodologiche sull'utilità delle tecnologie di business intelligence in questa ottica computazionale della segmentazione.

Queste considerazioni teoriche delineano una visione microeconomica del data warehousing e del data mining, considerati congiuntamente come strumento di ottimizzazione delle decisioni. Questa visione è approfondita nel capitolo seguente.

Nel capitolo 4 *Valore e business intelligence* si affronta in linea teorica il tema della qualità di un sistema analitico-decisionale di business intelligence e si confronta il classico approccio conoscitivo (qualità come correttezza di risposta) con un approccio microeconomico (qualità come fonte di utilità).

Si formula un modello microeconomico del data mining secondo un approccio ispirato alla teoria delle decisioni. Le risposte di un algoritmo di classificazione, le regole di decisione dell'impresa, le funzioni di utilità economica dei rapporti fra impresa e cliente e la composizione della popolazione dei clienti compaiono tutti come variabili o parametri in un problema di ottimizzazione lineare.

Definito il modello, si discute di come un sistema di business intelligence può essere inquadrato in questa visione e di come può contribuire all'ottimizzazione economica.

Si mostra che il modello può essere usato per trattare la dimensione temporale, e quindi per la stima dei risultati futuri, essenziale per la valutazione di un investimento.

Si suggerisce poi l'estensione del modello verso una visione del rapporto fra cliente e impresa come processo stocastico e come interazione fra automi.

Ancora, si discute di come il modello possa essere utilizzato per assegnare un valore economico ai sistemi di data mining e di conseguenza di come si possa usarlo per valutare i ritorni degli investimenti in data mining.

Il modello e il suo uso sono descritti su un esempio essenziale ma realistico di direct marketing.

Nel capitolo 5 *Riflessioni sulla customer intelligence* si svolgono alcune considerazioni metodologiche su temi di progettazione attinenti alle tematiche discusse precedentemente in un quadro teorico formale.

Si indicano alcune entità, sia fatti che dimensioni, che in un datawarehouse a supporto della customer intelligence servono a rappresentare i costrutti teorici della customer based view.

Si stabilisce una analogia fra segmenti di clientele e stati interni del cliente, entrambi definiti dalla funzione di risposta del cliente alle azioni dell'impresa. L'analogia è suggerita dallo studio condotto nei capitoli precedenti e contribuisce a chiarirne le implicazioni. Si suggerisce l'introduzione di questi concetti nel customer datawarehouse per la customer intelligence.

Si discute l'interesse di modelli multidimensionali dei dati in cui le aggregazioni sono create da algoritmi di data mining per segmentazione e clustering.

Si esemplifica un caso di inserimento nel customer datawarehouse di valori di attributi non osservabili della relazione fra cliente e impresa e del capitale relazionale, valori calcolati con metodi di data mining a partire da risposte a questionari.

Nel capitolo 6 *Autonomic customer intelligence* si descrivono obiettivi e modello computazionale di un programma pluriennale di ricerca industriale.

L'idea centrale della ricerca è un sistema di data mining per la gestione di campagne di marketing interattivo che in modo il più possibile automatico valuta in tempo reale l'andamento della campagna ed effettua le azioni correttive di ri-pianificazione.

La tecnologia sottostante prevede una popolazione di algoritmi di predizione del comportamento dei clienti che apprende per competizione, in modo ispirato alla programmazione evolutiva e genetica.

Un prototipo del sistema è stato sperimentato su dataset sintetici di campagne di tele-marketing simulate, con risultati positivi. Un'altra versione è stata sperimentata sul campo in un portale web a larghissimo accesso, ancora con esito positivo. È ora in via di realizzazione un prototipo più avanzato da impiegare in reali campagne di direct marketing postale. Nel programma di ricerca ci si è ispirati alla visione microeconomica della business intelligence presentata nei capitoli precedenti. Essi si sono dimostrati fertili di implicazioni interessanti e meritevoli di approfondimenti anche sotto il profilo teorico.

Capitolo 1

L'impresa proattiva

1. Costrutti teorici

Il concetto di impresa proattiva è formulato in [Valdani 2000] per descrivere un'impresa che deve essere capace di:

- gestire eventi e conseguenze di eventi che non si sono ancora manifestati;
- progettare il suo futuro mentre governa il presente;
- definire il proprio destino perché conosce il destino del suo settore;
- scoprire nuove soluzioni perché sa guardare lontano, oltre la tradizione;
- innovare e creare proposizioni di valore per la sua clientela;
- creare valore stimolando lo spirito imprenditoriale e la disciplina della creatività;
- stimolare il lavoro di gruppo e la disponibilità alla collaborazione;
- apprendere l'arte della coevoluzione;
- attivare processi di leverage e building delle sue risorse e capacità market driving;
- rigenerare se stessa e il suo mercato;
- inventare nuovi mercati.

L'impresa proattiva tende per sua natura ad assumere una configurazione organizzativa reticolare, flessibile, innovativa: una molteplicità di attori cooperanti che possono con facilità ristrutturare i propri ruoli e relazioni al fine di operare con flessibilità e velocità in un contesto ambientale continuamente in cambiamento. Valdani caratterizza l'impresa proattiva come *frattale, virtuale ed eterarchica*.

Frattale perché ogni parte dell'impresa è capace di intraprendere come se fosse essa stessa un'impresa, e come nei frattali matematici le configurazioni si ripresentano simili

ai vari livelli dimensionali. *Virtuale* perché si articola in team virtuali che si aggregano e disaggregano costantemente attorno ai processi della gestione, per trovare di volta in volta i modi più flessibili ed efficaci per affrontarli. *Eterarchica* perché non c'è più la tradizionale piramide dell'autorità, ma un reticolo di centri decisionali intelligenti, capaci di operare sia separatamente che in modo integrato, senza necessità di cambiare nei due casi la propria natura.

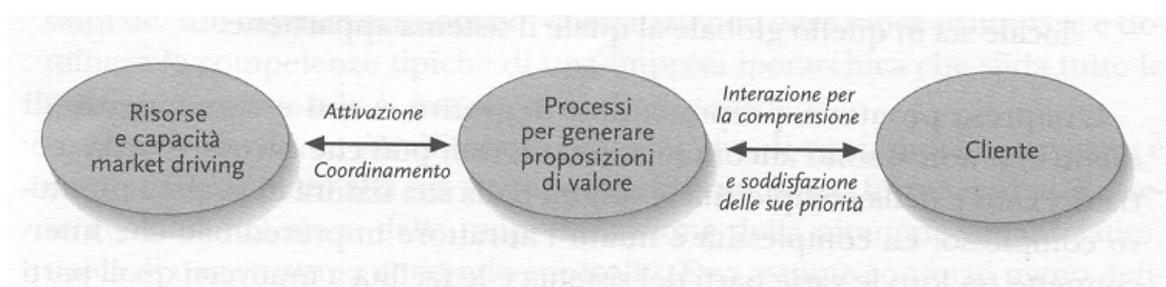
La teoria dell'impresa proattiva può essere sistematizzata in:

- tre costrutti;
- tre paradigmi;
- una metateoria.

I tre costrutti sono:

- la clientela;
- i processi per generare e trasferire alla clientela una proposizione di valore;
- lo stock e il network delle risorse e capacità market driving.

Figura 1 I costrutti dell'impresa proattiva (fonte Valdani 2000)



Il primo costrutto richiama l'attenzione sui *clienti* dell'impresa: in questa visione teorica, l'impresa esiste per servire i suoi clienti, e ogni altro suo aspetto deriva da questo. La clientela è la fonte primaria di generazione del valore per l'impresa e in un certo senso è il suo unico e vero asset patrimoniale. Non è certo un caso che le applicazioni di BI siano in misura preponderante applicazioni di analisi dei dati relativi alla clientela: sono

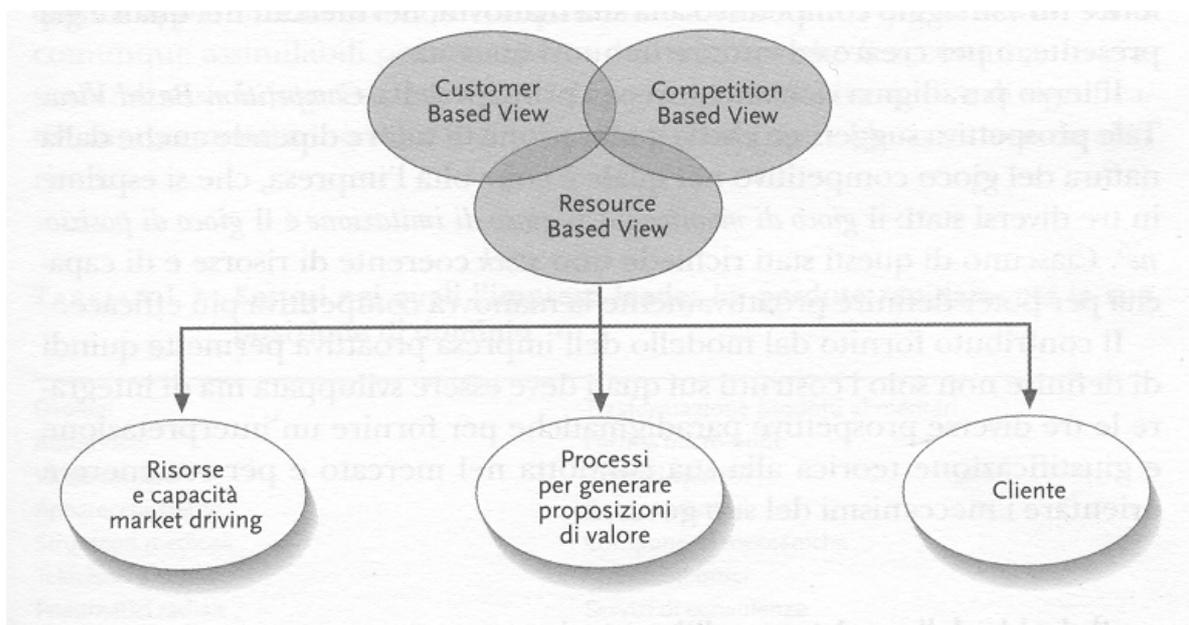
le imprese più proattive ad usare più intensamente e più consapevolmente la BI, e per loro natura sono orientate a conoscere e valorizzare al meglio la clientela.

Il secondo costrutto focalizza l'attivazione e il coordinamento dei *processi* necessari per definire e creare le proposizioni di valore da trasferire alla clientela. Queste proposizioni di valore esprimono la capacità competitiva dell'impresa di raccordare le esigenze della domanda con le risorse e le competenze che devono essere impiegate per generare prodotti e servizi capaci di offrire soluzioni di valore. L'impresa proattiva interagisce con i clienti per comprenderne le priorità, al fine di soddisfarle in modo efficace e competitivo rispetto alle altre imprese.

Il terzo costrutto enfatizza il sistema delle *risorse e capacità market driving* alle quali l'impresa può accedere, o perché le possiede al suo interno (stock) o perché sono disponibili tra fornitori e partner (network). Sono queste risorse e capacità che permettono di attivare e coordinare i processi per generare, offrire e trasferire alla clientela le proposizioni di valore.

I tre costrutti sono utilizzabili in tre diversi paradigmi concettuali.

Figura 2 I paradigmi dell'impresa proattiva (Fonte: Valdani 2000)



Il primo paradigma richiama la prospettiva della *Customer Based View*. L'enunciato fondamentale è che il valore generato da una impresa dipende dalla sua capacità di: 1) acquisire nuovi clienti, 2) sviluppare la redditività dei clienti in portafoglio, 3) fidelizzare e quindi mantenere i clienti. Nella *Customer Based View* i tipici problemi da affrontare sono: a) come le scelte e le preferenze dei clienti influenzano la possibilità di generare valore, b) come la proposizione di valore debba essere progettata e offerta per incrementare il valore della relazione con la clientela, c) come il valore risultante dalla relazione debba essere suddiviso fra acquirente e venditore.

Il secondo paradigma è propugnato dalla scuola di pensiero della *Resource Based View*. La tesi di fondo è che il valore generato dall'impresa dipende dalla specificità, qualità e varietà delle risorse e competenze che può utilizzare. La condotta dell'impresa sul mercato è quindi condizionata dalla sua capacità di *leverage* e di *building* delle risorse e delle competenze possedute per conferire alle sue manovre un vantaggio competitivo.

Il terzo paradigma richiama i principi della *Competition Based View*. In quest'ottica si analizza la generazione di valore alla luce del gioco competitivo nel quale è coinvolta l'impresa, il quale si esprime in tre stati: 1) *gioco di movimento*, 2) *gioco di imitazione*, 3) *gioco di posizione*. In ciascuno di questi tre stati l'impresa ha la necessità di costruirsi un opportuno stock e network di risorse e competenze.

Infine, dopo i costrutti e i paradigmi teorici, abbiamo una *metateoria* dell'impresa proattiva. Alcuni studiosi vedono un valido fondamento teorico per lo studio dell'impresa proattiva nella teoria della complessità, più precisamente nella teoria dei sistemi adattivi complessi. Le caratteristiche di questi sistemi più suggestive nello studio dell'impresa proattiva sono: 1) numerose e complesse interazioni fra il sistema di impresa e l'ambiente, 2) numerose e complesse interazioni tra i sottosistemi dell'impresa, 3) adattamento e innovazione continui in assenza di equilibri stabili, 4) evoluzione con discontinuità, 5) criticità dell'incorporazione nel sistema di conoscenza sull'ambiente. L'impresa proattiva è quindi un sistema complesso di relazioni fra agenti che grazie

all'apprendimento evolve in modo incessante e talvolta discontinuo reagendo e ai cambiamenti dell'ambiente e tentando di provarli a proprio vantaggio.

2. Formula imprenditoriale

La proattività si esprime in un formula imprenditoriale che possiamo riassumere con le cosiddette 6C:

1. clienti,
2. concorrenza,
3. capacità,
4. cambiamento,
5. consenso,
6. commitment.

L'impresa proattiva è tesa ad ascoltare sempre con la massima attenzione la voce dei *clienti* per soddisfarli al meglio; a questo fine essa deve rilevare le loro esigenze e preferenze presenti e anticipare quelle future, soddisfare e coltivare quelle attuali e suscitare e dirigere quelle potenziali. In questo suo sforzo, l'impresa deve tenere conto della *concorrenza*, per comprendere e anticiparne le mosse e le tendenze evolutive. Per riuscire l'impresa deve sviluppare *capacità* distintive che le diano vantaggi competitivi e le consentano di dominare i fattori critici di successo nel suo mercato. Deve essere orientata al *cambiamento*, che nell'epoca attuale è ormai un dato acquisito del gioco competitivo, accettando come normale e vitale l'innovazione di idee, capacità, comportamenti, modelli organizzativi, ruoli delle persone, valori. Deve cercare il *consenso* di tutti gli agenti rilevanti, dal personale alle istituzioni alla collettività sociale. Deve avere un forte e deciso *commitment* a perseguire la missione che si da.

La formula imprenditoriale delle 6C sintetizza ciò che una impresa proattiva deve saper fare per sostenere con successo il gioco competitivo nei mercati iperdinamici con i quali oggi deve misurarsi.

L'implementazione della formula imprenditoriale delle 6C richiede lo sviluppo continuo delle capacità market driving, le quali consentono di sviluppare continuamente le proprie condotte sul mercato e creare continuamente nuovi vantaggi competitivi, coltivando ed espandendo i mercati nei quali l'impresa opera già e aprendone di nuovi.

Alle capacità market driving sono associate in modo naturale quattro *economie*:

- economia di varietà;
- economia di velocità;
- economia di qualità;
- economia di apprendimento.

L'impresa realizza *economie di varietà* nella misura in cui riesce a realizzare la personalizzazione dei beni e servizi che offre ai suoi clienti mantenendo prezzi competitivi.

L'impresa può perseguire l'economia della varietà con tre diverse condotte:

la progettazione modulare del prodotto o dei processi di un servizio;

la standardizzazione del prodotto o del servizio intermedio;

il ricorso a impianti flessibili di produzione.

In ogni caso, il fine perseguito è la personalizzazione dell'offerta, la cosiddetta *mass customization*, che richiede capacità conoscitive per capire il cliente e la sua domanda espressa o inespressa, e capacità tecniche produttive per dare una risposta il più possibile aderente alla domanda nelle innumerevoli varianti in cui questa si può presentare.

Se l'economia della varietà consiste nel dare al cliente il prodotto nelle varianti in cui lo desidera (mantenendo condizioni di economicità), l'*economia della velocità* consiste nel dare al cliente il prodotto quando lo desidera, il che per lo più significa velocemente.

È noto che in molti dei mercati attuali (quelli che più interessano parlando di impresa proattiva) la velocità è un fattore determinante di vantaggio competitivo. Essere i primi a introdurre una innovazione nell'offerta è spesso decisivo per il successo. La velocità può essere misurata in componenti quali:

- *time to market*: tempo che intercorre fra la nascita di una idea e la sua com-

mercualizzazione;

- *throughput time*: tempo che intercorre fra l'inizio e la fine di un ciclo di produzione;
- *delivery time*: tempo che intercorre fra l'acquisizione di un ordine e la sua soddisfazione;
- *sub process time*: tempi consumati nelle diverse attività che compongono un processo.

La velocità consente a una impresa sia di migliorare la sua efficacia verso i clienti sia di aumentare la sua efficienza nei processi di produzione, diminuendo i costi.

L'*economia di qualità* si realizza nella misura in cui l'impresa riesce a far coincidere le prestazioni del prodotto con le aspettative del cliente (dimensione esterna della qualità) e con le specifiche progettuali (dimensione interna).

Una classificazione delle componenti della qualità è presentata in **Tabella 1**

Tabella 1

Dimensione interna	
Qualità fisica	Affidabilità Conformità alle specifiche Durata
Qualità funzionale	Caratteristiche tecniche e funzionali Performance Disponibilità di servizi di supporto
Dimensione esterna	
Qualità simbolica	Estetica e design Immagine Altri valori simbolici
Qualità garantita	Soddisfazione percepita Idoneità all'uso Valore riconosciuto

L'impresa realizza l'*economia di apprendimento* quando è capace di creare conoscenza e utilizzarla, il che implica:

- la sperimentazione continua di nuove idee e modalità di comportamento;
- l'apprendimento dall'esperienza;
- l'apprendimento da ciò che accade all'esterno dell'impresa;
- il trasferimento tempestivo ed efficace della conoscenza appresa a chi deve decidere ed agire.

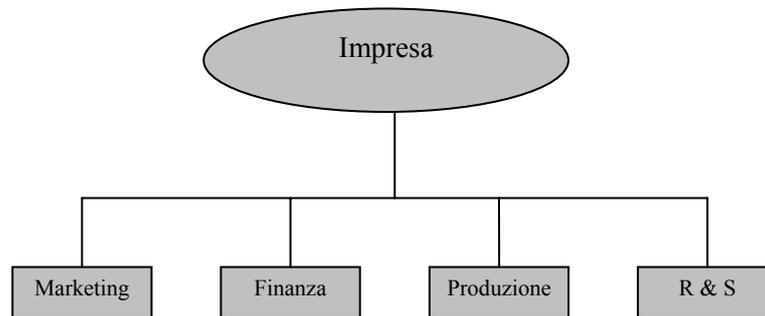
L'impresa proattiva è un sistema adattivo complesso basato sull'apprendimento e quindi non può che annettere grande importanza alla capacità di raccogliere, coltivare, trasferire e applicare la conoscenza.

Abbiamo detto che le 6C della formula imprenditoriale ispirata alla proattività richiedono le capacità market driving, che si esprimono nel conseguimento delle economie di varietà, velocità, qualità e apprendimento. Le capacità market driving richiedono poi in modo naturale due filosofie operative:

1. l'architettura reticolare e frattale;
2. la gestione per processi.

L'impresa proattiva mal si concilia con la tradizionale architettura gerarchica dell'impresa, la quale si riassume di regola in diagrammi come quello della Figura 3

Figura 3 L'impresa funzionale

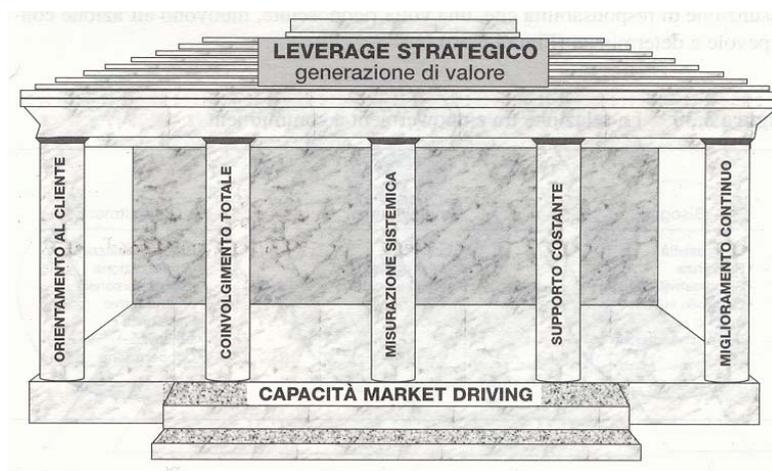


L'architettura dell'impresa proattiva non può essere semplicemente gerarchica, ma deve piuttosto essere una rete di unità capaci di agire come piccole imprese. L'intelligenza e la capacità decisionale devono essere distribuite, come accade nei sistemi adattivi complessi. L'idea è che il controllo dei processi dell'impresa sia nelle mani di coloro che concorrono alla generazione e al trasferimento delle proposizioni di valore alla clientela. In questa ottica i tradizionali compiti del management (pianificazione, organizzazione, coordinamento e controllo) assumono profili diversi, essendo distribuiti in tutta l'impresa, non più concentrati in un ristretto gruppo di individui.

La gestione dell'impresa proattiva è centrata sui processi, definibili come sequenze coordinate di attività finalizzate alla creazione di valore. L'idea di processo, in questo contesto, è orientata al valore, che è poi valore per il cliente. Alcuni processi collegano direttamente l'impresa con il cliente, per esempio i processi di gestione degli ordini, altri sono interni all'impresa, ma in ogni caso creano le potenzialità per generare valore per il cliente.

Per concludere questa descrizione dell'impresa proattiva, ricorriamo alla metafora del tempio del valore e dei suoi pilastri, illustrata graficamente in **Figura 4**

Figura 4 Il tempio e i pilastri dell'impresa proattiva [Fonte: Valdani 2000]



Delle *capacità market driving*, le fondamenta, abbiamo già detto. Esse permettono di adattare continuamente l'impresa ai cambiamenti già in atto nel suo ambiente e di provocare i cambiamenti stessi per creare un ambiente più favorevole. Le capacità market driving si esprimono nelle economie di varietà, velocità, qualità e apprendimento.

L'*orientamento al cliente*, il primo pilastro, richiede che in ogni sua azione l'impresa tenga sempre fissata l'attenzione sulla voce del cliente (le sue preferenze e aspettative), sulla migliore proposizione di valore che si può offrire al cliente, sulla soddisfazione che il cliente ha avuto dalle transizioni passate e che potrà avere da quelle future con l'impresa. Si tratta quindi di sapere come sono e cosa vogliono ora i clienti, prevedere come saranno e cosa vorranno in futuro, capire come agire per soddisfarli meglio dei competitori.

Il *coinvolgimento totale*, secondo pilastro, richiede che l'impresa sviluppi il consenso interno, la condivisione di valori ed obiettivi, e il commitment, l'impegno deciso e costante a conseguire gli obiettivi. Questo implica che l'architettura dell'impresa e la sua organizzazione, cioè il sistema di ruoli e di aspettative reciproche degli attori, siano progettati in modo consono alla filosofia della proattività, che facilitino la circolazione della conoscenza, la distribuzione del potere decisionale, l'evoluzione per apprendimento, in generale che facilitino la crescita dell'impresa come sistema adattivo complesso.

Il terzo pilastro è la *misurazione sistemica*. L'idea è di misurare i fenomeni di impresa ogni volta che questo è possibile, sensato e utile. Le misurazioni dovrebbero essere orientate alla creazione di valore, e quindi avere valenza sia conoscitiva, sia decisionale, sia motivazionale. Si tratta quindi di costruire sistemi di indicatori con certe caratteristiche, dei quali si trova un esempio nella *balanced scorecard*, sulla quale torneremo.

Il quarto pilastro è il *supporto costante*. Ciò significa che l'impresa proattiva le risorse disponibili sono reinvestite costantemente per potenziare i vantaggi competitivi attuali e crearne di nuovi. L'impresa ha quindi una visione orientata al futuro, che mira a una crescita sostenibile nel tempo

Infine, il *miglioramento continuo*, il quinto pilastro. Nell'impresa proattiva il miglioramento è dovuto alla continua innovazione. Nell'impresa proattiva l'innovazione è sistematica: non occasionale e volontaristica, ma frutto di un ciclo ininterrotto di sperimentazione, misurazione dei risultati, valutazione degli esiti, apprendimento e nuova sperimentazione.

Da ultimo, il tetto del tempio del valore, la generazione del valore per la clientela e il leverage strategico.

Generare valore per la clientela significa massimizzare il rapporto tra benefici e sacrifici che i clienti percepiscono. In un semplice modello aritmetico, se indichiamo con B il bisogno che il cliente ha di un certo prodotto (bene o servizio), con A l'abilità del prodotto dell'impresa di soddisfare questo bisogno, con C il costo o sacrificio che il cliente sostiene per acquisire e usare il prodotto, allora il valore offerto è

$$V = \frac{A \times B}{C} = \frac{I}{C}$$

dove I indica l'importanza del prodotto per il cliente.

Per massimizzare la generazione del valore si dovrà quindi individuare i bisogni, progettare e realizzare prodotti atti a soddisfarli e contenere i costi di produzione a un livello giustificato dall'importanza del prodotto per il cliente.

Il leverage strategico si può definire in forma aritmetica come

$$\text{leverage strategico} = \text{libertà di condotta} \times \text{risultati attesi}$$

La libertà di condotta è il margine di manovra dell'impresa nel cambiare la sua posizione nel mercato rispetto ai suoi competitori; i risultati attesi sono i valori che il mercato riconosce in risposta a tali cambiamenti. Il leverage strategico è il frutto delle capacità market driving e dei pilastri del valore, ed esprime la capacità dinamica dell'impresa di generare valore mantenendosi aperte le opzioni che le consentiranno di generarne ancora in futuro.

3. Customer Based View

La Customer Based View è un paradigma di analisi dell'impresa che si basa su questo assunto fondamentale: *la creazione del valore è influenzata in misura determinante dalle relazioni con la clientela.*

Il rapporto fra relazioni con i clienti e valore dell'impresa è sintetizzato da sei proposizioni:

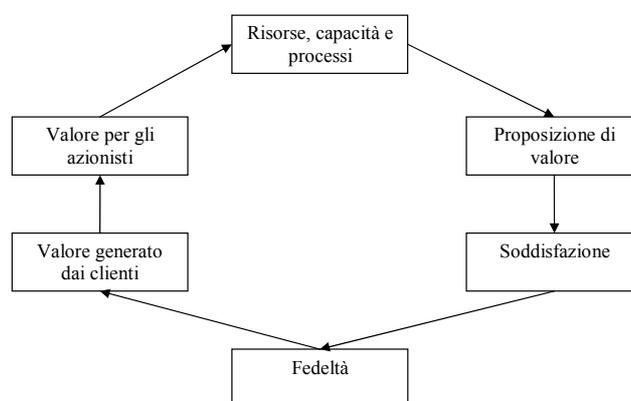
1. Il valore dell'impresa è funzione del valore delle sue relazioni con i clienti.
2. Il valore delle relazioni con i clienti è funzione della fedeltà dei clienti.
3. La fedeltà dei clienti è funzione della soddisfazione dei clienti.
4. La soddisfazione dei clienti è funzione della proposizione di valore offerta dall'impresa.
5. La proposizione di valore offerta dall'impresa è funzione delle risorse, delle capacità e dei processi dell'impresa.
6. Le risorse, le capacità e i processi dell'impresa sono funzione del valore dell'impresa.

Dunque, si afferma che il patrimonio fondamentale dell'impresa è la sua clientela. Tutti gli altri asset sono strumentali alla creazione e allo sviluppo delle relazioni con i clienti: sono queste relazioni che generano valore. Il valore delle relazioni dipende a sua volta da ampiezza, intensità e qualità della rete di relazioni con i clienti, e di queste si focalizza un aspetto critico: la fedeltà del cliente. Si afferma poi che la fedeltà del cliente dipende essenzialmente dalla soddisfazione che il cliente sperimenta nel godimento dei

prodotti dell'impresa. La soddisfazione dipende dall'adeguatezza della proposizione di valore, che dipende dalle capacità e risorse dell'impresa e dai processi che utilizzano queste capacità e risorse. Ma risorse e capacità devono essere create e continuamente potenziate dagli investimenti degli azionisti, i quali sono motivati a farlo dalla crescita attuale e potenziale del valore dell'impresa.

Si realizza così il ciclo del valore raffigurato in **Figura 5** Il ciclo del valore nella Customer Based View

Figura 5 Il ciclo del valore nella Customer Based View



Questo quadro teorico è teso a mettere il cliente e la relazione con il cliente al centro del processo conoscitivo, valutativo e decisionale. Gli assunti alla base del paradigma della Customer Based View non sono affatto così scontati come si potrebbe essere indotti a pensare. In effetti, la centralità del cliente per l'impresa è un concetto che si è affermato soltanto in tempi piuttosto recenti nella letteratura scientifica e di business, e che nella pratica non è ancora sempre e compiutamente applicato nelle imprese. Del resto, altri paradigmi dell'impresa, che descriveremo in seguito, vedono il valore generato dal cliente come funzione di altri fattori logicamente prioritari, per esempio la dotazione di risorse e le manovre competitive.

Nella concreta applicazione del paradigma concettuale, ciò che deve essere empiricamente stimato sono le intensità dei legami funzionali espressi dagli assunti di base. È intuitivo che la soddisfazione del cliente ne aumenti la fedeltà, ma non si può sapere a priori di quanto. In mercati altamente competitivi il cliente tende costantemente a cercare alternative, e quindi un piccolo incremento di soddisfazione genera un piccolo incre-

mento di fedeltà del cliente. In mercati regolamentati, in condizioni di monopolio o quasi monopolio, la fedeltà del cliente è più o meno forzata, e un piccolo incremento di soddisfazione genera un forte incremento di fedeltà. Si hanno quindi gradi diversi di intensità del legame *soddisfazione* → *fedeltà*, essendo diversa di caso in caso l'elasticità della risposta del cliente. Analogamente, non si può sapere a priori quanto la proposizione di valore influisca sulla soddisfazione del cliente, o quanto la fedeltà influisca sulla redditività delle relazioni, questa sulla motivazione degli azionisti a investire nel potenziamento della capacità dell'impresa, e quanto queste influiscano sul miglioramento della proposizione di valore e quindi sulla soddisfazione della clientela. Tutte queste stime devono essere basate su dati empirici.

4. Resource Based View

La Resource Based View è un paradigma di analisi dell'impresa che si basa su questo assunto fondamentale: *la creazione del valore è influenzata in misura determinante dalle risorse dell'impresa.*

Una *risorsa* è genericamente qualunque cosa dia un vantaggio competitivo a una impresa. Per dettagliare questa definizione introduciamo alcuni termini in uso fra i teorici del management.

Chiamiamo *risorsa (asset)* qualunque cosa una impresa possiede o controlla. Le risorse nel senso di asset possono essere:

- *finanziarie*, p.e. capacità di reperire fondi;
- *fisiche*, p.e. impianti, macchine, accesso alle materie prime;
- *umane*, p.e. esperienza, competenze, capacità decisionali;
- *organizzative*, p.e. cultura, sistemi di reporting, sistemi di controllo, relazioni informali.

Chiamiamo *capacità (capability)* un complesso di risorse che serve a compiere un processo di impresa. Per esempio, il processo di sviluppo di un prodotto implica la concettualizzazione, il progetto, il test pilota, il lancio in produzione, il perfezionamento eccetera.

Tutte le imprese hanno delle capacità; tuttavia, una impresa si focalizza su certe capacità che sono coerenti con le sue strategie. Per esempio, una impresa che persegue una strategia di differenziazione si focalizzerà sulle capacità per lo sviluppo di nuovi prodotti, mentre una impresa che persegue una strategia di leadership di costo si focalizzerà sulle capacità utili a migliorare l'efficienza del processo produttivo. Chiamiamo *competenze* (*competency*) le capacità strategicamente più importanti di una impresa.

In questo senso tecnico, una competenza è una capacità che l'impresa usa meglio delle altre; una *competenza essenziale* (*core competency*) è una competenza centrale, critica, per la strategia, la competitività e la redditività dell'impresa; una *competenza distintiva* (*distinctive competency*) è una competenza che l'impresa sa usare meglio dei suoi concorrenti.

Siamo arrivati al concetto chiave della Resource Based View, il *vantaggio competitivo sostenibile*. Questo è un vantaggio dell'impresa rispetto ai suoi concorrenti che non può essere facilmente imitato. È il vantaggio competitivo sostenibile la fonte ultima del valore, il concetto che gioca il ruolo analogo alle relazioni con la clientela nella Customer Based View. L'assunto fondamentale nella Resource Based View è che il vantaggio competitivo sostenibile deriva dal possesso di competenze distintive non imitabili dai concorrenti.

Parlando di risorse in senso ampio, comprendendo quindi asset, capability e competency, possiamo indicare [Barney 1991] delle caratteristiche che le risorse devono avere per supportare un vantaggio competitivo sostenibile. Esse devono:

1. apportare valore significativo;
2. essere rare per i competitori attuali e potenziali dell'impresa;
3. essere imperfettamente imitabili;
4. non essere sostituibili da altre risorse che apportano valore significativo ma non sono né rare né difficilmente imitabili per i concorrenti attuali e potenziali dell'impresa.

Un'altra caratterizzazione delle risorse che apportano un vantaggio competitivo sostenibile è data da questi requisiti [Peteraf 1993]:

1. eterogeneità: le risorse si differenziano da quelle dei concorrenti nel settore industriale;

2. limiti ex post alla concorrenza: per i concorrenti è difficile imitarle;
3. limiti ex ante alla concorrenza: l'impresa le acquisisce a prezzo privilegiato;
4. immobilità: le risorse non possono essere spostate dall'impresa ai concorrenti.

L'intuizione alla base della Resource Based View è che l'impresa di successo è tale perché possiede competenze diverse e superiori rispetto ai suoi concorrenti, che le danno un vantaggio competitivo che può essere difeso nel tempo, poiché i concorrenti hanno difficoltà a imitare quelle competenze.

Si parla quindi di una *asimmetria competitiva* a vantaggio dell'impresa di successo rispetto ai concorrenti.

5. Competition Based View

La Competition Based View è un paradigma di analisi dell'impresa che si basa su questo assunto fondamentale: *la creazione del valore è influenzata in misura determinante dalle manovre competitive dell'impresa.*

Per l'impresa proattiva la competizione va vista come indissolubilmente legata alla innovazione. Per avere successo l'impresa deve saper dominare il *ciclo dell'innovazione*, che schematizziamo in quattro fasi:

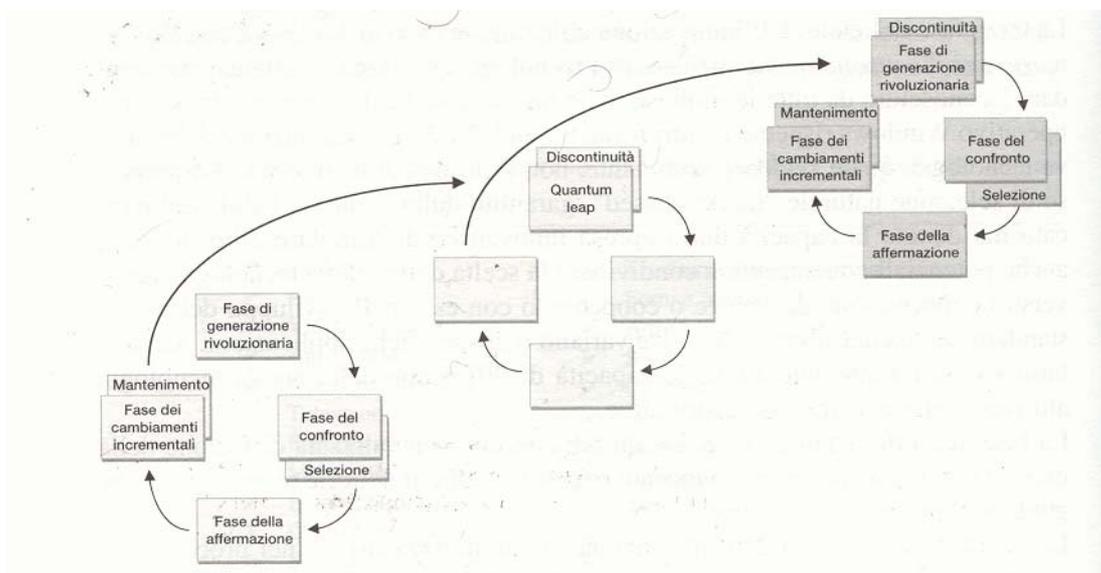
1. *Fase di generazione rivoluzionaria*: l'impresa crea una discontinuità nel gioco competitivo, introducendo una innovazione di qualsiasi genere (tecnologica, di marketing oppure organizzativa, di prodotto oppure di processo) che potenzialmente le dà un vantaggio sui suoi concorrenti.
2. *Fase di confronto*: l'impresa cerca di affermare la sua innovazione, diffonderla, farne percepire il valore, prendere il sopravvento sulle imprese non innovative e sulle innovazioni concorrenti; l'impresa è impegnata ad uscire vincente dalla selezione e rendere concreto il potenziale vantaggio competitivo.
3. *Fase di affermazione e consolidamento*: l'innovazione riesce in effetti ad affermarsi, i clienti percepiscono correttamente il valore dei prodotti che ne derivano, eventualmente altre imprese ne riconoscono la sua validità e l'adottano stipulando accor-

di con l'impresa innovatrice.

4. *Fase dei miglioramenti incrementali*: l'impresa migliora continuamente offerta basata sull'innovazione, potenziandone i benefici ed estendendoli a nuovi clienti; l'impresa è impegnata nel mantenimento del vantaggio competitivo.

Il ciclo dell'innovazione si ripete continuamente, presentando delle discontinuità, cosiddetti *quantum leap*, nella fase di generazione rivoluzionaria. La fase di mantenimento va a finire quando la stessa impresa o un suo concorrente attuano una innovazione che avvia un nuovo ciclo, come in Figura

Figura 6 [Fonte: Valdani 2000]

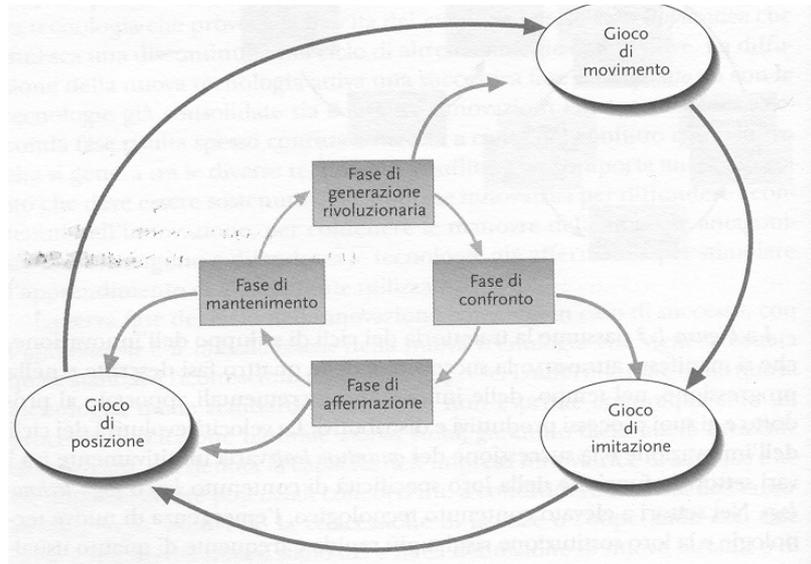


Le imprese si muovono nel ciclo dell'innovazione con manovre che si possono far rientrare in tre tipi di gioco competitivo:

1. Il *gioco di movimento* con il quale l'impresa si sottrae al confronto competitivo con i suoi rivali, spostandosi su nuovi mercati oppure soddisfacendo nuovi bisogni degli attuali mercati, in ogni caso portandosi verso un nuovo terreno competitivo.
2. Il *gioco di imitazione*, con il quale le imprese non innovatrici cercano di recuperare lo svantaggio competitivo, migliorando la propria offerta o imitando quella dell'innovatore.
3. Il *gioco di posizione*, con il quale le imprese competitive, che ormai condividono

l'innovazione affermatasi, cercano di migliorare incrementalmente le loro offerte per ottimizzare le proprie quote di mercato.

La relazione fra il ciclo dell'innovazione e i giochi competitivi è illustrata in Figura



La Competition Based View richiede di analizzare le dinamiche di questi giochi per individuare le fonti del vantaggio competitivo e quindi per aiutare la formulazione di decisioni basate su una visione più consapevole delle manovre in atto da parte dell'impresa e dei competitori e dello stato dei giochi competitivi.

6. Verso un quadro di riferimento per la business intelligence

Il quadro dell'impresa proattiva fin qui delineato secondo le tre ottiche della Customer Based View, della Resource Based View e della Competition Based View offre una attraente base teorica per sistematizzare la progettazione di un sistema di business intelligence. Come sempre accade con le applicazioni della Information Technology, anche per la business intelligence si presenta il difficile problema di comunicare al management aziendale il significato e l'utilità delle tecnologie disponibili e dei sistemi o progetti che si vanno a proporre. La visione dell'impresa che hanno i manager e quella che hanno i progettisti spesso non sono per così dire sulla stessa lunghezza d'onda. Non

intendiamo dire che si contraddicono: quando questo accade, il progetto di business intelligence resta lettera morta. Però manager e progettisti informatici usano costrutti concettuali diversi e la difficoltà di condividerli è un serio ostacolo al successo dei progetti.

In questo lavoro vogliamo delineare almeno per alcuni aspetti un sistema di business intelligence fondato con successo su teorie dell'impresa condivisibili dal management. Il nostro sistema ideale dovrebbe contenere quantomeno tre sottosistemi:

1. Customer intelligence
2. Resource intelligence
3. Competition intelligence

Questi tre sistemi dovrebbero servire al management per comprendere meglio il business aziendale e quindi prendere decisioni più consapevoli. Poiché un sistema decisionale deve avere la possibilità di elaborare il feed-back, cioè di esaminare i risultati delle decisioni già prese e delle azioni effettuate, il sistema di business intelligence dovrebbe avere un quarto sottosistema

4. Performance intelligence

che ha il compito di aiutare a comprendere gli andamenti dell'impresa.

Nel prosieguo di questo lavoro ci concentreremo sulla customer intelligence e sulla performance intelligence. La scelta della customer intelligence come argomento primario è dettata sia da una oggettiva immediatezza di applicabilità dei principi teorici sia dall'esperienza professionale dello scrivente. La scelta della performance intelligence, invece, è suggerita dall'opportunità di trattare anche un argomento per sua natura distinto dagli altri enunciati e caratterizzato dall'applicazione di concetti ben noti nell'economia aziendale, per i quali è certamente interessante indagare i rapporti con le moderne tecnologie di business intelligence.

La vastità e complessità del compito ovviamente non permettono di trattare esaurientemente l'argomento: tuttavia cercheremo di delineare almeno alcuni elementi di un quadro di riferimento generale per la progettazione di sistemi di business intelligence guidata da principi teorici tratti dalle discipline manageriali, indagando anche possibili direzioni di indagine nell'ambito di un programma di ricerca più generale, che mira a tro-

vare e formalizzare collegamenti fra due le due aree delle scienze dell'informazione e delle scienze del management di impresa, nel contesto tematico della business intelligence.

Capitolo 2

Il capitale relazionale

1. Il concetto di capitale relazionale

Il concetto di *capitale relazionale* è da alcuni anni all'attenzione degli studiosi di management [Costabile 2001, 2003], coerentemente con una generale tendenza all'attribuzione di importanza vitale alle *risorse intangibili* dell'impresa come critiche per l'acquisizione e il mantenimento di vantaggi competitivi, il conseguimento di performance elevate e la generazione del valore. Il concetto di capitale relazionale è stato definito in molti modi; per gli scopi di questa analisi, possiamo definire in prima istanza il capitale relazionale come il sistema delle relazioni che l'impresa ha con attori esterni e che essa può utilizzare per generare valore. Si vedono quindi le relazioni come *risorse*, delle quali interessano il numero, lo stato e altre molteplici caratteristiche. In questa accezione, il capitale sociale è considerato da molti come la fondamentale ricchezza dell'impresa, al quale ricondurre ogni altra forma di capitale, sia esso fisico, economico o intellettuale.

Le relazioni hanno come controparte uno *stakeholder*, un soggetto portatore di interessi verso l'impresa: i clienti, i fornitori, le istituzioni, i soggetti sociali, i lavoratori. Nel presente lavoro gli stakeholder a cui siamo interessati sono i clienti. Ci concentriamo quindi sulle relazioni che l'impresa ha con i suoi clienti, considerati ciascuno nella sua individualità e anche in forma aggregata, quindi come mercato di riferimento o suoi sottoinsiemi.

Il capitale relazionale come lo vediamo in questo contesto è dunque dato dalla quantità e qualità delle relazioni con i clienti, e quindi dalle capacità tecnologiche e di management di rapportarsi con i clienti.

Il capitale relazionale così inteso merita il nome di capitale perché quantità e qualità delle relazioni con i clienti si accumulano nel tempo, generano rendimenti e sono degli stock generati da flussi. Questo ultimo punto non è immediatamente intuitivo: che cosa

sono questi stock e da quali flussi sono generati? Una risposta è che il capitale relazionale è basato su atteggiamenti e convinzioni (fiducia e lealtà) i quali sono stock di predisposizioni originate da flussi di percezioni di soddisfazione o insoddisfazione dei clienti. Abbiamo quindi flussi di percezioni e stock di percezioni, che hanno natura di predisposizioni, atteggiamenti. L'impresa genera e accresce il suo capitale relazionale se riesce a generare soddisfazione nei suoi clienti e di conseguenza a indurre nei clienti disposizioni favorevoli all'acquisto dei prodotti dell'impresa.

Questi concetti di natura psicologica e comportamentale devono essere formalizzati e resi misurabili per poter arrivare ad utilizzare metodi e tecnologie di business intelligence. C'è l'esigenza di modelli formali che esprimano cosa è una relazione, quali sono i fattori che la fanno nascere ed evolvere, come si misura il valore di una relazione. I fattori in gioco saranno dunque le percezioni e gli atteggiamenti, concettualizzati in modi opportuni. Quando poco sopra abbiamo detto che gli atteggiamenti favorevoli dei clienti dipendono dalla loro soddisfazione, abbiamo espresso un enunciato intuitivo e ragionevole, ma non affatto scontato come può sembrare. Che la soddisfazione del cliente migliori il suo atteggiamento è esperienza comune, ma un sistema di business intelligence, per essere utile come supporto alle decisioni, deve poter indicare *come* la soddisfazione genera l'atteggiamento positivo, e soprattutto *quanto* contribuisce a generarlo, e quanto invece l'atteggiamento positivo dipende da altri fattori.

2. Il ciclo di vita della relazione

La letteratura propone molti modelli di rappresentazione del ciclo di vita della relazione tra l'impresa e il cliente. Uno dei modelli più noti è quello di Ford [Ford 1980], che identifica cinque stadi di sviluppo della relazione:

1. la *pre-relazione*, durante al quale si raccolgono le informazioni e si valutano i possibili contenuti della relazione, mantenendo ancora un basso *commitment*;
2. l'*avvio*, nel corso del quale si sostengono gli investimenti tangibili e intangibili necessari al processo di scambio;

3. lo *sviluppo*, caratterizzato da reciprocità nei processi di apprendimento e nell'investimento, quindi nel commitment;
4. la *stabilità*, caratterizzata dalla continuità degli scambi e da comportamenti che tendono anche a diventare routinari;
5. l'*istituzionalizzazione*, quando si adottano routine di gestione della relazione che minimizzano le incertezze.

Un modello più recente [Wilson 1995] propone una visione più dinamica e più operativa, articolata anch'essa in cinque fasi, alle quali sono associate delle variabili critiche:

1. *Ricerca e selezione del partner*. Si valutano le performance dei potenziali partner nella relazione per arrivare a selezionarne uno. Le variabili critiche sono la reputazione e la soddisfazione, sulle quali nelle prime interazioni si costruisce la fiducia.
2. *Definizione della portata della relazione*. Scelto il partner, si definiscono la portata e gli obiettivi della relazione, chiarendo anche le reciproche convenienze. Le variabili critiche sono la convergenza degli obiettivi e la fiducia.
3. *Specificazione dei confini dell'interazione*. Si approfondisce la reciproca conoscenza, si definiscono i confini dell'impegno reciproco, si chiarisce la plausibilità dell'interazione e della sua portata. Le variabili critiche sono la reciprocità e il commitment.
4. *Creazione di valore*. Si apprendono meglio e si potenziano le capacità dispiagate nella interazione, si crea il pieno valore della relazione e si sperimenta la sua distribuzione. Le variabili critiche sono la dipendenza, il potere, l'influenza, la reciprocità.
5. *Mantenimento*. La relazione si stabilizza e diventa routinaria. Le variabili critiche sono il coinvolgimento reciproco e la cooperazione.

Descriviamo ora un modello [Costabile 2001, 2003] che ha avuto applicazioni nell'economia digitale e che offre interessanti motivi di riflessione nell'ottica del presente lavoro.

Concetto centrale è la *percezione del valore da parte del cliente*. Tale valore può essere distinto in:

- *valore atteso*: il rapporto tra i benefici che ci si attendono dall'acquisto e dal godimento del prodotto e i costi che si ritiene di dover sostenere, anche in relazione alle alternative disponibili;
- *valore percepito dopo l'acquisto e l'uso*: deriva dalla soddisfazione o insoddisfazione e viene confrontato con il valore atteso;
- *valore monadico*: il valore percepito dopo le prime esperienze d'uso in termini comparativi rispetto alle alternative considerate; è detto monadico perché interno all'individuo cliente, che compara le esperienze certe maturate nella relazione con l'impresa con le aspettative ipotetiche verso le alternative;
- *valore diadico*: rapporto tra il valore che il cliente ritiene di avere ottenuto e quello che ritiene di aver generato per l'impresa; è detto diadico perché interno alla diade cliente-impresa, ed è anche detto *valore di equità*.

Queste quattro forme del valore sono percepite diversamente nel corso della relazione fra il cliente e l'impresa, e tali variazioni danno al modello il suo carattere dinamico, poiché le diverse configurazioni del valore percepito distinguono le varie fasi della relazione:

1. Fase della soddisfazione e della accumulazione di fiducia

La preferenza del cliente per una certa scelta di acquisto è fondata sulla percezione del valore differenziale che il cliente ritiene di poter ottenere rispetto alle alternative. Il valore atteso per il cliente è il rapporto tra i benefici attesi dall'acquisto e i sacrifici attesi per l'acquisto

$$V = \frac{B}{S}$$

La scelta d'acquisto si fonda sulle aspettative di valore e la percezione di soddisfazione dipende dalla congruenza tra il valore atteso e il valore percepito, come dimostrato da numerose ricerche empiriche. Tale percezione può essere vista come il “flusso” di cui si è parlato, che alimenta lo “stock” della fiducia, intesa come atteggiamento di pregiudizio positivo verso l'impresa o il suo prodotto.

2. Fase della fedeltà comportamentale

Le esperienze di acquisto e consumo caratterizzate da soddisfazione del cliente alimentano la tendenza al riacquisto e progressivamente portano a far evolvere la relazione verso la fedeltà, intesa come tendenza comportamentale verso il riacquisto.

Una interpretazione di questo fenomeno è che la fiducia diminuisce i costi di transazione, per cui il riacquisto diventa economicamente più conveniente. La fiducia diminuisce:

- i *costi cognitivi*, perché il cliente ha minore bisogno di raccogliere informazioni sulle alternative di acquisto;
- i *costi emotivi*, perché il cliente percepisce minore incertezza e rischio;
- i *costi operativi*, perché il cliente non deve spendere tempo e denaro nell'esame delle alternative;
- i *costi strutturali* del cambiamento, perché il cliente non deve riattrezzarsi per la fruizione di beni diversi, p.e. cambiando interfacce o accessori.

La seconda fase del ciclo di vita della relazione ha una durata che dipende dal tipo di prodotto o servizio, dalle caratteristiche del cliente, dal suo coinvolgimento, dalla competizione nel mercato, dallo stato della tecnologia incorporata nel prodotto o servizio. Durante una serie di esperienze il cliente aggiorna la sua valutazione comparativa con le alternative di acquisto. Se l'esito è positivo la relazione entra nella sua terza fase.

3. Fase della fedeltà mentale

La fedeltà mentale è la convinzione del cliente che l'impresa sarà capace di mantenere nel tempo un differenziale di valore positivo nei confronti dei suoi concorrenti. In questo stadio il cliente è molto stabile e disponibile a mantenere e ampliare la portata della relazione. La sua attenzione verso le offerte della concorrenza si riduce e può anche essere disposto a mantenersi fedele nonostante l'emergere di influenze contrastanti, come nuovi evidenti vantaggi nell'offerta di concorrenti.

4. Fase della lealtà

In questa quarta fase l'attenzione del cliente si sposta in misura maggiore verso il valore diadico o valore di equità: il cliente considera attentamente il rapporto fra il valore che percepisce di ricevere dall'impresa e quello che ritiene di generare per l'impresa. Il cliente vede il valore per sé come il rapporto B/S fra benefici e sacrifici derivanti dall'acquisto e uso del prodotto, e il valore per l'impresa come il rapporto R/C fra ricavi e costi per l'impresa, e si attende che i due rapporti siano più o meno di pari entità:

$$\frac{B}{S} \cong \frac{R}{C}$$

Se il cliente ritiene che lo scambio sia equo, perviene alla lealtà, alla cosiddetta *customer loyalty*. Si entra allora in un processo di co-evoluzione dell'impresa e del cliente.

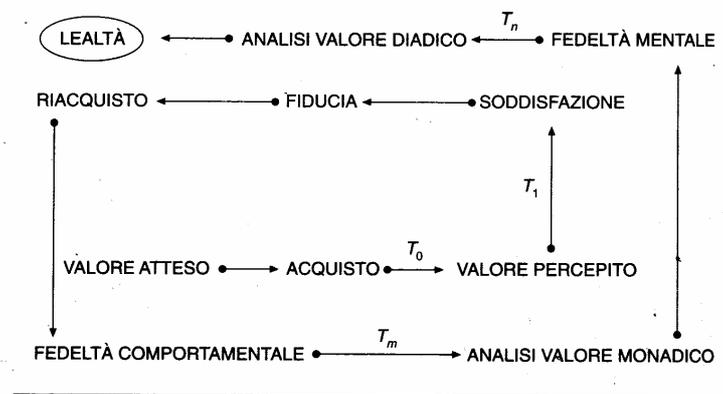
Ricapitoliamo alcune definizioni importanti:

- La *soddisfazione* è il risultato di valutazioni/interazioni post-acquisto nelle quali il valore percepito è in linea con il valore atteso.
- La *fiducia* è un pregiudizio dei clienti riguardo alla capacità dell'impresa di offrire un valore in linea con le proprie aspettative. E' uno stock che nasce da flussi ripetuti di customer satisfaction.
- La *fedeltà comportamentale* consegue alle economie della fiducia (riduzione costi cognitivi, emotivi, strutturali e operativi).
- La *fedeltà mentale* nasce dalla convinzione di costante superiorità del valore offerto dall'impresa rispetto ai concorrenti.
- La *lealtà* segue la fedeltà mentale e ha origine dalla convinzione che l'impresa sia equa e non opportunistica (valore ottenuto coerente con il valore generato).

Possiamo sintetizzare dicendo che in questo modello la customer loyalty è un costrutto multidimensionale, con due dimensioni comportamentali (fedeltà comportamentale e comportamenti cooperativi) e tre dimensioni cognitive (fiducia, valore monadico e valore diadico).

Riportiamo in figura una rappresentazione del modello della customer loyalty tratta da [Costabile 2003]

Figura 7



Nella figura sono indicati i tempi di sviluppo della relazione, per dare una idea visiva del progresso nelle varie fasi.

Riepilogando, possiamo suddividere la vita della relazione in cinque fasi:

1. soddisfazione
2. fiducia
3. fedeltà comportamentale
4. fedeltà mentale
5. lealtà

A questo modello possiamo utilmente connetterne un altro, che rappresenta il comportamento del consumatore in modo che si potrebbe definire canonico, essendo alla base della gran parte degli studi:

1. percezione del bisogno
2. ricerca delle informazioni
3. valutazione delle alternative
4. decisione di acquisto
5. valutazione post-acquisto

Non commentiamo questo modello, per il quale esiste una letteratura vastissima, perché in questa formulazione estremamente sintetica appare intuitivamente naturale.

3. Modelli della relazione e tecnologie di business intelligence

Il modello della relazione fin qui descritto presenta alcuni aspetti interessanti nell'ottica della business intelligence. I concetti utilizzati appaiono passibili di traduzione in termini operazionali, è cioè possibile tradurli in misure e correlazioni tra misure. Inoltre, essendo dinamico, il modello permette di concettualizzare l'evoluzione nel tempo dei rapporti con i clienti, aspetto importante in sede di predizione del comportamento e di valutazione del capitale relazionale, passi cruciali del momento decisionale strategico.

Negli anni '90 le tecnologie di business intelligence hanno trovato applicazioni interessanti ma per certi aspetti pionieristiche e talvolta ingenua, come del resto è perfettamente naturale per tecnologie nuove e in sviluppo. Tali applicazioni non avevano, spesso, basi teoriche sufficientemente forti, e del resto non era questa la giusta priorità al momento. Oggi è invece realistico, a nostro giudizio, rivedere l'impostazione dei progetti di business intelligence, e segnatamente della progettazione di data warehouse e di applicazioni di data mining, allo scopo di renderli più coerenti con i principi teorici delle discipline manageriali.

Prendiamo come esempio le applicazioni di data mining nel campo dell'analisi delle vendite. L'esempio è particolarmente significativo, dato che proprio queste sono le applicazioni più note, sia nella letteratura manageriale che in quella scientifica. Tipicamente, le analisi si sono basate su dati comportamentali: quali clienti acquistano quali articoli e quali articoli compaiono frequentemente insieme nelle stesse transazioni (la ben nota *market basket analysis*). Uno dei possibili risultati di un'analisi comportamentale è un sistema di regole associative, delle quali diamo due esempi formalmente e algoritmicamente simili ma concettualmente diversi, tratti da [Han 2001]:

- $age(X, "30...39") \wedge income(X, "42K...48K") \Rightarrow buys(X, "high resolution$

TV”)

- $buys(X, \text{“computer”}) \Rightarrow buys(X, \text{“financial management software”})$

dove X rappresenta un generico cliente e ciascuna regola è corredata di valutazioni sulla sua significatività statistica. La prima regola individua una tendenza di acquisto di una certa categoria di clienti, denotati in base alle loro caratteristiche descrittive intrinseche, l'età e il reddito; la seconda individua una tendenza di acquisto dei clienti che hanno un certo comportamento (l'acquisto di computer), e quindi fa riferimento al solo comportamento dei clienti, senza evidenziarne le caratteristiche descrittive.

Le due tipologie di regole fanno dunque riferimento a informazioni sul cliente di due nature diverse: descrittive e comportamentali. Entrambe le tipologie di dati sono *osservabili*, cioè possono essere rilevate oggettivamente. Se vogliamo utilizzare modelli come quelli delineati, è però utile, a volte forse necessario, trattare anche dati non osservabili, che servono a ricostruire i processi di generazione di quelli osservabili. In altri termini, appare desiderabile che i data warehouse e le applicazioni di data mining diventino più coerenti con i modelli teorici delle discipline economiche e aziendali, in particolare del marketing, e che i dati e gli algoritmi siano più chiaramente progettati su basi teoriche.

Nella fattispecie, inserire i modelli teorici nel data warehouse e nel data mining può significare introdurre una rappresentazione esplicita degli stati della relazione e del comportamento del cliente. Come il cliente ha un attributo *età* così potrà avere l'attributo *stato della relazione*, attributo non direttamente osservabile che ha un senso definito solo all'interno di un modello teorico.

Torneremo in seguito su questi problemi. Prima dobbiamo parlare di un concetto essenziale: il valore delle relazioni. Se infatti vogliamo che le applicazioni di business intelligence siano fondate su concetti di teoria economico-aziendale, non possiamo prescindere da questo che è in ultima analisi l'obiettivo di ogni analisi, decisione e azione aziendale: creare valore.

4. Il customer equity aggregato

Oltre a modellare la natura, le fasi e soprattutto i fattori determinanti delle relazioni con i clienti, per prendere decisioni consapevoli nella gestione delle relazioni stesse è necessario riuscire a stimare il valore delle relazioni stesse, sia quelle in essere sia quelle potenzialmente attivabili.

Un concetto fondamentale nel *relationship marketing* è il valore del cliente esteso all'intero suo ciclo di vita in quanto cliente, noto come *customer lifetime value* oppure come *customer equity*.

Un modello di customer equity ben noto è quello di Wayland e Cole [Wayland 1997]:

$$CE = \sum_{t=1}^n (Q_t M_t) d^t - \sum_{t=1}^n (S_t + F_t) d^t - A_1$$

In questa formula CE rappresenta la customer equity, il valore del cliente, n è il numero di periodi di tempo da considerare, Q_t è la quantità di acquisti al tempo t , M_t è il corrispondente margine, d è il tasso di attualizzazione, S i costi di sviluppo della relazione, F i costi di fidelizzazione, A_1 i costi di acquisizione del cliente, costi che qui si presuppone siano sostenuti interamente nel primo periodo.

Il significato del parametro d è ben noto dalla matematica finanziaria: un valore di 0.9 per questo parametro indica che un euro del periodo t è considerato di valore *attuale* equivalente a 0.9 euro del periodo $t-1$.

Il significato dei parametri Q e M è intuitivo, il che beninteso non implica che la loro stima sia semplice. Anche il significato del parametro A_1 è abbastanza intuitivo. In tutti i tre casi la stima puntuale di questi parametri richiede un sistema di determinazione dei costi sofisticato. Occorre però riconoscere che l'utilità di modelli come questo non viene annullata dalla impossibilità di stime precise dei parametri: anche stime approssimate e soggettive possono comunque aiutare il processo decisionale,

Non è invece del tutto ovvio il significato dei parametri S e F . C'è una distinzione fra sviluppo della relazione e fidelizzazione che deve essere chiarito. Una interpretazione è che S rappresenti i costi da sostenere per sviluppare la portata della relazione, e questo può significare che il cliente acquista maggiore quantità dei prodotti che già acquista, oppure prodotti simili ma di più elevato valore, oppure ancora prodotti diversi. Il parametro F rappresenta allora i costi da sostenere per condurre il cliente verso lo stadio di lealtà nella relazione. La distinzione è comunque sottile e nell'applicazione pratica della formula di Wayland e Cole il problema di definire con precisione i due concetti è critico.

Un altro elemento di ambiguità nella formula è dato dai parametri temporali n e t . In primo luogo, è da stabilire l'orizzonte temporale entro il quale si intende calcolare il valore della relazione con il cliente. In secondo luogo, è da stabilire quale debba essere la grana temporale da usare nella formula: può essere un periodo prefissato, come l'anno o il mese, o può essere qualcosa di più complesso.

Questi problemi connessi all'ambiguità della formula saranno affrontati in seguito. Qui vogliamo solo evidenziare che il calcolo del valore di un cliente non è semplice come potrebbe sembrare ad una prima considerazione, in virtù della semplicità algebrica della formula.

Veniamo ora ad un secondo famoso modello per la customer equity, quello di Blattberg e Deighton [Blattberg 1996], che mira a determinare l'allocazione ottimale del budget per la fidelizzazione della clientela:

$$(am - A) + a \cdot \left(m - \frac{R}{2}\right) + \frac{r^n}{1-r^n}$$

Qui a è il tasso di acquisizione di nuovi clienti (rapporto fra acquisti e clienti potenziali), m è il margine economico medio su ciascuna transazione, A è il costo di acquisizione per cliente, R è il costo annuo di mantenimento per cliente (il periodo di riferimento è l'anno perché il modello parla di budget), r è il tasso di ritenzione della clientela (rapporto fra clienti mantenuti a fine anno e totale clienti a inizio anno), r^n è $r/(1+d)$, d è il tasso di sconto. In questa formula sono da determinare i parametri a e r di acquisizione e ritenzione.

5. Il customer equity individuale

I due modelli presentati sono ritenuti tra i più interessanti e utili per le applicazioni reperibili nella letteratura. Senza volerne assolutamente sminuire l'interesse, in questo lavoro sosteniamo la tesi che questi modelli possono essere resi più utili dal punto di vista applicativo se si colgono certe opportunità offerte dalle nuove tecnologie per la business intelligence. Proviamo quindi a riconsiderare questi modelli alla luce delle nuove tecnologie.

Una prima considerazione è che questi modelli sono relativi all'intera popolazione di clienti, non distinguendo fra singoli clienti. Le differenze individuali contribuiscono alla definizione dei parametri del modello perché questi sono grandezze di insieme: il tasso di *retention* è un rapporto fra due statistiche (numero di clienti a fine e a inizio periodo), il costo di acquisizione è un costo medio e così via. Se però si dispone di validi *customer database*, allora questa impostazione è inutilmente restrittiva. I modelli del valore della base clienti possono essere interpretati non più in termini di grandezze di insieme, ma in termini di probabilità individuali. Per esempio, il tasso di retention diventa la probabilità che un determinato individuo che è cliente all'inizio del periodo resti tale alla fine. La nuova interpretazione mantiene le proprietà formali della grandezza statistica di insieme ma offre anche nuove opportunità.

Se nel data warehouse ogni cliente è una entità autonoma descritta con ricchezza di dati e se con il data mining si ha a disposizione un sistema di predittori, allora si può pensare di calcolare un customer equity individuale e muovere le leve decisionali in modo differenziato per gruppi di clienti, o al limite proprio per singoli clienti. Svolgiamo alcune considerazioni in merito.

Supponiamo di voler lanciare una campagna di marketing per l'acquisizione di nuovi clienti. Disponiamo di un *customer data warehouse* che contiene dei potenziali clienti (detti *prospect*) oggetto della nostra campagna. Vogliamo stimare il customer equity creato dalla campagna (non quello totale, ma quello addizionale) per valutare la convenienza economica ad attuare la campagna stessa.

Per un qualsiasi prospect indichiamo con p_a la probabilità di acquisizione in risposta alla nostra campagna. Se disponiamo di modelli predittivi possiamo stimare p_a in base alle informazioni descrittive o comportamentali sul prospect; potremmo avere regole del tipo $X \rightarrow A$ che indicano la tendenza di chi possiede certe caratteristiche X (età, comportamenti passati eccetera) a rispondere positivamente a campagne come quella che intendiamo attivare diventando nostro cliente. Supponiamo che

1. il prospect in esame abbia le caratteristiche X ;
2. questa regola abbia nel data warehouse un *supporto* significativo, cioè che sia presente in un numero statisticamente significativo di casi;
3. questa regola abbia una *confidenza* del 30%, cioè che statisticamente un prospect con caratteristiche X risponda positivamente nel 30% dei casi.

È allora ragionevole attribuire al prospect in esame una probabilità $p_a = 0.3$ di diventare nostro cliente. Le cose si complicano se il prospect ha caratteristiche simili ma non proprio uguali a X . Se la caratteristica X è di avere età compresa fra 30 e 39 anni e il prospect ne ha 40, si pone il problema se ignorare la regola oppure usarla in versione più debole, per esempio diminuendo a al 25%. Ancora, può darsi che la regola abbia nella sua parte di destra A una campagna passata non proprio analoga a quella che intendiamo lanciare, e anche in questo caso la regola dovrebbe essere indebolita in qualche modo. È poi possibile che per il prospect si possano applicare più regole, e in questo caso occorre decidere come combinarle. Si possono porre problemi dei più vari generi, e occorre definire algoritmi per stimare p_a in base alle regole disponibili.

Un caso realistico è che la caratteristica X consista nell'appartenere a un certo gruppo di prospect, individuato in precedenza, che mostra omogeneità di risposta alle campagne di marketing. Se così accade, occorre avere criteri algoritmici per decidere se il prospect può a buon diritto essere classificato in questo gruppo e di conseguenza se la regola può essere applicata.

In tutti questi esempi gioca un ruolo critico il concetto di *similarità* (o il concetto complementare di *distanza*), fra prospect ma anche fra campagne: per poter usare la conoscenza disponibile è necessario avere criteri per determinare se e quanto il caso in esame rientra nei casi già noti. È anche possibile usare misure di similarità per estendere modelli di predizione al di là del loro campo di applicazione naturale: se il prospect in esa-

me non rientra abbastanza bene in nessuna delle regole disponibili, non è possibile attribuirgli direttamente un valore di a ; se però somiglia abbastanza a prospect per i quali il valore è calcolabile, allora si può assegnargli un valore per analogia con loro. Per esempio, se il prospect P è simile ai prospect P_1 e P_2 e si situa a metà strada fra loro, e se gli altri due hanno valori $p_a = 0.50$ e $p_a = 0.60$, allora è ragionevole assegnare a P un valore $p_a = 0.55$, anche se non ci sono i requisiti per stimare direttamente la sua probabilità di acquisizione.

Ancora una possibilità, interessante quanto complessa: potremmo avere regole che caratterizzano i prospect in base al loro stato X nella relazione con l'impresa. La regola di associazione potrebbe essere del tipo "se il prospect ha un grado 3 di fedeltà comportamentale, allora la probabilità che diventi cliente in risposta alla campagna è del 75%", qualunque cosa significhi avere grado 3 di fedeltà comportamentale nel modello di relazione adottato dall'impresa. Si può capire come la rappresentazione nel data warehouse del grado di fedeltà sia questione complessa e impegnativa; tuttavia la prospettiva di avere informazioni di questo genere è estremamente attraente per un esperto di marketing che deve prendere decisioni sul lancio di una campagna di acquisizione clienti.

Ammettiamo che in qualche modo il sistema di business intelligence sia in grado di stimare il valore del parametro p_a per ogni prospect. Indichiamo ora con Q_a e M_a la quantità e il margine unitario prevedibili per il primo acquisto del prospect (che così diventa cliente). Anche queste quantità potranno essere stimate con metodi simili a quelli indicati per p_a (con maggiore o minore credibilità, naturalmente). Si noti come in questa stima siano compresi fenomeni complessi come la propensione all'acquisto e la sua elasticità rispetto al prezzo.

Dunque, la prima vendita al prospect porta un margine atteso di

$$p_a Q_a M_a$$

In un approccio classico il parametro a avrebbe il significato di una frazione e dovrebbe essere moltiplicato per il numero N dei prospect; i parametri Q_a e M_a sarebbero stimati come medie sullo storico dei nuovi clienti, oppure il loro prodotto sarebbe sostituito dalla differenza tra il fatturato dell'impresa e i suoi costi, estraendo le informazioni di-

rettamente dal sistema informativo dell'impresa (questo è quanto accade in [Blattberg 2001]). La differenza fra le due impostazioni è lieve sul piano formale nei modelli di alto livello, ma diventa netta quanto si entra nei dettagli sostanziali. L'approccio classico presuppone la contabilità e il bilancio, l'approccio che stiamo presentando presuppone un data warehouse che descrive dettagliatamente ciascun cliente e prospect e algoritmi di data mining che forniscono predizioni a livello individuale per i clienti e i prospect. È evidente che senza questi requisiti si deve adottare l'approccio classico; ma se i requisiti sono soddisfatti, allora il metodo classico mostra il suo limite. Questo limite non è concettuale nel senso economico, manageriale e di marketing: i modelli classici sono del tutto apprezzabili, e il nostro non aggiunge niente da questo punto di vista. Ma il punto è che i modelli classici *accettano a priori le approssimazioni dovute all'uso di grandezze di insieme e non individuali*.

In un contesto stazionario questo non comporta errori rilevanti nelle stime: se i clienti nuovi entranti di oggi sono simili a quelli di ieri, allora predire per ciascuno di loro un valore uguale a quello medio dei clienti di ieri risulta corretto per motivi statistici. Calcolare un valore atteso per ogni nuovo cliente e poi sommare su tutti i prospect porta allo stesso risultato che si ottiene calcolando direttamente una proporzione sui dati storici aggregati. La disaggregazione e la successiva riaggregazione dei dati portano a un risultato che era ottenibile in modo più semplice. Se la nuova campagna colpisce un numero di clienti che è la metà della precedente, è ragionevole prendere i numeri complessivi della precedente, dimezzarli e usare i risultati come predizioni. Può accadere però che un numero rilevante dei prospect della nuova campagna siano per qualche aspetto diversi dalla media dei prospect della campagna precedente. In tempi di cambiamenti rapidi questo può accadere molto spesso. Se così è, allora la disaggregazione permette di tenere meglio conto delle differenze individuali e quindi di effettuare stime più puntuali, e in tal caso la riaggregazione porta a risultati diversi da quelli ottenibili con un approccio più di insieme.

Questa esigenza di comprensione puntuale, diciamo *disaggregata*, dei clienti rientra in modo naturale nella filosofia dell'impresa proattiva e sta alla base di tendenze attuali del marketing come il *customer relationship management* e il *marketing one-to-one*. La tesi che qui sosteniamo è che i metodi classici sono troppo pessimisti sotto il profilo compu-

tazionale: il loro retroterra culturale risale a tempi (recenti) nei quali la disponibilità di dati e la potenza di calcolo non consentivano analisi disaggregate molto fini. Oggi la situazione è diversa, se non sempre almeno in molti casi pratici, ed è quindi possibile rivedere certi concetti pensando a come disaggregare il processo di analisi.

Tornando al modello, abbiamo detto che la prima vendita al prospect neo-cliente porta un margine atteso uguale a $p_a Q_a M_a$. Dobbiamo però considerare il costo della campagna da imputare al prospect, cioè il costo dell'azione di marketing che ha trasformato il prospect in cliente. Questo può essere fatto sottraendo al margine il costo della campagna di acquisizione C_a diviso per il numero di destinatari N_a . Il valore derivante dall'acquisizione del nuovo cliente diventa allora

$$V_a = p_a Q_a M_a - \frac{C_a}{N_a}$$

Si noti che il costo totale della campagna viene diviso per il numero di tutti i prospect destinatari, non solo di quelli che rispondono positivamente acquistando: il giudizio economico di convenienza della campagna deve essere globale, non individuale.

Il valore trovato può essere interpretato come il valore atteso per il prospect *limitatamente al primo periodo considerato*. Il valore atteso totale per l'intera popolazione di prospect sarà allora la sommatoria di questi valori calcolati per ogni prospect:

$$V_a^T = \sum_{i=1}^{N_a} V_a^i$$

dove V_a^T è il valore atteso totale della campagna e V_a^i è il valore atteso per il primo periodo calcolato sul singolo prospect di indice i .

A fini decisionali dobbiamo confrontare i costi della campagna con il customer equity addizionale che la campagna crea. Dobbiamo allora stimare non soltanto il valore creato direttamente dalla campagna nel primo periodo, ma anche i valori futuri.

Facciamo un esempio. Supponiamo che la campagna di acquisizione abbia un costo totale $C_a = 100.000$ €. Analizziamo un singolo prospect e, usando qualche metodo di

predizione, stimiamo che la probabilità di acquisirlo sia $p_a = 0.25$, cioè che su 4 prospect “simili” ci si possa attendere che uno sia acquisito come cliente. La stima del margine atteso è $Q_a M_a = 200$ €, vale a dire che se acquista allora ci porta un margine (ricavi meno costi) di 200 €. Il valore atteso per questo prospect è allora di 50 €. Sommando su tutti i prospect otteniamo un valore atteso totale di 50.000 €. In termini classici troveremmo questo risultato in un altro modo. Diremmo per esempio che ci attendiamo statisticamente di acquisire 1 cliente ogni 8, e poiché la popolazione target è di 8.000 prospect, allora supponiamo di acquisire 1.000 clienti, ciascuno dei quali in media apporta 50 € di margine di vendita. Ragioneremmo quindi *top-down*, partendo da aggregati di prospect e applicando con grandezze statistiche, essenzialmente delle medie. Procedendo invece *bottom-up* partiamo da individui e applichiamo probabilità che derivano da algoritmi di predizione.

Avendo ottenuto un ritorno di 50.000 € potremmo concludere che la campagna non è economicamente giustificata. Questo sarebbe errato, perché i ritorni della campagna non si esauriscono con il primo periodo, cioè con l’acquisizione del nuovo cliente: dopo la *acquisition economics* occorre considerare la *retention economics*. Una volta che il prospect è diventato cliente ci attendiamo che iteri i suoi acquisti e che quindi porti ulteriore valore.

Per stimare tale valore ci interessa non più la probabilità di acquisizione ma quella di ritenzione. Formalmente le cose non cambiano e il valore atteso da un cliente in un certo periodo è

$$V_r = p_r Q_r M_r - \frac{C_r}{N_r}$$

dove la sostituzione dell’indice r all’indice a indica che stiamo trattando il marketing di ritenzione, non quello di acquisizione. Perciò, C_r rappresenta il costo di marketing per il mantenimento del cliente.

La probabilità che il prospect diventi cliente alla fine del periodo 0 (quello della campagna) è p_a ; la probabilità che sia ancora cliente alla fine del periodo t è $p_a p_r^t$. Se è ancora cliente, allora apporta un margine di vendita $Q_r M_r$. Per essere più rigorosi, diciamo

che la probabilità di ritenzione dalla fine del periodo $t - 1$ alla fine del periodo t è $p_r(t)$, ammettendo che possa variare nel tempo. Se parametrizziamo anche le altre grandezze rispetto al tempo, e in definitiva rappresentiamo il valore del cliente i nel periodo t come

$$V_r(t) = \left(\prod_{k=0}^t p_r(k) \right) Q_r(t) M_r(t) - \frac{C_r(t)}{N_r(t)}$$

dove per uniformità di notazione abbiamo posto $p_a = p_r(0)$.

Abbiamo dunque una successione temporale di valori del cliente, nella quale essendo realisticamente $p_r < 1$ i valori tendono a decrescere. A questo punto applichiamo ad ogni valore della successione un fattore di attualizzazione e arriviamo ad una formulazione definitiva del valore del cliente:

$$V = \sum_{t=0}^{\infty} V(t)(1+d)^{-t}$$

dove d è il tasso di attualizzazione.

Il giudizio di convenienza sull'intera campagna sarà effettuato verificando che la somma dei valori così calcolati per tutti i clienti sia positiva. Una alternativa aritmetica è di scorporare dall'espressione del valore i costi di marketing e far coincidere il valore con il margine di vendita; poi si confronta la somma dei margini di vendita con la somma dei costi di marketing. I due procedimenti di calcolo sono equivalenti. Il secondo metodo è più immediato se si pone $C_r(t) = 0$ per tutti i $t > 1$, il che significa che la campagna è *una tantum* e che la retention è spontanea, ipotesi che può essere applicabile in un certo numero di casi pratici.

Per esempio, se la somma dei margini di vendita futuri attualizzati estesa a tutti i prospect è 80.000 €, questo valore deve essere sommato ai 50.000 € di margini di vendita da acquisizione e in definitiva si ha un margine di vendita totale attualizzato di 130.000 € che supera il costo una tantum della campagna fissato a 100.000 €. Il giudizio di convenienza è dunque positivo, anche se nel primo periodo il saldo era negativo.

Il metodo bottom-up qui illustrato può giungere a conclusioni simili a quelle di un classico metodo top-down, se le predizioni individuali dell'uno sono congruenti alle ipotesi statistiche aggregate dell'altro. Abbiamo già accennato ad alcune differenze nelle implicazioni dei due metodi. Un'altra considerazione interessante è che procedendo bottom-up viene naturale ordinare i clienti secondo il valore che ci si attende da loro e quindi concentrare l'attenzione sui più redditizi. Questo può essere molto importante in presenza di campagne a budget limitato o comunque di convenienza non evidente. Nel caso specifico del direct marketing il problema dell'individuazione dei clienti più promettenti è un tema cruciale, sia per gli operatori che per i ricercatori. Tradizionalmente viene affrontato con metodi aggregati, ma l'uso del data warehousing e del data mining è visto oggi come il più promettente per l'ottimizzazione delle campagne.

6. Il customer equity potenziale

La valutazione dei clienti sulla sola base dei flussi di cassa attesi attualizzati appare limitativa. Questo non perché non siano importanti i flussi di cassa, ma perché il procedimento appare troppo lineare: il futuro della relazione è visto come una sequenza che ad ogni passo genera un flusso di cassa, fino al passo in cui si interrompe in modo definitivo. Questo modello prevede sì un ruolo per la probabilità, ma in un certo senso è lo stesso deterministico: il processo può solo progredire su un binario rettilineo, con l'unica alternativa di fermarsi.

Si avverte l'esigenza di inserire nella valutazione anche dei benefici potenziali, che possono manifestarsi o meno nel corso della vita della relazione e che non sono così strettamente connessi all'acquisizione e al mantenimento di un cliente.

Una prima categoria di benefici potenziali riguarda l'influenza delle relazioni sulle risorse immateriali. Nell'impresa proattiva le risorse immateriali hanno importanza vitale, come condizioni per la generazione di valore in un contesto ipercompetitivo e costantemente teso all'innovazione. In [Pels 1992] si sottolinea l'importanza di valutare le relazioni sia riguardo alla capacità di generare flussi di cassa sia riguardo alla capacità di

potenziare la dotazione di risorse immateriali dell'impresa. Un cliente può infatti contribuire ad accrescere la conoscenza dell'azienda (che impara a soddisfare meglio le sue richieste, che rappresentano o anticipano quelle degli altri clienti) e la sua reputazione (che beneficia del capitale relazione del cliente stesso, le cui relazioni giocano a favore dell'impresa).

Questi concetti sono interessanti, ma non facilmente traducibili in termini operativi rilevanti per le tecnologie di business intelligence. La cosa non è impossibile: per esempio le tecniche di *link analysis* usate nel data mining potrebbero essere di aiuto nel valutare il network di relazioni di un cliente come ulteriore risorsa dell'impresa. Tuttavia non ci sono note applicazioni concrete di questa tecnica nella direzione qui accennata.

Invece un'altra categoria di benefici potenziali delle relazioni costituisce un classico campo di applicazione per il data warehousing e il data mining. Si tratta delle potenzialità di crescita del valore della relazione. Una classificazione possibile prevede quattro *opzioni di ampliamento della relazione*:

1. *up selling*: aumento di volume degli acquisti, derivante da economie di scala del cliente nell'acquisto e nel consumo;
2. *cross selling*: aumento della varietà di prodotti acquistati, derivante da economie di scopo del cliente;
3. *trading up*: acquisto di prodotti di maggiore qualità e complessità e a più alto margine per l'impresa, derivante da economie di esperienza e apprendimento del cliente;
4. *business extension*: ampliamento del portafoglio di offerta dell'impresa, derivante dall'aumento del portafoglio di business presidiati, derivante a sua volta dalla crescita del livello di fiducia e dalle economie di transazione che si creano nelle relazioni.

Si noti che la business extension differisce dalle altre opzioni di ampliamento in quanto dipende dall'intero portafoglio clienti dell'impresa, mentre le altre opzioni possono valere anche per il singolo cliente.

È noto che con tecniche statistiche e di data mining si può indagare sui presupposti che consentono le opzioni di up selling, cross selling e trading up. Se c'è una tendenza ricorrente di certe tipologie di clienti a scegliere queste opzioni, allora c'è quantomeno la speranza di individuare tali tipologie e quindi tenerne conto nella valutazione dei clienti. Se si riesce a stimare che un certo cliente avrà una probabilità $p_o(t)$ di esercitare una

certa opzione di ampliamento della relazione nel periodo t e si riesce a determinare il valore $\Delta V_o(t)$ che consegue dall'esercizio dell'opzione, allora nella customer equity del cliente dovrà essere incluso anche l'addendo $p_o(t)\Delta V(t)(1+d)^{-t}$.

Le considerazioni appena svolte si riconnettono ai concetti tipici delle *opzioni reali* [Amram 2000, Marzo 2003], una metodologia nata per valutare le decisioni di investimento, e di conseguenza il valore che si può razionalmente attribuire a un progetto quale una azione di marketing. Un'opzione reale è un'opportunità discrezionale di investimento e si sostanzia nella possibilità, ma non nell'obbligo, di effettuare un investimento o di modificarne le caratteristiche in corso di esecuzione.

Le più tradizionali tecniche di capital budgeting, quali ad esempio quelle fondate sul valore attuale netto dei flussi di cassa futuri scontati (tecnica qui adottata nei modelli descritti), evidenziano in sostanza un'unica fonte del valore di un progetto di investimento: i flussi di cassa *direttamente* derivanti dal progetto. Viene trascurato il valore che deriva dalle *opportunità* che il progetto offre all'impresa. Per esempio, la costituzione di una filiale in un mercato estero per seguire gli attuali clienti non solo genera flussi di cassa, ma permette anche di ottenere una maggiore conoscenza delle caratteristiche di quel mercato. Tali conoscenze, infatti, potranno consigliare nuove e specifiche strategie per acquisire una maggiore quota del valore economico di quel mercato, quota che può essere ben maggiore di quella riveniente dai clienti attuali. Strategie del genere sono state ad esempio perseguite da alcune società di consulenza che hanno stabilito i loro uffici nei paesi dell'Europa dell'Est subito dopo la caduta del muro di Berlino. I profitti attesi sarebbero difficilmente derivati dalle condizioni in cui quei mercati si trovavano, ma le opportunità insite nell'essere un first mover erano ritenute molto attraenti. In base a questa logica, il valore di una campagna di acquisizione di clienti non sta solo nei flussi di cassa ma anche nel fatto che il portafoglio di clienti acquisiti potrebbe in futuro aumentare il suo valore grazie ad altre azioni di marketing, distinte dall'attuale campagna ma proprio da questa rese possibili. Per fare un esempio: un normale rapporto di servizio di una banca (contro corrente, bancomat) con un giovane specializzando in chirurgia plastica non può essere valutato equivalente a un rapporto, pur identico sotto tutti gli aspetti operativi, con un altrettanto giovane impiegato. Esistono offerte che la banca potrà proporre in futuro all'uno ma non all'altro cliente. La possibilità, non la

certezza, che tali offerte siano recepite dal primo cliente assicura a quel rapporto un valore potenziale maggiore, di cui la banca deve tenere conto fin da oggi.

Questo esempio è del tutto intuitivo, fino alla banalità, ma per quanto sorprendente dal punto di vista di un decisore razionale, la logica ad esso sottesa spesso non è recepita ed applicata dai decisori aziendali. Ciò avviene in parte per motivi psicologici (eccesso di avversione al rischio), ma in parte per motivi tecnici e metodologici: la valutazione delle opzioni reali è un compito decisamente complesso, e non solo per la complessità matematica insita nella metodologia, ma anche per la difficoltà di raccogliere i dati e stimare le probabilità degli eventi: come stimare i futuri redditi del chirurgo plastico, la probabilità che resti nostro cliente una volta diventato ricco e così via?

La logica delle opzioni reali include una visione per *stati del mondo*. Supponiamo che al tempo $t = 0$ siamo in uno stato del mondo S che può evolvere al tempo $t = 1$ in due modi: con probabilità p_u può evolvere nello stato S_u che porta a un valore più alto per l'impresa V_u e con probabilità $p_d = 1 - p_u$ può evolvere nello stato S_d che comporta un valore più basso V_d (gli indici u e d per tradizione rappresentano gli stati *up* e *down*). Una razionale stima del valore atteso al tempo $t = 1$ è dunque la media pesata dei due valori possibili $V_1 = p_u V_u + p_d V_d$. Se ora estendiamo l'analisi fino al tempo $t = 2$ e supponiamo per semplicità che le probabilità restino p_u e p_d , allora i possibili stati del mondo sono quattro, perché lo stato S_u può evolvere nel caso migliore verso lo stato S_{uu} e nel caso peggiore verso lo stato S_{ud} , mentre lo stato S_d può evolvere verso S_{du} oppure verso S_{dd} . Il valore atteso alla fine del secondo periodo sarà allora

$$V_2 = p_{uu} V_{uu} + p_{ud} V_{ud} + p_{du} V_{du} + p_{dd} V_{dd}$$

Si può estendere l'analisi nei periodi successivi, avendo cura di attualizzare con un opportuno tasso i valori futuri per renderli comparabili a quelli presenti.

Con questo metodo si crea un *albero di decisione*, nel quale ogni nodo è uno stato del mondo dal quale si diramano due archi etichettati con probabilità che portano a due stati del mondo nel periodo successivo. Naturalmente gli archi che escono da ogni nodo possono essere più di due e non è necessario che le probabilità restino costanti. La logica è la stessa: ad ogni stato del mondo sono associate la probabilità di raggiungerlo e il valore stimato per quello stato.

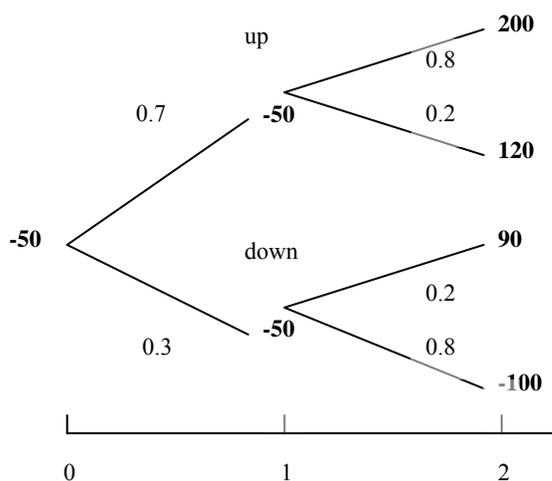
Vediamo un semplice esempio di valutazione di un investimento con la *decision tree*

analysis. Una impresa di moda deve decidere se introdurre una nuova linea di abbigliamento. Il progetto richiede due anni per la completa attuazione, con un investimento iniziale $I_0 = 50$ e un investimento $I_1 = 50$ dopo un anno. Il tasso di attualizzazione è fissato a $d = 0.1$, perché per esperienza il 10% è un valore normale di ritorno su questo tipo di investimenti. Il valore attualizzato di I_1 è

$$A(I_1) = \frac{I_1}{1+d} = \frac{50}{1.1} = 45.5$$

Introduciamo l'incertezza, al fine di creare vari scenari, cioè vari stati del mondo.

L'impresa non può essere certa del successo della nuova linea di prodotto: dipende dall'andamento della moda. L'opinione del management (stiamo parlando ovviamente di stime soggettive di persone competenti, di *educated guess*) è che ci sia una probabilità del 70% che nel primo anno il mercato della moda vada nella direzione "buona", quella *up*, cioè verso il gusto cui si ispira la nuova linea. Nel secondo anno si stima una probabilità dell'80% che il mercato prosegua nella tendenza del primo anno, sia essa *up* o *down*. Il ritorno dalle vendite dopo due anni varia nei quattro scenari possibili: si va da ricavi di 200 nell'ipotesi più favorevole (il mercato mantiene per due anni la tendenza sperata) alla perdita di 100 nel caso più sfavorevole, come illustrato nell'albero di decisione in figura



Ciascuno dei 7 possibili flussi deve essere pesato con la probabilità che si verifichi lo stato del mondo corrispondente e attualizzato in base al periodo in cui si realizza, portano a un valore atteso

$$-50 + \frac{-50}{1.1} + \frac{0.7 \cdot (0.8 \cdot 200 + 0.2 \cdot 120) + 0.3 \cdot (0.2 \cdot 90 - 0.8 \cdot 100)}{1.1^2} = -4.38$$

In base alle ipotesi del management, il valore atteso è negativo, quindi si dovrebbe decidere di non attuare il progetto.

Questo procedimento è più articolato e significativo del tradizionale metodo dei flussi di cassa scontati, con il quale si sceglie un unico scenario e una successione temporale di flussi di cassa. Possiamo dire che nel metodo tradizionale si lavora con uno *scenario aggregato*, mentre qui si ha un albero di *scenari individuali*: si ripresenta la questione dell'aggregazione, tema che emerge continuamente quando si riflette sull'utilità del data warehousing e del data mining. Svolgeremo altrove delle considerazioni in merito. Qui ci interessa invece puntare l'attenzione su un limite del metodo seguito: esso non considera alcuna decisione dopo quella iniziale di attivare l'investimento. Gli scenari alternativi sono causati da fattori esogeni, le tendenze della moda, ma non da decisioni dell'impresa. Per superare questo limite si introduce il concetto di opzione reale, che illustriamo proseguendo l'esempio.

Supponiamo che l'impresa abbia la possibilità di scegliere se impiegare o meno la seconda tranche dell'investimento dopo un anno. Supponiamo cioè che essa abbia la cosiddetta *opzione di abbandono* dell'investimento. Allora un decisore razionale che dopo un anno constata di essere nello stato *down* deve abbandonare il progetto. Infatti il valore atteso del progetto, condizionato all'evento "transizione down fra $t = 0$ e $t = 1$ " è

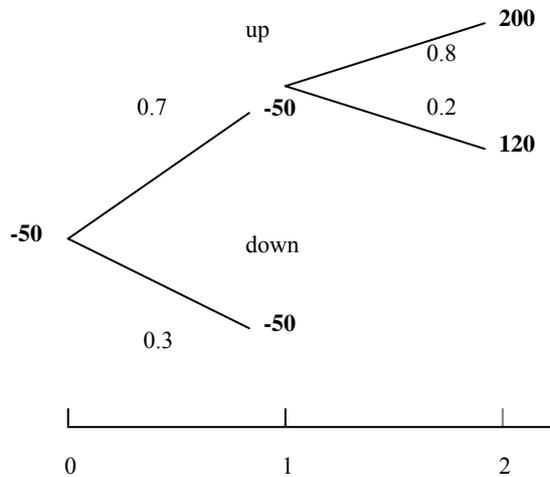
$$\frac{0.2 \cdot 90 - 0.8 \cdot 100}{1.1} = -56.36$$

e il progetto deve essere abbandonato. Si noti che ciò non dipende necessariamente dalla presenza di prospettive di perdita, come il -100 dello scenario peggiore: anche se il valore fosse stato positivo ma minore di 50 il decisore avrebbe dovuto comunque decidere per l'abbandono, perché proseguire ha comunque un costo di 50. Perciò se anche lo stato *down-down* al tempo $t = 2$ avesse un valore di +30 anziché di -100, il valore atteso dalle vendite sarebbe

$$\frac{0.2 \cdot 90 + 0.8 \cdot 30}{1.1} = -38.18$$

e il valore atteso dal progetto sarebbe $-50 + 38.18 = -11.82$, quindi negativo.

Poiché per ipotesi il decisore dispone dell'opzione di abbandono, il “vero” albero che ha a disposizione non è quello prima visto ma questo



dove viene eliminata la biforcazione a valore negativo.

Con il nuovo albero di decisione si può calcolare il valore atteso del progetto in presenza dell'opzione di abbandono:

$$-50 + \frac{0.7 \cdot (-50)}{1.1} + \frac{0.7 \cdot (0.8 \cdot 200 + 0.2 \cdot 120) + 0.3 \cdot (0)}{1.1^2} = +20$$

Pertanto il progetto deve essere intrapreso.

La presenza dell'opzione di abbandono ha fatto passare il valore del progetto da -4.38 a +20, quindi possiamo dire che l'opzione ha essa stessa un valore e questo valore si può quantificare in 24.38.

Questa ultima affermazione in realtà deve essere precisata. L'opzione di abbandono comporta un beneficio di 24.38; ma comporta o no dei sacrifici? La risposta dipende dal contesto. Ci possono essere dei sacrifici di natura commerciale e contrattuale. Supponiamo che l'investimento di 100 consista di acquisiti di macchinari e materiali da uno o pochi fornitori: se l'impresa si riserva di fermarsi a metà acquisto, allora è verosimile che i fornitori pratichino dei prezzi superiori e che quindi l'investimento diventi maggiore di 100. Oppure può darsi che la rinuncia alla seconda parte dell'investimento costi all'impresa il pagamento di una penale. Perciò è verosimile che l'opzione abbia un costo: può essere un costo da sostenere per *creare* l'opzione (caso dell'aumento dei prezzi di acquisto) oppure un costo per *esercitare* l'opzione (caso della penale).

Che l'opzione abbia un costo di esercizio risulta ancora più naturale se si pensa che in

un certo senso essa è simile a una polizza assicurativa che protegge contro un evento rischioso: l'evento di movimento del mercato della moda verso la direzione non desiderata. È naturale aspettarsi che il premio della polizza debba pur essere pagato a qualcuno (negli esempi, ai fornitori di macchine e materiali). Del resto, la teoria delle opzioni reali nasce dalla matematica finanziaria degli investimenti in titoli: le *option* sono facoltà di comprare o non comprare, vendere o non vendere dei titoli, e tali facoltà possono essere usate proprio per proteggersi da rischi eccessivi, pagando in cambio una somma certa ma limitata.

Abbiamo parlato di costi contrattuali dell'opzione di abbandono. Nel contesto del nostro studio ci interessa però ancora di più un altro genere di costo dell'opzione, il suo *costo conoscitivo*. Perché l'impresa possa decidere inizialmente per la convenienza di creare l'opzione di abbandono, ed eventualmente pagare il costo annesso, deve poter stimare i parametri dell'albero di decisione, e questo può avere un costo non trascurabile (p.e. uno studio del mercato). Ancora, perché l'impresa possa decidere per la convenienza di esercitare l'opzione di abbandono occorre che dopo un anno possa sapere che il mercato nel frattempo si è mosso nella direzione *down*, e di nuovo questo può comportare un costo.

Porsi la questione dei costi conoscitivi dell'opzione significa porsi la questione del valore della conoscenza: siamo disposti a pagare 10 per avere le conoscenze necessarie perché queste ci danno una opzione che vale 24.38 e quindi resta un utile di 14.38. Anche questo è un argomento su cui torneremo.

Abbiamo parlato dell'opzione di abbandono; esistono vari altri tipi di opzione, relative all'espansione o contrazione, al differimento dell'inizio o alla temporanea sospensione, al cambiamento dei mix di input o di output eccetera. L'idea di base è quella già vista: un investimento dà origine sia a conseguenze non controllabili, con annessi rischi, sia a conseguenze che possono essere scelte, e che quindi possono essere viste come opportunità che in quanto tali hanno un valore, perché si è disposti a pagare per averle.

Questa visione delle cose può essere applicata anche al capitale relazionale: la relazione con un cliente può portare valore sia direttamente sia indirettamente con la creazione di opportunità: abbiamo detto delle opportunità di up selling, cross selling, trading up e

business extension. Questa considerazione suggerisce un tema di ricerca interdisciplinare molto interessante: posto che le tecnologie di business intelligence sono fertili di applicazione alla comprensione e allo sviluppo delle relazioni con i clienti, possiamo indagare sulla possibilità di usarle per supportare anche l'innovativo approccio basato sulle opzioni reali al capitale relazionale.

Nella successiva trattazione torneremo su questo tema. Possiamo però anticipare che la riflessione sulle opzioni reali, e in generale sul valore potenziale delle relazioni con i clienti, nell'ottica della business intelligence pone dei requisiti informativi interessanti. Il data warehouse dovrebbe prevedere le informazioni necessarie a identificare e valutare scenari e opzioni, e il data mining dovrebbe comprendere algoritmi capaci di "ragionare" in questi termini. Ad esempio: che cosa vuol dire rappresentare nelle tabelle di un data warehouse gli stati del mondo dei decision tree e delle opzioni Oppure, quali informazioni servono al data mining per valutare le probabilità di transizione fra stati del mondo?

Un principio di particolare interesse è questo: il sistema di business intelligence di una impresa proattiva dovrebbe comprendere le decisioni manageriali come entità di prima categoria, con la stessa importanza e dignità delle entità clienti o delle entità risorse. L'impresa proattiva per sua natura deve prendere molte decisioni in condizioni di incertezza e in tempi rapidi; le decisioni dell'impresa proattiva influiscono molto sugli andamenti dell'impresa, più che per una impresa soltanto reattiva; cosa forse persino più importante del punto di vista modellistico e informatico, l'impresa proattiva ha a sua disposizione uno *spazio delle decisioni* particolarmente grande e articolato, nel quale deve risolvere problemi di ottimizzazione di estrema complessità e aleatorietà.

In questo contesto, le tecnologie di business intelligence possono aiutare a rendere più trattabili i problemi conoscitivi e decisionali dell'impresa proattiva, e al tempo stesso possono porre vecchi problemi in un'ottica nuova, suggerendo nuove strade che potrebbero portare a nuove prospettive. Nel prosieguo di questo studio si cercherà di delineare alcune direzioni in cui la ricerca interdisciplinare sulla business intelligence può muoversi.

7. Customer equity e decisioni di marketing

Esamineremo ora un metodo di valutazione del customer equity, ispirato a una ricerca reale [Rust 2001], che esemplifica l'interesse applicativo del concetto e suggerisce utili spunti di riflessione per il progettista di un sistema di business intelligence.

Nel modello vediamo il customer equity come funzione di tre *driver*:

1. *value equity*: il valore che il cliente assegna all'offerta dell'impresa in base a considerazioni oggettive e razionali, riassunte nella percezione di quanto riceve e quanto dà in cambio.
2. *brand equity*: il valore che il cliente assegna all'offerta dell'impresa in base a considerazioni soggettive e emotive, riassunte nella attitudine verso il marchio.
3. *relationship equity*: il valore che il cliente assegna all'offerta dell'impresa in base a considerazioni sulla qualità della relazione con l'impresa stessa.

Il valore che il cliente attribuisce alla nostra offerta dipende dunque, in questo modello, dalle qualità oggettive del prodotto o servizio percepite dal cliente, dalle qualità che il cliente soggettivamente attribuisce al marchio e dalle qualità della relazione intrattenuta con l'impresa.

Analizzando più in dettaglio, indichiamo dei *subdriver*:

1. il driver value equity è determinato dai subdriver *quality*, *price* e *convenience* (intesa come sacrifici diversi del prezzo, p.e. il tempo speso per l'acquisto o per imparare a usare il prodotto);
2. il driver brand equity è determinato dai subdriver *brand awareness*, *brand attitude* e *corporate ethics* (si tratta chiaramente di caratteristiche soggettivamente sentite, anziché oggettivamente percepite, come le precedenti);
3. il driver relationship equity è determinato dai subdriver *loyalty programs*, *special recognition and treatment*, *affinity*, *community building*, *knowledge building* (sono qualità proprie della relazione e anche leve di marketing che l'impresa può usare per fidelizzare il cliente, farlo sentire riconosciuto come individuo e trattato in modo particolare, dargli un senso di affinità con la marca e con gli altri clienti dell'impresa, costruire una comunità di clienti, creare, diffondere e condividere co-

noscenza sull'impresa e sui suoi prodotti e servizi).

Queste considerazioni si riassumono in un modello formale del Customer Equity:

$$CE = CE(VE, BE, RE)$$

$$VE = VE(Qua, Pri, Con)$$

$$BE = BE(Awe, Att, Eth)$$

$$RE = RE(Loy, Spe, Aff, Com, Kno)$$

Per quanto generico, il modello comincia ad assumere una forma che suggerisce le informazioni utili a un sistema di business intelligence. Abbiamo 11 subdriver per i quali si possono immaginare delle osservazioni empiriche. Per lo più tali informazioni sono però non direttamente osservabili; perciò esse saranno desumibili dai dati osservabili (come il comportamento del cliente) da una o entrambe queste fonti:

- inferenze dai comportamenti osservabili interpretati in qualche modello teorico;
- inferenze da questionari somministrati a clienti o osservatori.

Per introdurre nel modello del customer equity i concetti di fedeltà, mantenimento, abbandono, rientro, numero atteso di clienti, valore atteso degli acquisti, ricorriamo a uno strumento matematico che generalizza quelli visti, la *switching matrix*. Questa è una matrice bidimensionale, il cui elemento di indici i e j rappresenta la probabilità che un cliente che all'ultimo acquisto ha scelto la marca i scelga adesso la marca j .

Un esempio di switching matrix fra due marche è

$$S = \begin{bmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{bmatrix}$$

La probabilità che il cliente dopo l'acquisto della marca A compri ancora A è 0.7, mentre la probabilità che passi all'acquisto successivo alla marca B è 0.3. Se invece il cliente ha acquistato B l'ultima volta, il prossimo acquisto ha la stessa probabilità di essere

della marca A o della marca B . Essendo una matrice markoviana, la switching matrix ha una proprietà importante: le probabilità dello i -esimo prossimo acquisto sono rappresentate dalla i -esima potenza della matrice. La seconda potenza di S è

$$S^2 = \begin{bmatrix} 0.64 & 0.36 \\ 0.60 & 0.40 \end{bmatrix}$$

La probabilità che il cliente, dopo avere scelto la marca A , al secondo prossimo acquisto scelga ancora A è 0.64.

È chiaro che disporre della switching matrix significa poter calcolare il numero di clienti al tempo t in modo più raffinato che non usando semplici tasso di mantenimento e abbandono; in più, si può modellare il rientro dopo un temporaneo abbandono. L'uso della matrice anziché dei tassi o probabilità di migrazione ha anche il vantaggio di specificare le probabilità del prossimo acquisto senza specificare quando avverrà questo acquisto. Perciò non costringe a usare una sola unità di tempo in tutto il modello e per ogni cliente; è possibile seguire l'evoluzione nel tempo degli acquisti di un cliente senza imporre una frequenza predeterminata. Un accorgimento utile è l'inserimento di una marca fittizia, che rappresenta il non acquisto. Questa permette anche di rappresentare la mancanza di acquisto in un certo periodo di tempo, se si preferisce impostare l'analisi in questo modo. Infine, la switching matrix permette di modellare simultaneamente varie marche, dell'impresa ma anche altre, e quindi permette di disaggregare in modo elegante le previsioni di acquisto. Disponendo della switching matrix di un cliente si possono stimare gli acquisiti futuri di quel cliente per ogni marca; assegnando ad ogni acquisto una stima del valore e attualizzandolo in base al tempo di acquisto si arriva dunque a valutare il customer equity di quel cliente per ogni marca.

Non entriamo nei dettagli di calcolo. Piuttosto, ci chiediamo come si possono stimare i valori della switching matrix di un certo cliente. Ipotizziamo che la probabilità che un cliente acquisti la marca B (sta per l'inglese *brand*) è funzione dell'utilità che la marca B ha per lui (nello stato in cui si trova in quel momento). Definiamo quindi il concetto di utilità: indichiamo con U_{ijk} l'utilità del brand k per il cliente i che nell'ultima occasione ha acquistato il brand j . Usiamo questo modello:

$$U = I + VE + BE + RE$$

dove U è l'utilità che il cliente trova in un certo brand e I è una misura dell'inerzia che porta un cliente a ripetere gli acquisti di una marca già acquistata.

Modelliamo l'utilità con un modello di regressione logistica, di uso comune quando la funzione da predire è una funzione di probabilità.

Ricordiamo che il modello logistico ha questa forma:

$$\text{logit}(\pi(X_1, \dots, X_n)) = b_0 + b_1 X_1 + \dots + b_n X_n$$

con

$$\pi(X_1, \dots, X_n) = \text{prob}(Y = 1 \mid X_1, \dots, X_n)$$

e

$$\text{logit}(p) = \frac{p}{1-p}$$

Il logit è una trasformata della probabilità con alcune proprietà utili. Dato il logit si può trovare la probabilità, cioè il logit è una trasformazione invertibile. Quindi usando un modello di regressione logistica si può stimare la probabilità che Y abbia il valore 1 o 0 (che accada o non accada) in base a certe variabili indipendenti, i suoi predittori.

Per esemplificare, supponiamo di voler stimare la probabilità che un cliente risponda a una comunicazione di direct marketing. Un modello logistico può essere di questo tipo:

$$L = -10.83 + 0.28 \times \text{Età} + 2.30 \times \text{Genere}$$

dove *Genere* è una variabile con valori 0 o 1 (maschio o femmina) e L è il logit.

Se il cliente è un maschio di 40 anni, si ha

$$L(40, 0) = -10.83 + 0.28 \times 40 + 2.30 \times 0 = 0.37$$

e analogamente per una femmina di 40 anni

$$L(40, 1) = 2.67$$

Dato il logit, si ricava la probabilità di una risposta, che è 0.59 nel primo caso e 0.93 nel

secondo; infatti $\log\left(\frac{0.59}{0.41}\right) = 0.37$.

Tornando alla variabile utilità, la vogliamo descrivere con un modello logistico in questo modo:

$$U_{ijk} = B_{0k} LAST_{ijk} + \overrightarrow{B_{1k}} \overrightarrow{VE}_{ik} + \overrightarrow{B_{2k}} \overrightarrow{BE}_{ik} + \overrightarrow{B_{3k}} \overrightarrow{RE}_{ik} LAST_{ijk} + \varepsilon_i$$

Nel modello ε_i è un errore, una variabile casuale con distribuzione propria del cliente i .

$LAST_{ijk}$ è una variabile fittizia (*dummy variable*) che vale 1 se $j = k$ e 0 altrimenti. Questa variabile fittizia viene introdotta per tenere conto dell'inerzia che porta il cliente a

ripetere l'ultima scelta di acquisto e del Relationship Equity solo quando il cliente mantiene la scelta precedente (compra la stessa marca), mentre le due grandezze spariscono quando il cliente cambia marca. B_{0k} è il coefficiente del fattore inerzia e dipende dal brand, ma non dal cliente: è *market-level*, non *individual-level*. B_{1k} , B_{2k} e B_{3k} sono tre vettori colonna i cui elementi sono i pesi del Value Equity, del Brand Equity e del Relationship Equity. Anche questi sono *market-level*, indipendenti dagli individui. Invece sono *individual-level* i valori di VE , BE e RE .

Costruiamo ora una matrice delle utilità di transizione tra brand

$$U = \begin{bmatrix} I + RE_1 + VE_1 + BE_1 & VE_2 + BE_2 & VE_3 + BE_3 \\ VE_1 + BE_1 & I + RE_2 + VE_2 + BE_2 & VE_3 + BE_3 \\ VE_1 + BE_1 & VE_2 + BE_2 & I + RE_3 + VE_3 + BE_3 \end{bmatrix}$$

La probabilità che il cliente i passi dal brand j al brand k^* è

$$S_{jk}^i = \frac{\exp(U_{ijk}^*)}{\sum_k \exp(U_{ijk})}$$

dove l'esponenziale è introdotto per invertire la funzione logit, che è logaritmica.

A questo punto la switching matrix è espressa in funzione dei driver del Customer Equity, come desiderato.

Finalmente possiamo stimare il customer equity per il cliente i e la marca k :

$$CE_{ik} = \sum_{t=0}^{T_i} (1 + d_k)^{-t} v_{it} m_{ikt} B_{ikt}$$

dove T_i è il numero di acquisti fino all'orizzonte temporale assunto per l'analisi, v_{it} è il volume previsto nella transazione di acquisto t , m_{ikt} è il margine unitario previsto per l'acquisto t della marca k da parte del i , B_{ikt} è la probabilità che il cliente i compri la marca k nell'acquisto t , probabilità ottenuta dalla potenza t -esima della switching matrix.

Si pone ora la questione del reperimento dei dati empirici da inserire nel sistema di business intelligence e da elaborare per stimare le grandezze VE , BE , RE in funzione degli 11 subdriver prima introdotti. Per raccogliere i dati dovremo somministrare degli opportuni questionari ai clienti e procedere con le necessarie elaborazioni statistiche. Noi concludiamo qui concludiamo l'esposizione degli aspetti matematici del modello, che ai

fini del presente lavoro sono meno importanti delle considerazioni metodologiche sui contenuti e la strutturazione del datawarehouse e sulle tipologie di algoritmi di data mining che andranno a comporre il sistema di business intelligence.

Qui evidenziamo che disponendo dei dati empirici per applicare il modello, si ha a disposizione un quadro di supporto decisionale improntato a criteri microeconomici ben fondati. Supponiamo di saper trattare in modo analitico o tabellare le funzioni delle tre equazioni

$$VE = VE(Qua, Pri, Con)$$

$$BE = BE(Awe, Att, Eth)$$

$$RE = RE(Loy, Spe, Aff, Com, Kno)$$

Se l'impresa effettua un investimento finalizzato a migliorare un subdriver, poniamo la qualità, allora in base alla prima equazione si potrà stimare l'impatto dell'investimento sul value equity VE , Questo influisce sulle probabilità di scelta di ciascun cliente, la sua switching matrix, e questo a sua volta influisce sul customer equity di ciascun cliente, e in definitiva sul customer equity totale dell'impresa. Abbiamo allora uno strumento per prevedere l'impatto dell'investimento nella qualità di un prodotto sull'intero capitale relazionale dell'impresa. L'impresa ha quindi a disposizione uno strumento di business intelligence che supporta le decisioni in base a informazioni (data warehousing) e modelli matematici (data mining). Avendo descritto il modello, lasciamo aperta la questione del reperimento e della rappresentazione delle informazioni, problema su cui torneremo in seguito.

Capitolo 3

Segmentazione e ottimizzazione del valore

1. Sul concetto di segmentazione

Il concetto di segmentazione riveste una importanza cruciale nella teoria e nella pratica del marketing. Rimandando alla vastissima letteratura sull'argomento, vogliamo qui indagare sull'impatto che le tecnologie per la business intelligence possono avere sul concetto stesso di segmentazione, prima ancora che sulla sua elaborazione metodologica e sulla sua applicazione pratica.

Intuitivamente, la segmentazione di un insieme di clienti (attuali o potenziali) consiste nella suddivisione dell'insieme in sottoinsiemi (segmenti) spesso ma non necessariamente disgiunti che presentino caratteri di omogeneità al loro interno e disomogeneità gli uni rispetto agli altri. Quindi i clienti di uno stesso segmento sono “simili” mentre quelli di due segmenti diversi sono “dissimili”. Naturalmente occorre dire in quale senso due clienti possono essere giudicati simili o dissimili, e come si possano stabilire graduatorie per dire che il cliente A è più simile al cliente B che non al cliente C. In termini di data mining è chiara la parentela di questo concetto con quello di clustering di un dataset; vogliamo qui approfondire il senso del concetto di segmentazione in ottica aziendale, ma cercando di indicarne i collegamenti con concetti propri delle tecnologie informatiche e in particolare con le tecnologie per la business intelligence.

Premettiamo che in questa sede supponiamo di avere un dataset che descrive clienti: il dataset sarà in sostanza una tabella di dati nella quale a ogni riga corrisponde un cliente e a ogni colonna un attributo dei clienti. Fare segmentazione significherà dividere l'insieme delle righe in sottoinsiemi disgiunti in base a un qualche criterio di similarità, per il momento non definito. Questa visione dei dati è semplificata ma realistica in un contesto di data mining, per il quale i dati provenienti da un database più complesso vengono preventivamente rielaborati e collocati in una tabella di questo genere.

Una volta che sono stati individuati i segmenti, l'impresa elabora una strategia per "agire" su tutti o su alcuni di questi segmenti; per ogni segmento si decideranno azioni di tipo diverso, calibrate sulle caratteristiche dei clienti di quel segmento. Tenendo presente questo concetto fondamentale, ci dobbiamo chiedere come dovrebbero essere fatti dei "buoni" segmenti di clientela.

Si usa dire che una buona segmentazione dovrebbe rispettare una serie di requisiti requisiti:

- risposta differenziata
- identificabilità
- misurabilità
- rilevanza
- stabilità
- accessibilità

I segmenti devono quindi presentare una risposta differenziata: due clienti di due segmenti diversi rispondono diversamente alle azioni di marketing dell'impresa, mentre due clienti di uno stesso segmento rispondono in modi simili. Questo criterio indica perciò un criterio di marketing per definire il concetto di similarità fra clienti, concetto che in qualche modo si potrà ritrovare in una elaborazione di data mining (con algoritmi di clustering o di genere diverso) per effettuare una segmentazione sul dataset dei clienti. Il criterio di risposta differenziata dice che ciascun segmento deve avere una sua "individualità" che lo distingue dagli altri.

Il criterio di identificabilità richiede che sia possibile assegnare ciascun cliente al suo segmento di appartenenza (oppure ai suoi segmenti se si è scelto un metodo di segmentazione che ammette la sovrapposizione fra segmenti).

Il criterio di rilevanza richiede che le dimensioni e il potere di acquisto dei clienti del segmento giustifichino economicamente l'adozione di una politica di marketing specifica per quel segmento. Il segmento deve quindi "valere la pena" di essere trattato in modo differenziato.

Il criterio di misurabilità richiede che le caratteristiche che distinguono il segmento e i suoi potenziali di acquisto siano quantificabili in modo affidabile ed agevole. Per esempio, se i clienti sono distinti secondo la loro ricettività all'innovazione (criterio interessante per chi opera in mercati di prodotti ad alta tecnologia) probabilmente sarà necessario condurre una costosa ricerca di mercato ad hoc per prendere decisioni sulle campagne di marketing da intraprendere per aggredire in modi differenziati i vari segmenti.

Il criterio di accessibilità richiede che sia in effetti possibile agire in modo mirato sui clienti di un segmento e raggiungerli con le azioni di marketing a loro specificamente dedicate.

A nostro parere, queste quattro condizioni, tutte interessanti in teoria e tutte importanti nella pratica, hanno uno status concettuale non omogeneo. Il criterio di risposta differenziata è più primitivo degli altri, nel senso che li precede e in un certo modo li giustifica: l'essenza della segmentazione sta proprio in questo criterio. Prima di tutto la segmentazione dovrebbe rendere conto della elasticità di risposta dei clienti rispetto alle azioni dell'impresa. Gli altri criteri sono più pertinenti all'utilità di una certa scelta dei segmenti rispetto ad un'altra, mentre questo coglie la stessa ragione d'essere della segmentazione. Peraltro, quando la segmentazione sia condotta su un *customer database* di dimensioni e qualità adeguate i criteri di rilevanza, misurabilità e accessibilità sono soddisfatti spesso in modo naturale; invece il criterio di risposta differenziata richiede elaborazioni anche molto complesse sui dati disponibili.

Nel presente lavoro si propone un concetto di segmentazione che non coincide con quelli generalmente in uso nella letteratura di marketing, ma che non di meno è con essi compatibile. La nostra definizione di segmentazione tenta di formalizzare matematicamente alcuni aspetti del concetto di marketing allo scopo di poter applicare meglio alla segmentazione le nuove tecnologie per la business intelligence. Il prezzo da pagare è la perdita di aspetti del concetto usuale di segmentazione che sono certamente interessanti e fruttuosi. Riteniamo che questa operazione di astrazione (e quindi simultaneamente di formalizzazione e di limitazione) possa essere molto utile per i progettisti e gli analisti dei dati aziendali, purché questi abbiano l'accortezza di non utilizzarla in opposizione o sostituzione ai metodi tradizionali del marketing, la cui bontà è comprovata da decenni di applicazione. Piuttosto, il concetto di segmentazione che andiamo a delineare deve

essere considerato come uno strumento conoscitivo e decisionale complementare a quelli tradizionali.

In sintesi, vediamo qui la segmentazione come il processo euristico di soluzione di un problema matematico di ottimizzazione: in linea di principio vorremmo scegliere per ogni cliente l'azione di marketing ottimale per lui; poiché questo in pratica è possibile soltanto in alcuni casi, in generale ricorriamo alla segmentazione per approssimare la soluzione ottimale del problema con una soluzione soddisfacente e praticabile. In sostanza, segmentare significa ridurre lo spazio di ricerca.

La segmentazione ha quindi un aspetto computazionale che è utile tenere presente, unitamente a quelli ben noti dalla letteratura di marketing.

Esponiamo ora in modo semi-formale questo concetto di segmentazione, ispirato all'articolo seminale di [Kleinberg 1998].

Chiamiamo D l'insieme delle decisioni di marketing che l'impresa può prendere. Non chiariamo che cosa siano queste decisioni, ci limitiamo a dire che esiste uno *spazio delle decisioni* nel quale cercare soluzioni ottimali. Indicando con $x \in D$ una generica decisione e con f una funzione di utilità che esprime il valore di una decisione, possiamo pensare che l'impresa debba massimizzare la funzione di utilità f sullo spazio delle decisioni, risolvendo il problema

$$\max_{x \in D} f(x)$$

Nella realtà economica lo spazio delle decisioni D è in buona misura sotto il controllo dell'impresa mentre la funzione f dipende moltissimo da altri attori: clienti, competitori, stakeholder di ogni genere. Qui ci interessano in particolare i clienti.

In pratica, la definizione dello spazio delle decisioni D è un problema a volte semplice a volte complesso, ma comunque ragionevolmente risolvibile. Invece la definizione esplicita e precisa della funzione f può essere di complessità estrema, spesso del tutto impraticabile. Ciò non significa che l'impresa non tenti di risolvere il problema di ottimizzazione, ma significa che lo fa ricorrendo a euristiche più o meno intuitive. Consapevolmente o meno, i decisori dell'impresa affrontano questo problema, formalizzandolo ed esprimendolo matematicamente solo nella misura del possibile.

Chiamiamo C l'insieme dei clienti dell'impresa ed esprimiamo l'utilità $f(x)$ di una decisione come la somma delle utilità apportate da ciascun cliente in seguito a quella decisione, indicando con f_i la singola funzione di utilità

$$f(x) = \sum_{i \in C} f_i(x)$$

Per esemplificare, possiamo pensare che una f_i esprima il margine economico derivante dagli acquisti che il cliente i effettua a seguito della campagna di direct marketing x , una delle campagne possibili.

Ora assumiamo che il contributo di ciascun cliente all'utilità totale sia una funzione dei dati disponibili per quel cliente. In altri termini, assumiamo di avere a disposizione per ciascun cliente dati adeguati a spiegare il suo contributo all'utilità totale in conseguenza di ogni possibile decisione. Con questa assunzione, possiamo riscrivere la funzione di utilità individuale così:

$$f_i(x) = g(x, y_i)$$

dove y_i rappresenta i dati sul cliente i e g è un funzione che ha come parametri la descrizione di una decisione e la descrizione di un cliente. Il problema di ottimizzazione diventa quindi

$$\max_{x \in D} \sum_{i \in C} g(x, y_i)$$

Abbiamo così una formulazione generale di tipici problemi di ottimizzazione studiati dalla ricerca operativa e dalla microeconomia.

Il punto critico della questione è che in generale questo problema non può essere risolto in modo naturale, trovando le funzioni di utilità di tutte le decisioni possibili e scegliendo il massimo, per tre motivi:

- I. non si sa esprimere la funzione g ;
- II. non si hanno sufficienti dati y_i ;
- III. non si ha potenza computazionale sufficiente.

Per questi motivi convenzionalmente i problemi di ottimizzazione di questo tipo sono risolti in modo sub-ottimale ricorrendo a euristiche che restringono in modo drastico lo spazio di ricerca delle soluzioni.

In primo luogo si cerca di esprimere la funzione g in forma analitica semplice, tipicamente in forma lineare, usando una funzione approssimata \hat{g} .

In secondo luogo anziché usare i dati y_i si usano dei dati aggregati \hat{y} che sono già di per sé estesi all'insieme dei clienti (p.e. domanda aggregata di un prodotto); in questo modo non si risolve $\max_{x \in D} \sum_{i \in C} g(x, y_i)$ bensì $\max_{x \in D} \hat{g}(x, \hat{y})$.

Con questi accorgimenti si risolvono i primi due problemi esposti (difficoltà di esprimere la funzione di utilità e mancanza di dati sui clienti) e implicitamente anche il terzo (complessità computazionale, ora drasticamente ridotta). Il prezzo da pagare è evidentemente la perdita di valore della soluzione trovata al confronto di quella ottimale. Questa sub-ottimalità deriva dal grado di non-linearità della funzione $g(x, y_i)$: le euristiche descritte introducono due approssimazioni, una sulla forma della funzione di utilità, una sulla precisione delle informazioni sui clienti.

Per certi aspetti la segmentazione dei clienti è una euristica che consiste nel considerare un piccolo numero ristretto di *stereotipi* di clienti, i segmenti, descritti da dati aggregati, per esempio assegnando a un segmento l'età media dei clienti che gli appartengono. Nella formula, la segmentazione è visibile nella scelta di usare i dati aggregati \hat{y} anziché i dati individuali y_i . Chiaramente questa è una definizione molto restrittiva di segmentazione, che serve solo a scopi di formalizzazione quantitativa in vista della progettazione di sistemi di business intelligence basati su data warehousing e data mining. L'utilità di questa definizione è che indica vie per collegare gli algoritmi matematici usati per segmentare i clienti con le motivazioni economiche aziendali che stanno alla base della segmentazione.

Definiti in questo modo certi problemi decisionali di marketing, possiamo cercare modi per risolvere il problema di ottimizzazione in modi soddisfacenti trovando euristiche capaci di fornire soluzioni ottimali o vicine all'ottimo che non introducano eccessive complessità computazionali, di modellazione o di reperimento dei dati.

Una considerazione generale che possiamo fare è questa: con le moderne tecnologie di business intelligence e con la recente esplosione di disponibilità delle informazioni, oggi è possibile risolvere il problema di ottimizzazione descritto utilizzando grandi quantità di dati espresse a livello di dettaglio molto disaggregato, utilizzando funzioni non lineari

molto complesse e impiegando grandi risorse di potenza computazionale, tutto questo in misura non praticabile soltanto qualche anno fa.

L'esperienza testimonia anzi che i decisori di impresa, quando affrontano problemi di ottimizzazione di questo tipo, spesso si adeguano a soluzioni meno efficaci del possibile a causa dell'ignoranza delle opportunità offerte dalle tecnologie di business intelligence. A volte essi ricorrono a segmentazioni semplicistiche e ad alto grado di aggregazione semplicemente perché, in forza dell'abitudine, non pensano neanche che sia possibile fare di meglio.

2. Segmentazione come euristica di ottimizzazione

Il problema di ottimizzazione dell'utilità precedentemente formulato può essere espresso in modo più rigoroso come segue.

È assegnato un insieme $D \subseteq \mathbf{R}^d$ di possibili *decisioni*, ciascuna consistente nella scelta di d parametri nel campo dei numeri reali; è assegnato anche un insieme di n *clienti*, per ciascuno dei quali è data una funzione f_1, \dots, f_n con $f_i : D \rightarrow \mathbf{R}$.

Un *problema di segmentazione* è un problema di ottimizzazione espresso come segue: date le n funzioni f_i e un intero k , trovare k soluzioni $x_1, \dots, x_k \in D$ che massimizzano la funzione obiettivo

$$\sum_{i=1}^n \max_{1 \leq j \leq k} f_i(x_j)$$

In questo problema ad ogni cliente viene “applicata” una certa decisione e il cliente risponde apportando all'impresa una certa utilità, secondo la sua *funzione di risposta individuale*; l'obiettivo è massimizzare l'utilità totale scegliendo nel modo migliore quali decisioni attivare e per ciascuna a quali clienti applicarla. Non è immediatamente evidente perché questo sia un problema di segmentazione: lo si comprende meglio in una versione alternativa del problema, in cui si chiede di trovare una partizione di $\{1, \dots, n\}$ in k insiemi S_1, \dots, S_k che massimizza la funzione obiettivo

$$\sum_{j=1}^k [\max_{x \in D} \sum_{i \in S_j} f_i(x)]$$

Gli insiemi S_j sono i segmenti di clientela, che qui si suppongono a due a due disgiunti. Le due versioni del problema sono equivalenti, essendo della forma *max-max* dove l'operatore *max* commuta: per questo la prima formulazione del problema è in effetti un problema di segmentazione. L'equivalenza algoritmica delle due formulazioni significa che la complessità computazionale non deve necessariamente essere dell'ordine di $|D|2^{nk}$, come nella soluzione ingenua di elencare tutte le possibili partizioni di n clienti in k insiemi, collegarle in tutti i modi possibili alle decisioni e quindi scegliere il modo migliore. È invece possibile calcolare l'utilità di ciascuno degli n clienti rispetto a una scelta di k decisioni, con complessità $|D|^k n$. La complessità è quindi lineare nel numero dei clienti, che realisticamente è di regola molto più grande del numero di decisioni possibili. Possiamo dunque pensare ai problemi di segmentazione nella prima forma, sapendo che essa è equivalente alla seconda forma, più intuitivamente legata alla segmentazione come normalmente intesa.

Nelle due formulazioni precedenti abbiamo fissato il numero k dei segmenti; possiamo anche lasciare k libero di variare, e quindi parte della soluzione, in questa ulteriore versione del problema: date le n funzioni f_1, \dots, f_n e un intero γ , trovare un intero k e k soluzioni x_1, \dots, x_k che massimizzano

$$\left(\sum_{i=1}^n \max_{1 \leq j \leq k} f_i(x_j) \right) - \gamma k$$

In questa formulazione si può scegliere quanti segmenti definire, vale a dire quante diverse strategie di marketing attivare, accettando un costo per ogni nuova strategia.

L'interesse di queste formulazioni sta nel fatto che la segmentazione è vista come un clustering dello spazio dei clienti e lo stesso processo di clustering è influenzato, non a posteriori ma intrinsecamente, dall'obiettivo di massimizzare una funzione di utilità e di adottare un insieme di decisioni indipendenti per i vari clienti.

La complessità computazionale dei problemi di segmentazione è stata indagata con risultati in linea di principio negativi [Kleinberg 1998, Kleinberg 2004], perché i problemi sono NP-completi. Questo risultato non deve però essere interpretato come impossibilità pratica di affrontare e risolvere i problemi di segmentazione: le tradizionali segmentazioni della clientela sono utili anche se formulate in base a criteri euristici decisamente approssimativi, anche perché nella pratica non del tutto consapevoli. Vale

quindi la pena di utilizzare questi concetti per guidare l'impostazione dei problemi di segmentazione e la ricerca di euristiche per la loro soluzione approssimata.

La segmentazione tradizionale può essere vista come un processo euristico di soluzione del problema di ottimizzazione, nel quale si presuppone *che certi dati disponibili sui clienti siano buoni predittori della funzione di risposta del cliente stesso alle azioni di marketing e che se due clienti sono simili rispetto a questi dati allora sono simili anche rispetto alle loro funzioni di risposta*. Quando si segmentano i clienti in base alla zona di residenza (segmentazione a base geografica) si presuppone che due clienti residenti nello stesso quartiere abbiano significative somiglianze nella funzione di risposta. Quando si segmentano i clienti in base alle caratteristiche socio-demografiche si presuppone che due clienti di simile età, provenienza e stadio nel ciclo di vita siano anche significativamente simili nella loro risposta alle azioni di marketing dell'impresa. In questi assunti, del tutto ragionevoli e collaudati dall'esperienza, sono implicite varie ipotesi. Due le abbiamo enunciate: una, più evidente, è che le variabili poste a base della segmentazione (cioè i dati y_i delle nostre formule) siano *buoni predittori* della funzione di risposta; l'altra, più sottile, è che la funzione di risposta sia più o meno continua e *lineare* rispetto ai predittori. Un'altra ipotesi, correlata alla seconda, è che l'uso di dati aggregati non comporti una eccessiva perdita di informazione sulla funzione di risposta rispetto all'uso di dati disaggregati.

3. Il problema della segmentazione dei cataloghi

Per esemplificare il senso della formalizzazione della segmentazione come problema di ottimizzazione, esaminiamo ora un problema di segmentazione particolarmente interessante, il *problema della segmentazione dei cataloghi*. Supponiamo di dover creare un catalogo di prodotti da proporre ai nostri clienti, per esempio stampandolo e inviandolo per posta. Assumiamo che, per ragioni economiche e di altro genere, possiamo inserire nel catalogo un numero prefissato r di prodotti; ci poniamo l'obiettivo di inserire nel catalogo gli r prodotti per i quali è massimo il numero di clienti interessati a questi prodotti. Formalmente, se n è il numero dei nostri clienti e il cliente i è interessato

all'insieme I_i di nostri prodotti, il catalogo C contenente gli r prodotti p_1, \dots, p_r sarà scelto per massimizzare questa funzione

$$\sum_{i=1}^n |C \cap I_i|$$

Così enunciato il problema ha una soluzione banale: inseriremo nel catalogo gli r prodotti più popolari, cioè che interessano il maggior numero di clienti. Si noti che questo è un caso particolare del classico problema dello zaino, nel quale tutti i pesi degli oggetti da selezionare sono uguali a 1.

Il problema del catalogo diventa ben diverso, e più interessante, se assumiamo di dover creare due cataloghi, ciascuno di r prodotti, e di inviare ad ogni cliente uno fra i due cataloghi. Questo è un problema realistico e di estrema importanza economica: una accurata scelta dei cataloghi e dei loro destinatari può apportare benefici molto grandi. La definizione formale del problema dei due cataloghi diventa la seguente.

Dato un insieme U e n sottoinsiemi S_1, \dots, S_n di U , trovare due sottoinsiemi X e Y di U , ciascuno di cardinalità r , che massimizzano

$$\sum_{i=1}^n \max(|S_i \cap X|, |S_i \cap Y|)$$

L'algoritmo naturale per risolvere questo problema è il seguente. Si elencano tutte le partizioni dei clienti in due sottoinsiemi. Per ciascuno dei due sottoinsiemi si trovano gli r prodotti più popolari fra i clienti di quel sottoinsieme; questi prodotti costituiscono il catalogo ottimale per il sottoinsieme e la coppia di cataloghi così determinata è ottimale per quella partizione. Fra tutte le partizioni e le relative coppie di cataloghi si sceglie la migliore. Tutto ciò è molto semplice ma di complessità esponenziale nel numero dei clienti.

Purtroppo lo studio del problema dei cataloghi ha portato a vari risultati negativi: metodi per garantire soluzioni approssimate accettabili al problema di ottimizzazione sono stati trovati soltanto ponendo condizioni piuttosto forti.

Il problema dei cataloghi, anche nel caso base di due cataloghi, è NP-completo, così come molti altri problemi di segmentazione. Diventa risolvibile in tempo polinomiale a patto di fissare il numero di dimensioni dello spazio di ricerca: in concreto, ciò significa fissare il numero di dimensioni d dello spazio delle decisioni $D \subseteq R^d$. Inoltre, ciò vale solo se le funzioni di utilità f_i sono lineari. Sotto queste ipotesi, si conoscono degli algo-

ritmi approssimati, che però danno garanzie sui risultati soltanto con ulteriori ipotesi restrittive piuttosto forti e intuitivamente non soddisfacenti per realismo.

A nostra conoscenza, questi risultati scientifici non hanno ancora portato ad applicazioni pratiche, almeno non direttamente. Il nostro interesse in questa sede non è di approfondire gli aspetti algoritmici, per quanto certo interessanti [Alon 1999], ma di chiarire il problema per indicare quali possono essere le vie per impostarlo in modo da renderlo realisticamente risolvibile con metodi euristici.

Osserviamo che il problema dei cataloghi può essere formulato con più di due cataloghi in modo naturale: il prezzo computazionale è che le partizioni dei clienti sono dell'ordine di $|C|^n$, dove C è l'insieme dei cataloghi e n il numero di clienti.

Un'altra estensione è di considerare funzioni di utilità non costanti per tutti i prodotti e tutti i clienti. Indichiamo con $f_i(x)$ l'utilità apportata dal cliente i se trova l'articolo x nel suo catalogo (tipicamente, il margine economico derivante dal suo acquisto di quel prodotto in catalogo). Se C è un catalogo e S è un insieme di clienti (cioè un segmento di clientela), l'utilità derivante dalla proposizione del catalogo al segmento di clienti può essere rappresentata così

$$f(C, S) = \sum_{i \in S} \sum_{x \in C} f_i(x)$$

e questa è dunque la funzione da massimizzare con la scelta della partizione dei clienti in segmenti e con l'associazione fra segmenti e cataloghi.

4. Considerazioni metodologiche

Riflettiamo ora informalmente su alcune possibili implicazioni metodologiche di quanto precedentemente enunciato in modo formale.

È noto che il data mining offre due classiche metodologie per affrontare il problema della segmentazione, il clustering e gli alberi di classificazione. Possiamo dire che il clustering raggruppa i clienti in base a relazioni di distanza definite dall'analista (i clienti in un gruppo sono simili fra loro e dissimili rispetto a quelli degli altri gruppi), mentre un albero di classificazione li raggruppa in base a criteri che spiegano al meglio il valore che per ogni cliente ha una certa funzione obiettivo (p.e. crea un gruppo di uomini e uno

di donne perché il genere aiuta efficacemente a spiegare il comportamento di acquisto). Gli alberi di decisione danno un risultato più comprensibile, perché espresso in termini di valori degli attributi dei clienti, più o meno in termini di una query al database dei clienti; il clustering invece fornisce risultati non direttamente esprimibili in questi termini e quindi di meno agevole comprensione. Con entrambi i metodi i clienti sono raggruppati in insiemi omogenei: le classi degli alberi di decisione sono omogenee rispetto alla funzione criterio, i cluster rispetto ad attributi che dovrebbero, secondo l'analista, determinare il comportamento dei clienti.

Oltre a queste due tecniche, si possono usare le molte tecniche di classificazione e quelle di regressione. Con una tecnica di classificazione un cliente viene assegnato ad una classe in base al valore di una funzione discriminante, appresa dai dati stessi; con una tecnica di regressione si predice il valore ignoto di una certa funzione per un cliente per analogia dal valore noto di quella funzione per altri clienti.

La nostra formulazione dei problemi di segmentazione (dei cataloghi o di altro tipo) richiama in modo naturale concetti analoghi a questi. Il punto interessante per noi non è tanto nell'aspetto algoritmico della formulazione, cioè complessità, possibilità di approssimazione e simili. Il punto essenziale è che la suddivisione in gruppi è esplicitamente guidata da una funzione di utilità economica, e questa dipende dalle decisioni di marketing che è possibile prendere.

Il classico approccio alla segmentazione, tipicamente con il clustering, usa le funzioni obiettivo di tipo economico prima della segmentazione, nella scelta delle variabili da considerare, e dopo, nell'interpretazione dei risultati, ma non durante il processo stesso e comunque in modo non esplicito e formale. Esistono metodi di clustering guidati da funzioni obiettivo, ma sono orientati a criteri matematici, per esempio fondati sulla teoria dell'informazione, più che su criteri decisionali, quindi economici e di marketing.

Il punto di vista di questo lavoro, e dei progetti industriali ai quali esso si ispira, è che per mettere veramente all'opera il data mining nei processi decisionali, secondo la logica di chi in azienda prende le decisioni, sarebbe di grande beneficio trovare metodi per incorporare negli algoritmi di data mining i criteri decisionali, affinché l'intero processo di analisi e di estrazione della conoscenza ne benefici. È questo un obiettivo chiaramente ambizioso, al quale il presente lavoro intende contribuire con qualche riflessione metodologica.

Una direttrice di ricerca finalizzata a collegare più strettamente ai criteri decisionali del management di impresa gli algoritmi di data mining usati per la segmentazione consiste nell'usare a fini di segmentazione non soltanto variabili osservabili (attributi del cliente e storia del suo comportamento) ma anche variabili non osservabili che diano ragione della sua propensione all'acquisto, della sua fedeltà e in generale della sua *responsiveness* alle azioni di marketing dell'impresa. Storicamente si sono diffusi dapprima i metodi di segmentazione basati su variabili osservabili e con finalità descrittive; in una seconda fase si sono affermati metodi ancora basati su variabili osservabili, ma con finalità predittive. I metodi predittivi, certamente interessanti, fanno largo uso di data warehousing e data mining, che rendono possibili il *customer tracking* e il *customer profiling*. Tuttavia, anche questi metodi non focalizzano abbastanza nitidamente la questione centrale: la differenziazione della risposta dei clienti rispetto alle azioni di marketing dell'impresa.

Per rendere conto appieno della responsiveness dei clienti appare naturale e inevitabile includere nelle metodologie di segmentazione le determinanti cognitive dei comportamenti, e quindi le loro preferenze e le loro valutazioni, i loro grado di soddisfazione, tutti fattori che determinano la funzione di risposta dei clienti.

In sintesi, diciamo che i *customer database* o meglio i *customer data warehouse* contribuiscono ancora meglio alla efficace segmentazione della clientela se contengono anche variabili non osservabili utili a predire le differenziazioni nella funzione di risposta dei clienti. Ciò implica che il customer data warehouse dovrebbe fare esplicito riferimento a modelli di customer loyalty e customer satisfaction. Questo requisito poteva essere troppo restrittivo per molte imprese fino a qualche anno fa, ma oggi è abbastanza normale che un'impresa rilevi i dati sulla customer satisfaction e quindi è ragionevole richiedere che il data warehouse faccia riferimento a modelli come quelli descritti in precedenza.

5. Euristiche per il problema della segmentazione

Prendiamo in considerazione il problema dei cataloghi, esempio interessante per la sua importanza pratica che solleva questioni valide per la segmentazione in generale. Per esemplificare, limitiamoci a due cataloghi.

Abbiamo detto che il problema è in linea di principio computazionalmente arduo, e quindi richiede di essere affrontato con algoritmi euristici per arrivare a soluzioni approssimate. Ad oggi non sono noti metodi generali che diano soddisfacenti performance garantite neanche per algoritmi approssimati [Kleinberg 2004]. Possiamo dunque pensare a metodi che sfruttino la conoscenza sullo specifico dominio applicativo per ridurre lo spazio delle soluzioni e guidare gli algoritmi di ricerca.

Uno fra i classici requisiti per una buona segmentazione è la *rilevanza*, intesa come consistenza di dimensioni e di potenziali ritorni economici. Un criterio per ridurre lo spazio delle soluzioni per il problema dei cataloghi è di escludere dalla generazione di tale spazio le segmentazioni nelle quali uno dei segmenti è “non rilevante”. La non rilevanza di un segmento può essere dovuta alla troppo piccola cardinalità (troppo pochi clienti) oppure alla troppo piccola utilità attesa per il segmento.

L'accorgimento di non generare segmentazioni con segmenti troppo piccoli non apporta benefici significativi. Per esempio, supponiamo che i cataloghi siano due, e che quindi le partizioni possibili per n clienti siano 2^n , quindi un numero intrattabile, dato che i clienti saranno molte migliaia e potranno essere anche milioni. Imporre che ciascuno dei due segmenti sia composta da almeno m clienti, con un valore di m poniamo pari al 10% o 20% di n , esclude solo una piccola frazione delle segmentazioni e il problema resta in pratica intrattabile. Il punto è che la riduzione dello spazio di ricerca, esprimendosi in termini intuitivi, non deve essere “aritmetica” bensì “logaritmica”.

Vincolare la generazione delle segmentazioni candidate a produrre segmenti significativi per profitto atteso appare a prima vista più promettente. Un esempio di euristica potrebbe essere questo:

1. selezionare un piccolo sottoinsieme dei clienti che sia più promettente per profitto atteso;
2. segmentare in due parti questo sottoinsieme con un algoritmo di enumerazione come quello prima descritto;
3. associare gli altri clienti uno alla volta a uno dei due segmenti determinati al passo precedente, usando un criterio decisionale binario.

Con i primi due passi si cerca di acquisire qualche garanzia che i due segmenti abbiano una certa rilevanza di utilità. Con il terzo passo si dà per acquisita la segmentazione e ci si limita a scegliere per ogni altro cliente il migliore fra i due segmenti ormai acquisiti. In questo modo, se $m \ll n$ è il numero dei clienti più promettenti selezionato al passo 1, lo spazio delle soluzioni al problema del passo due ha dimensione 2^m , mentre il passo 3 ha complessità lineare nel numero $n-m$ dei clienti restanti.

L'accorgimento è interessante, anche se da solo non risolve il problema delle dimensioni eccessive dello spazio di ricerca: se n è dell'ordine di 100.000 clienti, scegliere m uguale all'1% di n ci porta a uno spazio di ricerca di 2^{1000} clienti, numero ancora intrattabile, costringendoci allo stesso tempo a determinare efficacemente la composizione del primo centile di clienti più promettenti, problema non banale di per sé.

Tuttavia, l'idea ha un suo senso economico: in molti problemi reali, in effetti, si riscontra il fenomeno proverbialmente noto come *principio 80-20*: l'80% dell'utile è dato dal 20% dei clienti. Esprimendosi in modo più rigoroso, il valore dei clienti molto spesso non è distribuito in modo uniforme, ma concentrato su una frazione relativamente piccola della clientela. Questo giustifica l'idea di effettuare la segmentazione, e di conseguenza progettare i due cataloghi, concentrandosi solo sui clienti più promettenti.

Anche se questi accorgimenti da soli non abbattano la complessità computazionale a livelli trattabili, tuttavia essi hanno una certa utilità pratica, se abbinati ad altri provvedimenti.

Allo stato delle cose, non ci sono metodi euristici formalizzati in grado di garantire buoni risultati in generale per i problemi di segmentazione. Tuttavia, nei casi concreti delle euristiche efficaci possono essere trovate pagando un ragionevole prezzo. Si possono trovare criteri per suddividere lo spazio delle decisioni in un numero ristretto di classi di quasi-equivalenza, all'interno delle quali il valore economico a parità di segmentazione dei clienti oscilla moderatamente. Rappresentando ciascuna classe con una singola decisione si arriva allora a dimensioni trattabili per il problema di ottimizzazione approssimata. Si tratta cioè di fare un clustering sullo spazio delle decisioni in base a un criterio economico. L'idea che suggeriamo è di valutare questo criterio in base a un ristretto campione dei clienti, dopo averli sottoposti a loro volta a clustering in base a criteri intuitivi ed avere selezionato i soli cluster a maggior valore. Con il doppio cluste-

ring su decisioni e clienti le dimensioni del problema collassano; questo non lo rende necessariamente di bassa complessità, ma permette di usare ulteriori tecniche, poniamo l'individuazione di fattori latenti, in grado di suggerire nuove euristiche.

L'idea è questa: non si può affrontare il problema di ottimizzazione in forma pura, perché la sua complessità è intrattabile; ma non si vuole neanche affidarsi totalmente alla segmentazione tradizionale, con la quale la ricerca dell'ottimo è totalmente incorporata nella scelta dei criteri di aggregazione (sia tale scelta a priori o a posteriori); allora si effettua una segmentazione con un numero non troppo ristretto di segmenti risultanti e si riformula il problema di ottimizzazione usando cluster di decisioni cluster di clienti, anziché singole decisioni e singoli clienti. Il risultato dipende certamente ancora dalle scelte di segmentazione, ma lascia spazio alle procedure di ottimizzazione. C'è quindi la possibilità di scegliere fra diverse segmentazioni quella che massimizza il valore atteso, realizzando così una procedura di soluzione approssimata guidata da euristiche.

Esemplifichiamo l'idea sul problema dei cataloghi.

Supponiamo di avere uno spazio delle decisioni D dove ciascuna decisione consiste in un catalogo di r articoli, e di avere prefissato un numero k di cataloghi che intendiamo proporre ai clienti. Prendiamo una soluzione candidata $C \in D^k$ consistente in un insieme di k cataloghi. Sappiamo che assegnati i cataloghi possiamo ottimizzare la segmentazione dei clienti accoppiando a ciascun cliente al catalogo che massimizza i suoi acquisti (per esempio quello che ha il massimo numero di articoli che piacciono al cliente) e che ciò si può fare con un algoritmo di complessità lineare nel numero dei clienti. Determiniamo la segmentazione $S(C)$ ottimale rispetto a C ; resta allora determinato il valore $V(C)$.

Ora apportiamo una modifica a C : sostituiamo nei vari cataloghi l'articolo a con l'articolo b . Otteniamo una nuova soluzione candidata $C_{b/a}$; calcoliamo la nuova segmentazione ottimale dei clienti $S_{b/a}$ e troviamo il nuovo valore $V(C_{b/a})$. Se $V(C_{b/a}) > V(C)$, abbiamo trovato una soluzione migliore e sappiamo che *in questo contesto* l'articolo b è migliore dell'articolo a . Ciò non significa che b sia sempre migliore di a : può accadere che invece a sia in media più gradito agli utenti di b ma che la preferenza per a su b sia concentrata in segmenti di clienti che apportano minor valore, in assoluto o rispetto alla segmentazione indotta dagli altri articoli nei cataloghi che compongono C . Sostituire un articolo con un altro in media più popolare è un ragionevole criterio

euristico ma non si ha garanzia che funzioni in generale. Possiamo essere in presenza di correlazioni non lineari, a causa delle quali il valore di ogni scelta risente delle scelte effettuate contestualmente.

Appurato che in C la sostituzione di a con b è una buona mossa nello spazio delle soluzioni, possiamo ripartire da $C_{b/a}$ per cercare altre mosse che incrementano il valore atteso, realizzando così una forma di *euristica greedy*, consistente nel cercare ad ogni passo una mossa che migliora il risultato, e proseguendo finché sono possibili mosse di questo genere, cioè fino a raggiungere un massimo locale nello spazio di ricerca. Si noti che una modifica moderata nella composizione dei cataloghi induce una modifica moderata nella composizione dei segmenti di clienti. Supponiamo che un cliente sia accoppiato al catalogo C_i e che apporti un valore v_i . Se il gradimento del cliente per a è minore o uguale al suo gradimento per b , allora non c'è motivo perché il cliente cambi segmento, in quanto il suo gradimento complessivo per il catalogo cresce e non crescerebbe di più dopo la modifica in un altro segmento. Se invece il cliente gradisce più a di b , allora può darsi che $C_{i[b/a]}$ non sia il catalogo per lui ottimale nella nuova soluzione $C_{[b/a]}$, perché il suo valore diminuisce di Δv . I nuovi segmenti candidati ad accogliere il cliente non sono però tutti, ma solo quelli che erano per lui peggiori di C_i , ma con uno scarto non maggiore di Δv . Si capisce quindi che un algoritmo accuratamente implementato può contenere la complessità dell'operazione di modifica della segmentazione, che in ogni caso è lineare nel numero dei clienti.

Su questa linea di ragionamento, si vede che il problema di ottimizzazione può essere affrontato con altre euristiche oltre alla ricerca greedy, e le prime che si propongono sono quelle che si rifanno a metodi come il *simulated annealing* e gli *algoritmi genetici*. Il concetto è in tutti i tre casi di creare varianti non troppo dissimili di soluzioni candidate e verificare se iterando questo procedimento con opportuni accorgimenti si può dirigere la ricerca verso regioni più promettenti dello spazio delle soluzioni.

Procedure ibride di segmentazione-ottimizzazione sono in sperimentazione in un programma di ricerca industriale del quale si tratterà successivamente.

Capitolo 4

Valore e business intelligence

1. Qualità di un sistema di data mining

Il problema della valutazione della qualità di un algoritmo o di un sistema di data mining è tradizionalmente affrontato in un'ottica cognitiva, derivante dal *machine learning*: detto in modo informale, un algoritmo è migliore di un altro se i suoi risultati contengono meno errori. Il criterio di giudizio fondamentale è l'esattezza congiunta all'affidabilità, in ultima analisi l'accrescimento di conoscenza.

Prendiamo il classico problema di valutare l'interesse di una regola associativa del tipo $A \rightarrow B$, per esempio nella market basket analysis: la regola dice che c'è una correlazione fra l'acquisto dell'articolo A e l'acquisto dell'articolo B . In generale, regole di questo genere correlano premesse e conseguenze. Misure universalmente note dell'interesse della regola sono di tipo probabilistico (dove le probabilità sono calcolate in base alle frequenze del dataset esaminato):

$$\text{supporto} = \text{prob}(A \wedge B)$$

$$\text{confidenza} = \text{prob}(B \mid A)$$

$$\text{lift} = \frac{\text{prob}(A \wedge B)}{\text{prob}(A) \cdot \text{prob}(b)}$$

Un supporto del 5% significa che nel 5% dei carrelli della spesa esaminati i due articoli si presentano insieme. Una confidenza del 40% significa fra i carrelli in cui è presente l'articolo A il 40% contiene anche B . Il supporto è una misura simmetrica, la confidenza è asimmetrica. Un criterio decisionale comunemente usato è di prendere in considerazione le regole che hanno sia supporto che confidenza maggiori di soglie prefissate, oppure quelle che hanno i più alti supporto e confidenza fra le regole trovate. Il supporto è inteso come misura di affidabilità, la confidenza come misura di utilità predittiva. Il lift misura l'interesse inteso come guadagno di informazione sulla correlazione esistente fra A e B .

Queste misure hanno un senso intuitivo e sono ragionevolmente utili, ma colgono solo in parte le esigenze di un decisore che cerca aiuto nel data mining.

Molte regole con supporto e confidenza alti sono banali, non hanno cioè *interesse*, sono credibili ma non aggiungono nulla a quanto già si sa. Per questo motivo sono state elaborate misure di interesse, per misurare quanto una regola sa *sorprendere*, aggiungere nuova conoscenza. Misure di questo genere sono certo utili, unitamente a supporto e confidenza, a filtrare regole insignificanti e selezionare regole forse interessanti. Però sono orientate a misurare l'interesse cognitivo delle regole, quanto esse servono a *capi- re*; non sono orientate a misurare quanto esse servono a *decidere*. Esistono nella lettera- tura misure assai più complesse di queste, ma sono comunque a orientamento cognitivo.

Anche per gli algoritmi di classificazione la situazione è simile. Un classificatore è mi- gliore di un altro se sbaglia meno, se più spesso colloca un oggetto nella giusta classe di appartenenza. Per esempio, se le classi sono due, chiamiamole *P* e *N*, *positivi* e *negati- vi*, si fa riferimento alla *matrice di confusione*

	P	N
P	tp	fn
N	fp	tn

Sulle righe ci sono le classi effettive, sulle colonne quelle assegnate dal classificatore. Con *tp* indichiamo il numero dei veri positivi, cioè degli oggetti classificati come positi- vi che sono veramente tali; con *fp* il numero dei falsi positivi, cioè degli oggetti classifi- cati come positivi che in realtà sono negativi; analogamente per *tn* e *fn*, i veri negativi e i falsi negativi. Il classificatore perfetto ha $fp = fn = 0$, e quindi $tp = pos$ e $tn = neg$, dove *pos* e *neg* sono il numero di positivi e negativi nella popolazione di casi. Esistono vari indici per misurare la qualità del classificatore a partire dalla sua matrice di confusione. I più noti (usati comunemente nei test clinici) sono

$$sensitività = \frac{tp}{pos} = \frac{tp}{tp + fn}$$

$$specificità = \frac{tn}{neg} = \frac{tn}{tn + fp}$$

Anche in questo caso esistono misure della qualità di una classificazione molto più sofisticate, che però sono ancora cognitive, non decisionali. Possono essere usate per supporto alle decisioni, e lo sono, ma non nascono esplicitamente con questo intento.

Per rispondere in modo più immediato alle esigenze decisionali sono state introdotte misure che tengono conto del costo degli errori di classificazione. In ambito clinico l'importanza di considerare il costo è particolarmente evidente: un falso positivo (un sano che risulta malato) comporta in genere costi economici per ulteriori esami, mentre un falso negativo (un malato che risulta sano) può avere costi anche molto alti, intesi come conseguenze per la salute del soggetto. Buon classificatore in questi casi è quello che minimizza il numero di falsi negativi (alta sensibilità) accettando in cambio più falsi positivi (bassa specificità).

Esistono anche misure che tengono conto del costo da sopportare per raccogliere le informazioni necessarie per la classificazione. Di nuovo, in ambito clinico è chiara l'esigenza a cui queste misure rispondono: certi test clinici sono estremamente costosi in senso economico, oppure comportano alti rischi per il soggetto, o richiedono molto tempo. In questi casi occorre capire se i risultati di un certo test, pesati con la loro affidabilità, sono tali da giustificare i costi del test stesso. Altri test potrebbero offrire un migliore rapporto fra costi della procedura e valore dei suoi risultati.

Nel caso delle regole di associazione occorre avere misure che tengono conto del valore economico delle conseguenze dell'applicazione delle regole. Se un algoritmo di market basket analysis fornisce in output queste due regole

profumo → rossetto

profumo → diamanti

quale delle due è più interessante per l'impresa?

La seconda regola appare interessante per il valore dell'articolo nella parte destra: se usando la regola ci si procura una vendita addizionale di un diamante ciò equivale a un gran numero di vendite addizionali di rossetto ottenute applicando la prima. Tuttavia, la prima regola potrebbe essere molto più efficace in termini di supporto e confidenza, e portare a volumi addizionali di vendita di rossetti tanto più alti da superare il valore dei diamanti addizionali. Oppure, più drasticamente, la prima regola potrebbe essere usata

per implementare una qualche azione di marketing, la seconda no: in tal caso la seconda regola avrebbe semplicemente valore nullo dal punto di vista decisionale.

Si avverte qui una esigenza precedentemente comparsa in questo lavoro: il sistema di business intelligence dovrebbe avere una rappresentazione esplicita delle decisioni possibili e del loro valore economico. Senza questo la componente di data mining rischia di produrre risultati significativi dal punto di vista conoscitivo ma non da quello decisionale.

2. Un modello microeconomico della classificazione

Nella definizione comune di metodo di classificazione si suppone di avere un insieme E di *esempi* e un insieme $S = \{s_1, \dots, s_{n_s}\}$ di classi che formano una partizione di E , cioè

tali che $i \neq j \Rightarrow s_i \cap s_j = \emptyset$ e $\bigcup_{i=1}^{n_s} s_i = E$. Un algoritmo di *classificazione* $C: E \rightarrow S$ è un

algoritmo che dato un esempio $e \in E$ lo assegna ad una delle classi s_i . Un esempio concreto può essere la suddivisione dell'insieme di clienti E nei segmenti s_i ; l'algoritmo può essere uno fra i molti noti nel data mining, per esempio un classificatore bayesiano, una rete neurale, un albero di decisione (non un algoritmo di clustering perché le classi sono definite a priori). Partendo da questa definizione si può studiare la qualità del metodo di classificazione esaminando quanto bene esso riesce a classificare ogni esempio nella sua giusta classe. La misura della sua qualità diventa allora una misura di *errore* da minimizzare. L'approccio può essere modificato ammettendo che errori di tipo diverso hanno conseguenze negative di peso diverso, e in questo caso la qualità è una misura di *costo* da minimizzare.

Vogliamo qui definire un diverso tipo di modello di algoritmo di classificazione, per il quale la qualità dell'algoritmo è una misura di *utilità* da massimizzare. Più rigorosamente, vogliamo definire un modello di processo analitico-decisionale che comprende il data mining per la classificazione come sua fase e che consente di ragionare in modo quantitativo sul valore che il data mining può apportare alle decisioni di impresa.

Nel nostro modello si immagina che il decisore usi l'algoritmo di classificazione per formulare ipotesi; in base a queste ipotesi (ma non soltanto a queste) egli prende decisioni; le decisioni portano a utilità diverse in dipendenza sia dalla bontà dell'algoritmo sia dalla efficienza economica dell'impresa.

Per esemplificare, supponiamo che gli esempi siano clienti e che i clienti siano suddivisi in due segmenti s_1 e s_2 . Abbiamo poi un insieme $H = \{h_1, \dots, h_{n_H}\}$ di ipotesi; nella versione più intuitiva $n_H = n_S$ e l'ipotesi h_i consiste nell'attribuire il cliente e alla classe s_i , ma sono possibili strutture più complesse, nelle quali non c'è corrispondenza diretta fra classi e ipotesi.

L'algoritmo di classificazione è ora visto come una funzione $E \rightarrow H$ che ad ogni esempio assegna una ipotesi. L'ipotesi trovata è poi usata da un processo decisionale $H \rightarrow A$, dove $A = \{a_1, \dots, a_{n_A}\}$ è un insieme di azioni che si possono intraprendere su e . L'utilità derivante dall'applicazione di una azione su un esempio è calcolata da una funzione $E \times A \rightarrow U$. Per esempio, E può essere un insieme di clienti, H l'insieme di ipotesi sull'appartenenza di un cliente a uno dei segmenti predefiniti, A un insieme di possibili offerte da proporre a un cliente, U l'insieme dei possibili utili economici derivanti dall'accettazione di una offerta da parte del cliente.

In questo modello l'algoritmo serve a generare ipotesi che guidano il processo decisionale ma non lo determinano: il passaggio dall'ipotesi all'azione non è determinato dall'algoritmo. La bontà dell'algoritmo, e quindi delle ipotesi che genera, influisce perciò in modo indiretto sull'utilità derivante dal processo analitico-decisionale.

Formalizziamo ora questo modello, introducendo funzioni aleatorie per rappresentare il fatto che i passaggi non sono determinati in modo univoco dagli elementi compresi nel modello, ma dipendono anche da fattori esogeni (per esempio, l'esperienza e l'intuizione dei decisori, o altre informazioni a loro disponibili e non trattate dall'algoritmo di classificazione).

Sia C una matrice di dimensioni $n_S \times n_H$ tale che C_{ij} rappresenti la probabilità che l'applicazione dell'algoritmo di classificazione a un esempio della classe s_i dia come risultato l'ipotesi h_j . Sia D una matrice di dimensioni $n_H \times n_A$ tale che D_{ij} rappresenti la probabilità che il processo decisionale, ottenuta dall'algoritmo l'ipotesi h_i , scelga di intraprendere l'azione a_j . La matrice CD è allora di dimensioni $n_S \times n_A$ e l'elemento di indici

i e j rappresenta la probabilità che in presenza di un esempio della classe s_i il processo decisionale intraprenda l'azione a_j .

Prendiamo una matrice 2×2

$$C = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

Per semplicità interpretiamo le due ipotesi (cui si riferiscono le colonne) come attribuzioni dell'esempio alle due classi (cui si riferiscono le righe): l'ipotesi h_1 asserisce che l'esempio e proviene dalla classe s_1 e l'ipotesi h_2 asserisce che proviene dalla classe s_2 . Perciò la matrice rappresenta il comportamento di un algoritmo di classificazione (che qui non ci interessa specificare) il quale in presenza di un esempio $e \in s_1$ risponde nel 90% dei casi con la classificazione corretta, mentre nell'altro 10% dei casi commette un errore classificando e come elemento di s_2 . Analogo senso ha la seconda riga, dalla quale risulta che l'algoritmo opera peggio sugli esempi di s_2 , che classifica correttamente nell'80% dei casi.

Si noti che in questa particolare interpretazione la matrice C è la cosiddetta *matrice di confusione* dell'algoritmo di classificazione, che contiene le frequenze delle possibile coppie input-output, con le frequenze delle classificazioni corrette sulla diagonale principale. In generale, però, essa è una *matrice stocastica*, nella quale gli elementi di ogni riga hanno per somma 1. Nel nostro esempio è quadrata, ma non è così in generale (il numero delle ipotesi può essere indipendente dal numero delle classi).

Prendiamo ora la matrice

$$D = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$$

anch'essa matrice stocastica. La matrice rappresenta un processo decisionale (che anch'esso non ci interessa qui identificare) il quale in presenza della ipotesi h_1 (che attribuisce l'esempio e alla classe s_1) intraprende l'azione a_1 nell'80% dei casi, mentre nel restante 20% dei casi intraprende l'azione a_2 ; in presenza dell'ipotesi h_2 intraprende l'azione a_2 nel 70% dei casi e l'azione a_1 nel restante 30% dei casi. Il processo decisionale di cui si descrive il comportamento tiene quindi conto di altri criteri oltre a quelli forniti dall'algoritmo di classificazione (forse di un altro algoritmo simile e alternativo, o forse di criteri del tutto diversi).

Il prodotto delle due matrici è

$$CD = \begin{bmatrix} 0.75 & 0.25 \\ 0.40 & 0.60 \end{bmatrix}$$

La prima riga dice che applicando l'algoritmo di classificazione a un esempio della classe s_1 e poi il processo decisionale all'ipotesi formulata dall'algoritmo, nel 75% dei casi si intraprende l'azione a_1 e nel 25% dei casi l'azione a_2 . Come è facile vedere, se prendiamo 100 esempi di s_1 , ci attendiamo che in 90 casi l'algoritmo suggerisca l'ipotesi h_1 e in 10 casi l'ipotesi h_2 (prima riga della matrice C). Dei primi 90 casi, ci attendiamo che l'80%, cioè 72, portino a decidere per l'azione a_1 e il 20% dei casi, cioè 18, per l'azione a_2 (prima riga della matrice D). Dei restanti 10 casi, ci attendiamo che il 30%, cioè 3, portino a decidere per l'azione a_1 e il 70% dei casi, cioè 7, per l'azione a_2 (seconda riga della matrice D). Sommando i casi si ottengono gli elementi della matrice CD .

Il prodotto di matrici esprime proprio questo comportamento. Le probabilità condizionate di intraprendere le azioni in dipendenza dalle classi di provenienza degli esempi sono gli elementi della matrice CD :

$$\begin{aligned} \text{prob}(a_1 | s_1) &= C_{11}D_{11} + C_{12}D_{21} = (CD)_{11} \\ \text{prob}(a_2 | s_1) &= C_{11}D_{12} + C_{12}D_{22} = (CD)_{12} \\ \text{prob}(a_1 | s_2) &= C_{21}D_{11} + C_{22}D_{21} = (CD)_{21} \\ \text{prob}(a_2 | s_2) &= C_{21}D_{12} + C_{22}D_{22} = (CD)_{22} \end{aligned}$$

Ora vogliamo rappresentare nel modello le utilità derivanti dalle azioni scelte dal processo decisionale sulla base (anche) del processo di classificazione. Per far questo definiamo una matrice di utilità U di dimensioni $n_A \times n_S$ in cui l'elemento U_{ij} rappresenta l'utilità derivante dall'applicare l'azione a_i a un esempio proveniente dalla classe s_j . Proseguendo il nostro esempio, la matrice dovrà essere 2×2 , poniamo

$$U = \begin{bmatrix} 10 & 6 \\ 5 & 8 \end{bmatrix}$$

Se applichiamo l'azione a_1 a un esempio proveniente dalla classe s_1 otteniamo una utilità di 10, mentre se l'esempio proviene dalla classe s_2 otteniamo una utilità di 6. Nell'esempio i valori sulla diagonale principale sono maggiori degli altri valori della rispettive righe perché immaginiamo che l'azione a_1 sia la migliore per gli esempi di s_1 e così per a_2 e s_2 , ma questo è solo un accorgimento espositivo.

Sappiamo che se l'esempio e proviene da s_1 abbiamo probabilità uguale a 0.75 di applicare l'azione a_1 e l'utilità attesa è quindi $0.75 \times 10 = 7.5$; abbiamo invece probabilità uguale a 0.25 di applicare l'azione a_2 , con utilità attesa uguale a $0.25 \times 5 = 1.25$. Pertanto l'utilità attesa per un esempio della classe s_1 è $7.5 + 1.25 = 8.75$. Allo stesso modo, l'utilità attesa per un esempio della classe s_2 è $0.40 \times 6 + 0.60 \times 8 = 7.20$.

Le utilità attese sono contenute nella matrice CDU di dimensione $n_S \times n_S$ (per costruzione quadrata), che nel nostro esempio è

$$CDU = \begin{bmatrix} 8.75 & 6.50 \\ 7.00 & 7.20 \end{bmatrix}$$

Gli elementi significativi di questa matrice sono soltanto quelli sulla diagonale principale, poiché un esempio appartiene a una e una sola classe. Dati i nostri processi di classificazione e di decisione per noi un esempio della classe s_1 è più utile di uno della classe s_2 . Se interpretiamo gli esempi come clienti, le classi come segmenti, le azioni come proposte e le utilità come margini di vendita, possiamo dire che un cliente del primo segmento ci porterà, in media, un margine maggiore di un cliente del secondo segmento. Si noti che questo non significa che il primo cliente è intrinsecamente più redditizio del secondo, ma solo che risponde in modo più redditizio ai nostri processi di analisi e decisione e alle offerte che ne derivano. È possibile che con altri processi di analisi e decisione un cliente del secondo segmento diventi in media più redditizio di uno del primo segmento.

Per valutare la qualità dei processi di analisi e decisione dobbiamo misurare il valore atteso su tutti gli esempi, non su uno soltanto. Dobbiamo quindi pesare la frequenza delle varie classi, per esempio la numerosità dei segmenti di clientela. A questo scopo costruiamo una matrice diagonale P di dimensione $n_S \times n_S$, in cui l'elemento P_{ii} è la probabilità che un esempio provenga dalla classe s_i . Se poniamo per esempio

$$P = \begin{bmatrix} 0.60 & 0 \\ 0 & 0.40 \end{bmatrix}$$

indichiamo che ci aspettiamo nel 60% dei casi che l'esempio provenga dalla classe s_1 e nel 40% dei casi dalla classe s_2 .

La matrice $CDUP$ di dimensione $n_S \times n_S$ contiene sulla diagonale principale il valore di utilità atteso considerando la numerosità attesa delle classi:

$$CDUP = \begin{bmatrix} 5.25 & 2.60 \\ 4.20 & 2.88 \end{bmatrix}$$

La somma degli elementi sulla diagonale, la cosiddetta traccia della matrice è la media pesata dei valori attesi per le due classi, cioè $5.25 + 2.88 = 0.60 \times 8.75 + 0.40 \times 7.20$, e in definitiva è il valore atteso dall'intero processo di analisi e decisione (e implementazione della decisione, naturalmente). Dunque, indicando con V il valore di utilità atteso e con τ la traccia di una matrice quadrata, possiamo scrivere

$$V = \tau(CDUP)$$

Il valore $V = \tau(CDUP)$ esprime l'utilità che ci attendiamo dal nostro processo analitico-decisionale. Lo scopo del decisore è naturalmente massimizzare V . Per far questo il decisore può modificare una o più delle matrici che compaiono nella formula. Si pongono quindi due questioni:

1. come determinare o stimare le matrici;
2. come intervenire sulle matrici per ottimizzarle.

3. Determinazione dei parametri del modello microeconomico

La matrice C è relativa all'algoritmo di classificazione, quindi la si può determinare in base all'analisi sia teorica che empirica dell'algoritmo stesso.

Per determinare la matrice C si può studiare il comportamento dell'algoritmo su un test set: sapendo quali sono in effetti le classi degli esempi disponibili, si registrano nella matrice le ipotesi prodotte come risultati dall'algoritmo. Così si determinerà in modo empirico che gli esempi provenienti dalla classe s_1 sono classificati correttamente nel 90% dei casi, o esprimendosi più rigorosamente che il risultato della classificazione è nel 90% dei casi l'ipotesi h_1 .

Oltre alle risultanze empiriche, possono essere interessanti quelle teoriche: per esempio, si potrebbe avere una garanzia a priori che l'algoritmo di classificazione riconosce correttamente la classe di provenienza in almeno l' $x\%$ dei casi. Questo potrebbe già essere utile per certe analisi sull'intero processo analitico-decisionale.

La matrice C , come sappiamo, è di dimensioni $n_S \times n_H$ e correla classi con ipotesi. In pratica, spesso ci sarà corrispondenza uno a uno fra classi e ipotesi e quindi la matrice potrà essere quadrata. In generale, però, per determinare la matrice occorrerà definire lo spazio delle ipotesi.

La matrice D descrive il criterio con il quale ad ogni ipotesi risultante dalla classificazione corrisponde una certa azione da effettuare. Perciò occorre definire sia quali sono queste azioni sia come sono correlate alle ipotesi. Chiaramente qui intervengono conoscenze e principi di tipo manageriale esogeni al modello.

Sotto l'aspetto formale, possiamo rilevare che la relazione tra ipotesi e azioni non è deterministica, per rappresentare il fatto che la scelta dell'azione dipende da considerazioni non rappresentate nel modello e non soltanto dai risultati della classificazione. Si deve tenere presente che, in base a certi risultati della teoria matematica delle decisioni, una matrice deterministica, nella quale cioè in ogni riga c'è un elemento uguale a 1 e tutti gli altri sono 0, può rappresentare ugualmente la correlazione fra ipotesi e azioni e offrire importanti vantaggi in termini computazionali. In pratica, potrà spesso essere naturale proprio usare matrici deterministiche, perché un certo processo decisionale è in effetti deterministicamente basato sui risultati della classificazione.

La costruzione della matrice D può essere un compito estremamente complesso anche dal punto di vista matematico e informatico per le sue implicazioni sull'ottimizzazione, qualora si intenda definire uno spazio delle azioni ottimale. In tal caso, si deve esplorare lo spazio delle azioni possibili alla ricerca del sottospazio ottimale. Se invece la gamma delle azioni possibili è predefinita, o quasi, allora l'associazione fra ipotesi e azioni si riconduce a una previsione dell'effetto di una certa azione su un certo esempio (p.e. di una azione di marketing su un cliente), e quindi diventa un problema di valutazione economica e di previsione della risposta del cliente. Vediamo dunque come la costruzione del modello per una applicazione implichi dei problemi vario genere: valutazioni economiche, previsioni intuitive, previsioni con algoritmi di data mining, ottimizzazione matematica, business intelligence.

La matrice U contiene utilità, e quindi deve essere costruita in base a valutazioni economiche. Nel caso che gli esempi siano clienti, le utilità potrebbero essere dei customer

equity individuali, con tutte le implicazioni e le opportunità di applicazione della business intelligence di cui si discute altrove in questo lavoro. Un aspetto interessante è l'uso del metodo delle opzioni reali nel contesto del modello microeconomico. Le opzioni reali potrebbero essere incorporate nei singoli elementi della matrice U , oppure usate per creare diverse matrici U da usare congiuntamente.

La matrice P contiene le probabilità di provenienza dell'esempio dalle varie classi, quindi le numerosità relative delle classi stesse. La rilevazione delle numerosità può essere determinata a campione (p.e. sul database dei clienti), se le classi stesse sono predefinite. È anche possibile definire le classi stesse in base a criteri di ottimizzazione del processo, e in questo caso le frequenze stesse contribuiscono alla definizione delle classi.

È importante rilevare che fra le quattro matrici possono sussistere delle dipendenze. Per esempio, il processo di classificazione e quello di decisione possono essere progettati separatamente, ma ci si può anche attendere che invece sia collegati da precise relazioni. Anche le utilità possono essere calcolate a posteriori una volta definite le decisioni, ma è anche possibile progettare le decisioni stesse tenendo presenti le utilità. Tutto ciò non è espresso nel modello ma in fase di progettazione dell'intero sistema analitico-decisionale deve essere tenuto presente.

Vediamo dunque che le tecnologie della business intelligence possono essere utilizzate nella varie fasi della determinazione dei parametri del modello: non solo la matrice C , ma l'intero modello è esso stesso un metodo di business intelligence.

4. Ottimizzazione nel modello microeconomico

Per introdurre le problematiche dell'ottimizzazione parliamo prima di tutto dell'uso del modello microeconomico per analisi di tipo what-if. Il modello può infatti essere utilizzato per valutare l'impatto dei vari elementi del processo analitico-decisionale sul risultato espresso in termini di utilità.

Il prodotto $CDUP$ è una espressione lineare costituita da una somma di prodotti, ciascuno dei quali contiene quattro fattori che sono elementi delle quattro matrici. Facendo variare singoli fattori si possono calcolare le derivate parziali del valore rispetto a tali elementi, mediante differenze finite.

Per esempio, se si modifica la matrice C inserendo nella prima riga i valori 1 e 0 anziché 0.9 e 0.1, il valore atteso V passa da 8.13 a 8.28. Questo è intuitivo, perché significa rendere più preciso il processo di classificazione. Però occorre ricordare che la matrice dell'esempio è stata costruita immaginando che ci sia corrispondenza uno a uno fra classi di provenienza e ipotesi. In matrici di dimensione molto più grande questa corrispondenza non c'è in generale e l'analisi di sensitività può essere molto più significativa. Possiamo vedere che se facciamo scendere il valore di C_{11} fino a 0.81, e quindi facciamo salire C_{12} fino a 0.19, V scende fino a 8. Supponiamo di dover prendere una decisione (p.e. la spedizione di un catalogo per corrispondenza o una campagna di telemarketing) che richiede una utilità di almeno 8 per essere valida (poniamo, per coprire almeno i costi di lancio della campagna); è frequente che nelle imprese le decisioni dipendano da una soglia prefissata. Per prendere la decisione dobbiamo quindi assicurarci che il nostro classificatore sappia classificare bene gli esempi della classe s_1 almeno nell'81% dei casi. Se possiamo garantire questo, allora possiamo prendere la decisione, senza bisogno di quantificare con precisione la qualità dell'algoritmo e quindi la sua matrice C .

Analogamente, possiamo modificare la matrice P portando gli esempi della classe s_1 al 70% del totale, anziché 60%: il valore V aumenta da 8.13 a 8.29.

Veniamo ora a considerare le problematiche di ottimizzazione del modello microeconomico.

È naturale pensare prima di tutto all'ottimizzazione della matrice D fissando le altre, essendo D l'unica sulla quale il decisore aziendale ha certamente piena libertà di azione. Per questo problema lo spazio delle possibili soluzioni è costituito da tutte le matrici stocastiche. Il criterio di ottimizzazione è

$$\max_{D_{ij}} \tau(CDUP)$$

e i vincoli, derivanti dalla definizione di matrice stocastica, sono

$$\sum_{j=1}^{n_A} D_{ij} = 1 \text{ per } i = 1, \dots, n_H$$

$$D_{ij} \geq 0 \text{ per } i = 1, \dots, n_H \text{ e } j = 1, \dots, n_A$$

Si tratta di un problema di ottimizzazione lineare al quale possono essere applicati i metodi noti.

Se l'ottimizzazione di D è il problema più naturale, tuttavia non è l'unico che può essere utilmente posto. Il decisore aziendale può infatti chiedersi come ottimizzare C , vale a dire come ottimizzare il suo processo di classificazione, che può consistere in un singolo algoritmo o in un processo più articolato.

Come abbiamo detto, è verosimile che spesso sia $n_H = n_S$ e che la matrice C rappresenti la cosiddetta matrice di confusione. In questo caso, il problema di ottimizzare C può sembrare indipendente dal modello microeconomico, in quanto si cercherà in ogni caso di ottimizzare la capacità dell'algoritmo di classificazione di operare quanto più correttamente è possibile. Se però assumiamo un'ottica di analisi economica la questione si

pone diversamente. Supponiamo di trasformare la matrice C da $\begin{bmatrix} 0.90 & 0.10 \\ 0.20 & 0.80 \end{bmatrix}$ a

$\begin{bmatrix} 0.95 & 0.05 \\ 0.25 & 0.75 \end{bmatrix}$, modificando in modo opportuno i suoi parametri per rendere più probabile la classificazione di e in s_1 : il valore di utilità V cresce da 8.13 a 8.19. Questo significa che "spingere" l'algoritmo verso la classificazione nella classe s_1 è economicamente vantaggioso, anche se questo non significa necessariamente che la nuova versione dell'algoritmo sia più corretta della precedente in termini di corretta classificazione. Se

arriviamo a trasformare C nella matrice $\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ vediamo però che l'utilità decresce fino

a $V = 7.96$, peggiore del valore originario. Questo indica che la strategia ottimale non è semplicemente quella di forzare in ogni caso l'attribuzione dell'esempio e alla classe s_1 : siamo davanti a un problema di ottimizzazione non banale, nel quale si deve cercare il punto di equilibrio fra la correttezza della classificazione e l'utilità delle varie possibili scelte di classificazione (o utilità perduta per i vari possibili errori). Qui vediamo come il concetto informativo e quello economico del valore della classificazione possano divergere.

Resta inteso che la matrice C ideale, sempre sotto l'interpretazione naturale, non è altro che la matrice identica, nel nostro esempio $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Anche per la matrice U delle utilità è ovvio che ciascun elemento singolarmente preso dovrebbe essere il più grande possibile, e che questo significa semplicemente che l'impresa dovrebbe ottimizzare i suoi processi e le sue performance. Però nella realtà occorre stabilire delle priorità e concentrare l'impegno verso il miglioramento di performance sotto alcuni scenari anziché altri. Il modello aiuta a capire quali vantaggi ci si possono attendere da un intervento migliorativo del sistema delle utilità: lasciando inalterate le altre matrici, si cercherà di scegliere la matrice U che massimizza il valore, sempre risolvendo un sistema lineare simile a quello visto per la matrice D .

Anche la matrice P può essere oggetto di ottimizzazione. È vero che il decisore non può modificare le proporzioni delle classi nella popolazione degli esempi, ma può concentrarsi verso sottoinsiemi della popolazione che offrono le migliori matrici P , ferme restando le altre matrici. Questo suggerisce per esempio la possibilità di selezionare vari sottoinsiemi da una popolazione di clienti e per ognuno procedere con un diverso processo analitico-decisionale, facendo in modo che per quel particolare processo il sottoinsieme offra la migliore matrice P .

È interessante riflettere su cosa accade se si fissano tre matrici del modello e ci si propone di intervenire sulla quarta (specie su D , ma anche sulle altre). Ancora più interessante, e purtroppo molto complesso, è riflettere sulla ottimizzazione simultanea di due o più matrici. Gli elementi delle varie matrici non sono realmente indipendenti: se le azioni nella matrice D consistono in offerte di sconto ai clienti, allora le utilità della matrice U sono funzioni di tali azioni, perché gli sconti influiscono sulle propensioni all'acquisto dei clienti. Ecco quindi che fra le variabili in gioco sussistono relazioni che possono richiedere l'inclusione nelle analisi di altri fattori e che possono rendere più adeguato, o necessario, un modello non lineare. Tutto ciò non significa certo che il modello non sia utile, ma solo che esso è un modello lineare di approssimazione a una realtà molto complessa. Il modello presuppone che le relazioni fra le variabili siano già in-

cluse nella formulazione lineare del problema, usando metodi opportuni, compresi quelli offerti dalle tecnologie di business intelligence.

5. Un esempio di applicazione del modello microeconomico

Useremo ora il modello microeconomico per impostare un processo decisionale, semplice ma non privo di realismo, relativo al targeting di una campagna promozionale.

Una impresa vuole promuovere un suo prodotto con una campagna di marketing telefonico, offrendo uno sconto. Il prezzo normale del prodotto è p , lo sconto offerto è s , il costo del contatto telefonico è c . L'impresa utilizza il suo customer database e un algoritmo di classificazione per individuare i clienti ai quali è conveniente proporre l'offerta. In un'ottica decisionale il criterio per individuare il target non è semplicemente "quali clienti accetteranno l'offerta?" bensì "quali clienti daranno un valore superiore ricevendo l'offerta piuttosto che non ricevendola?".

Ogni cliente ha una serie di attributi X_1, \dots, X_n riportati nel customer database, che sono utilizzati da un algoritmo di data mining. Il cliente ha poi due attributi comportamentali che sono ignoti: l'attributo binario A_0 che vale 1 se il cliente acquista in assenza di proposta e 0 se non acquista, e l'attributo A_1 che vale 1 se il cliente acquista in caso di proposta e 0 se non acquista. L'algoritmo di data mining dovrebbe predire i valori di questi due attributi ignoti; la predizione si risolve in questo caso in una classificazione, dove le classi sono quattro:

1. la classe S_{00} dei clienti che non acquistano né in assenza né in presenza di proposta (non sono in alcun caso interessati al prodotto) ;
2. la classe S_{01} dei clienti che acquistano in presenza di proposta ma non acquisterebbero in assenza di proposta (sono i clienti da conquistare con l'offerta);
3. la classe S_{10} dei clienti che acquisterebbero in assenza di proposta ma non acquistano più in presenza di proposta (con loro l'offerta è controproducente);
4. la classe S_{11} dei clienti che acquistano sia in assenza che in presenza di proposta (sono in ogni caso disposti all'acquisto e per loro l'offerta è superflua).

L'algorithmo di data mining esamina gli attributi X_1, \dots, X_n di un cliente e lo classifica in uno dei quattro segmenti.

Per costruire la matrice di classificazione C dobbiamo considerare che i segmenti non sono osservabili. Il punto è che noi non possiamo osservare su uno stesso cliente sia l'attributo A_0 che l'attributo A_1 , perché lo stesso cliente non può essere sia sottoposto sia non sottoposto alla proposta. Se proponiamo l'offerta a un cliente e questi l'accetta, allora sappiamo che è nell'unione $S_{01} \cup S_{11}$, ma non sappiamo in quale dei due segmenti stia; se invece non accetta l'offerta sappiamo che è nell'unione $S_{00} \cup S_{10}$, ma di nuovo non possiamo discriminare fra i due segmenti. Il problema non è risolvibile con la sola osservazione del comportamento nel corso di una campagna; tuttavia, se dal piano rigorosamente scientifico ci spostiamo pragmaticamente a quello operativo, il problema è superabile sia pure con accorgimenti che danno luogo a questioni sul piano metodologico. Prendiamo un campione di coppie di clienti e in ogni coppia presentiamo la proposta ad uno solo dei due. Se assumiamo che i due clienti siano "equivalenti", possiamo assegnare un valore all'attributo A_0 in base al comportamento del cliente che non ha ricevuto l'offerta e un valore all'attributo A_1 in base al comportamento dell'altro. Naturalmente questo assegnamento è affidabile nella misura in cui i due clienti sono davvero equivalenti, il che dipende da quanto bene abbiamo scelto il criterio di equivalenza, e questo potrebbe essere a sua volta un problema di data mining. In alternativa, potremmo ricorrere ad interviste ad un campione di clienti proponendo loro l'offerta e chiedendo esplicitamente quale sarebbe stato il loro comportamento in assenza di offerta.

Assumiamo dunque che con l'osservazione del comportamento di un campione di clienti si possa costruire la matrice C :

$$C = \begin{bmatrix} C_{00 \rightarrow 00} & C_{00 \rightarrow 01} & C_{00 \rightarrow 10} & C_{00 \rightarrow 11} \\ C_{01 \rightarrow 00} & C_{01 \rightarrow 01} & C_{01 \rightarrow 10} & C_{01 \rightarrow 11} \\ C_{10 \rightarrow 00} & C_{10 \rightarrow 01} & C_{10 \rightarrow 10} & C_{10 \rightarrow 11} \\ C_{11 \rightarrow 00} & C_{11 \rightarrow 01} & C_{11 \rightarrow 10} & C_{11 \rightarrow 11} \end{bmatrix}$$

dove l'indice $C_{ij \rightarrow kl}$ rappresenta la probabilità che un cliente del segmento S_{ij} sia classificato come appartenente al segmento S_{kl} . Usiamo quindi l'interpretazione naturale, per la quale le ipotesi sono in corrispondenza uno a uno con le classi.

Le possibili azioni sono due: a_0 che consiste nel non proporre l'offerta e a_1 che consiste nel proporla. La matrice D avrà dunque dimensione 4×2 . La si può costruire in modo ovvio perché proporre l'offerta è l'azione utile soltanto per i clienti del segmento S_{01} , per i quali a fronte di un costo c si ha un ricavo $p - s$ poiché si ottiene una vendita che altrimenti non si sarebbe stata, al prezzo normale p diminuito dello sconto s . Se $p - s > c$, come è necessario perché l'intera campagna di promozione abbia senso, conviene proporre l'offerta. Se invece il cliente appartiene al segmento S_{00} l'offerta non provoca l'acquisto e implica la perdita del costo di proposta c . La proposta implica una perdita di importo $c + p$ se il cliente appartiene al segmento S_{11} , perché la vendita ci sarebbe stata comunque. Se poi il cliente appartiene al segmento S_{10} allora la proposta provoca una perdita di importo $c + p$, perché implica sia il suo costo sia la mancata vendita a prezzo intero.

Perciò la matrice D è costruita così

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$$

(la seconda colonna è quella relativa all'azione a_1 di offerta e la seconda riga quella relativa al segmento S_{01}).

Il prodotto delle due matrici è

$$CD = \begin{bmatrix} 1 - C_{00 \rightarrow 01} & C_{00 \rightarrow 01} \\ 1 - C_{01 \rightarrow 01} & C_{01 \rightarrow 01} \\ 1 - C_{10 \rightarrow 01} & C_{10 \rightarrow 01} \\ 1 - C_{11 \rightarrow 01} & C_{11 \rightarrow 01} \end{bmatrix}$$

La matrice indica che si intraprende l'azione a_1 con una probabilità uguale alla probabilità che l'algoritmo di classificazione attribuisca il cliente in esame al segmento S_{01} .

La matrice U può essere costruita in base alla definizione del problema:

$$U = \begin{bmatrix} 0 & 0 & p & p \\ -c & p - s - c & -c & p - s - c \end{bmatrix}$$

La prima riga è relativa all'azione a_0 con la quale non si propone l'offerta scontata. Se il cliente non acquista (segmenti S_{00} in prima colonna e S_{01} in seconda colonna) l'utilità è

0, perché non ci sono né costi né ricavi. Se il cliente acquista (segmenti S_{10} in terza colonna e S_{11} in quarta colonna) si ha un ricavo p , il prezzo di vendita non scontato.

La seconda riga è relativa all'azione a_1 con la quale si propone l'offerta scontata. Se il cliente non acquista (segmenti S_{00} in prima colonna e S_{10} in terza colonna) l'utilità è $-c$, perché c è il costo della proposta. Se il cliente acquista (segmenti S_{01} in seconda colonna e S_{11} in quarta colonna) si ha un ricavo $p - s$, il prezzo di vendita scontato, e un costo c per la proposta, con una utilità totale $p - s - c$.

Utilizzando i risultati dell'indagine a campione costruiamo la matrice

$$P = \begin{bmatrix} f_{00} & 0 & 0 & 0 \\ 0 & f_{01} & 0 & 0 \\ 0 & 0 & f_{10} & 0 \\ 0 & 0 & 0 & f_{11} \end{bmatrix}$$

dove f_{ij} è la probabilità che un cliente provenga dal segmento S_{ij} (abbiamo usato la lettera f , per frequenza, perché la lettera p è stata usata per il prezzo di vendita).

La matrice $CDUP$ è dunque

$$\begin{bmatrix} -cf_{00}C_{00 \rightarrow 01} & 0 & 0 & 0 \\ 0 & (p-s-c)f_{01}C_{01 \rightarrow 01} & 0 & 0 \\ 0 & 0 & pf_{10}(1-C_{10 \rightarrow 01})-cf_{10}C_{10 \rightarrow 01} & 0 \\ 0 & 0 & 0 & pf_{11}(1-C_{11 \rightarrow 01})+(p-s-c)f_{11}C_{11 \rightarrow 01} \end{bmatrix}$$

L'azione a_1 di proposta avviene, come sappiamo, con la probabilità $C_{ij \rightarrow 01}$ che il cliente sia classificato dall'algorithm di data mining nel segmento S_{01} , qualunque sia il vero segmento S_{ij} di appartenenza del cliente. I valori sulla diagonale indicano le utilità per i clienti di ognuno dei quattro segmenti, utilità opportunamente pesate con le probabilità che la proposta avvenga o non avvenga. Tali valori sono quelli corretti, come è facile verificare.

La somma dei valori sulla diagonale è il valore atteso dalla campagna basata sul processo analitico-decisionale descritto dalle matrici C e D . Si può ora effettuare una analisi what-if per capire quali leve usare per migliorare il risultato. Per esempio, si può pensare di modificare qualche parametro dell'algorithm di classificazione e, in base al cam-

pione di clienti studiato, vedere come varia la matrice C e in definitiva V , tenendo fisse U e P (nel nostro esempio la matrice D è chiaramente ottimale in questa forma, indipendentemente dalle altre).

Si può anche pensare di intervenire su U variando lo sconto s : questa è una ipotesi interessante, che introduce nel modello la *elasticità della propensione all'acquisto rispetto al prezzo*. Se lo sconto aumenta, verosimilmente alcuni clienti migrano dal segmento S_{00} verso il segmento S_{01} , perché l'offerta con sconto $s + \Delta s$ è più attraente. Invece i clienti di S_{10} resteranno dove sono, e forse il segmento potrebbe attirare altri clienti agli occhi dei quali il prodotto è svalorizzato proprio dal maggiore sconto. Dunque, al variare di s variano in numerosità i segmenti stessi e perciò variano la matrice P e il valore atteso V . Per effettuare l'analisi what-if nell'ipotesi di variazioni di s occorre quindi stimare in qualche modo questa funzione di elasticità $\frac{\partial P}{\partial s}$.

6. Valore di un classificatore

Il modello microeconomico facilita la stima del valore che un classificatore apporta a un processo analitico-decisionale. Nell'esempio della campagna di telemarketing possiamo vedere come parametri le matrici D , U e P e confrontare due varianti $C^{(1)}$ e $C^{(2)}$ della matrice C , che descrivono due diversi classificatori, per stabilire quale sia il migliore studiando la disequazione

$$\tau(C^{(1)}DUP) \geq \tau(C^{(2)}DUP)$$

e determinando quali ipotesi sulle altre matrici rendono la disequazione vera o falsa (in questa operazione occorre utilizzare gli stessi spazi di classi, ipotesi e azioni per i due classificatori). Può accadere che una delle due matrici di classificazione sia sempre migliore dell'altra per qualunque scelta delle altre matrici: in questo caso uno dei due classificatori è in assoluto migliore dell'altro. Può invece accadere che il confronto fra i due classificatori dipenda dalla scelta delle altre matrici di caso in caso. Se così è, allora si può porre la questione di scegliere fra i due in base allo specifico problema decisionale.

Il confronto fra i valori attesi usando i due classificatori permette non solo di stabilire un ordine, ma anche di quantificare la differenza di valore fra i due classificatori. Fissate le altre matrici, se abbiamo un classificatore descritto dalla matrice C e ci chiediamo se investire per costruirne un altro più efficace, possiamo creare una matrice C' che descrive l'ipotetico futuro classificatore e valutare i benefici che esso potrebbe apportarci, per decidere se procedere o meno all'investimento. L'incremento di valore atteso

$$\Delta V = \tau(C'DUP) - \tau(CDUP)$$

deve coprire l'investimento previsto per la realizzazione del nuovo classificatore.

Il confronto fra due matrici di classificazione permette di assegnare ad ogni classificatore un valore relativo. Si può anche definire una misura di valore assoluto prendendo come classificatore di riferimento quello puramente casuale, nel quale tutti gli elementi sono uguali, e quindi tutti uguali a $\frac{1}{n_H}$ (la somma su ogni riga deve essere 1 e la matrice C ha n_H colonne). Il valore assoluto di un classificatore è il differenziale di valore rispetto al classificatore casuale basato sulle stesse classi e le stesse ipotesi.

7. Estensioni dinamiche del modello

Il modello microeconomico presentato può essere reso dinamico, cioè capace di gestire variazioni nel tempo dello stato del mondo.

Un modo di fare questo è parametrizzare le matrici sul tempo. Il valore atteso dal sistema analitico-decisionale al tempo t è allora

$$V_t = \tau(C_t D_t U_t P_t)$$

Il valore atteso su tutto l'orizzonte temporale è allora la somma attualizzata dei valori attesi nei vari periodi

$$V_\infty = \sum_{t=1}^{\infty} \tau(C_t D_t U_t P_t) (1+d)^{-t}$$

Con questa estensione si può usare il modello nel contesto di una analisi di customer equity e fra le varie possibilità una veramente stimolante è di considerare anche le opzioni reali. Le opzioni reali possono essere considerate nella valutazione delle utilità

nella matrice U , tenendo conto di opzioni di investimento per ampliare le relazioni con i clienti, oppure per migliorare i processi dell'impresa. Questo è per così dire un uso convenzionale della metodologia. Ma una possibilità innovativa è di usarle per valutare gli interventi di miglioramento del classificatore: ciò significa applicare una metodologia finanziaria e manageriale molto moderna e sofisticata alla microeconomia della business intelligence, tema di ricerca di stimolante attualità.

Un'altra estensione dinamica del modello microeconomico tiene conto dell'evoluzione del cliente, che con il tempo migra da una classe all'altra, in parte spontaneamente e in parte in risposta alle azioni che l'impresa attua nei suoi confronti (per concretezza interpretiamo gli esempi del classificatore come clienti).

Mettiamoci in un'ottica markoviana di processo stocastico, supponendo che la risposta del cliente all'azione dell'impresa dipenda solo dal suo stato attuale, cioè dal segmento cui appartiene al momento: la sua storia è tutta riassunta nel suo segmento di appartenenza, e quindi tutti i clienti di uno stesso segmento hanno la stessa funzione di risposta alle azioni di marketing.

Supponiamo di avere fissato un insieme di probabilità T_{ijk} del passaggio di un cliente dal segmento i al segmento j in risposta all'azione k . Quindi per ogni azione a_1, \dots, a_{n_A} abbiamo una matrice di dimensione $n_S \times n_S$ che esprime le probabilità che un cliente passi da un segmento all'altro, e per ogni segmento s_1, \dots, s_{n_S} abbiamo una matrice di dimensioni $n_A \times n_S$ che esprime la probabilità che un cliente passi a un certo segmento in risposta a una certa azione.

Se T_i è la matrice stocastica $n_A \times n_S$ di transizione segmento-segmento relativa all'azione a_i , allora la matrice

$$M = \sum_{i=1}^{n_A} CDT_i$$

ottenuta sommando elemento a elemento le varie matrici CDT_i è la matrice stocastica di transizione segmento-segmento del cliente in funzione del processo analitico-decisionale (oltre che delle transizioni spontanee). Essendo una matrice markoviana, le sue potenze esprimono le probabilità di transizione in tutti i periodi futuri.

Con questo strumento l'impatto del processo analitico-decisionale sul customer equity può essere studiato nella sua evoluzione temporale, utilizzando gli strumenti concettuali già descritti.

Fin qui abbiamo fatto riferimento all'interpretazione delle classi come segmenti di clienti. Possiamo anche interpretare gli s_i non come aggregati ma come stati individuali, lasciando inalterato l'apparato formale. In questa interpretazione trova allora posto l'uso di modelli della relazione fra impresa e cliente, come quello discusso nel capitolo 2. Per esempio, un stato può essere interpretato come uno stadio di avanzamento del cliente lungo il continuum che porta alla lealtà verso l'impresa.

Capitolo 5

Riflessioni sulla customer intelligence

Nella letteratura sui datawarehouse si incontra un gran numero di esempi di progettazione di *customer datawarehouse* o più in generale di *marketing datawarehouse*. Questi esempi sono alquanto simili fra loro e per lo più riproducono alcuni esempi classici [Kimball 2002, Adamson 1998] che, sebbene a carattere divulgativo, sembrano in effetti utili nel guidare i progettisti di data warehouse incontrano nella pratica professionale. Un discorso analogo vale per le applicazioni di data mining presenti nella letteratura. Qui il carattere didattico e divulgativo è ancora più marcato, e nel complesso la letteratura sembra offrire ai progettisti una guida meno puntuale che nel caso dei datawarehouse, cosa peraltro comprensibile per la natura stessa dell'argomento.

Senza voler entrare in particolari, per i quali si rimanda alla letteratura, e senza pretesa di completezza sistematica, tentiamo qui di enunciare alcuni principi di progettazione del customer datawarehouse che tengano conto delle considerazioni teoriche svolte in questo lavoro e che quindi arricchiscano, senza sostituirli, i principi assestati nella letteratura.

1. Azioni e risultati

Discutendo di segmentazione, abbiamo visto come in linea di principio sia desiderabile considerare congiuntamente clienti e azioni di marketing, perché questo permette di inquadrare la segmentazione in un contesto teorico di ottimizzazione, nel quale essa svolge un ruolo di euristica per ridurre gli enormi spazi di ricerca delle soluzioni e consentire utili approssimazioni limitando la complessità computazionale a dimensioni trattabili.

Questo è un principio metodologico che resta valido anche quando i progettisti non intendano arrivare a porre formalmente un problema di ottimizzazione decisionale sui dati contenuti nel datawarehouse. Il customer datawarehouse (o il customer database, secondo i casi) viene in pratica molto spesso utilizzato per la gestione di campagne di marke-

ting. Però di frequente le campagne stesse sono rappresentate nel datawarehouse in modo molto meno ricco e preciso di quanto accade per i clienti. L'esperienza dice però che i dati sui clienti perdono molto del loro valore potenziale se non sono correlati ai dati sulle campagne di marketing; più in generale, se non sono correlati ai dati sulle azioni e sulle decisioni di marketing.

Vedendo la questione in modo astratto, lo scopo primario di un sistema di business intelligence è aiutare il decisore umano nell'esplorazione di uno spazio di eventi possibili, dove ciascun evento è una sintesi di azioni e conseguenze. Una parte di questi eventi, quelli effettivamente accaduti, è rappresentata direttamente nel datawarehouse; gli altri, gli eventi che potrebbero accadere, sono in qualche modo generabili con simulazioni e analisi che usano tecniche di data mining. Se si assume questa ottica, allora si richiede al datawarehouse una rappresentazione esplicita:

- delle azioni possibili
- delle conseguenze possibili
- delle relazioni possibili fra azioni e conseguenze

Insistiamo sull'aggettivo *possibili* perché gli eventi storicamente accaduti devono essere visti come un minuscolo sottoinsieme di quelli possibili. Ciò può essere superfluo a fini di reporting, ma è essenziale a fini decisionali, specie in un contesto di impresa proattiva, nel quale proprio l'esplorazione del nuovo è il problema essenziale e l'attività caratteristica.

Poiché il decisore non vuole soltanto capire cosa è possibile, ma anche assegnare alle possibilità dei valori, ne consegue che il datawarehouse dovrebbe essere in grado di rappresentare in modo esplicito il valore, concetto centrale della business intelligence.

Proseguendo il ragionamento, possiamo dire che il datawarehouse dovrebbe rappresentare esplicitamente non soltanto i risultati ma anche i fattori che portano a quei risultati, in quanto questi fattori collegano le azioni ai risultati.

Prendiamo il più classico degli esempi di *fatto* di un datawarehouse: le vendite.

Una vendita è un fatto aggregato, nel senso che si scompone in altri fatti: nella stessa vendita si distinguono vendite di vari articoli, così come avviene tipicamente in un carrello della spesa. Partiamo dall'analisi di una vendita elementare, cioè di un solo articolo.

Prima di tutto ci chiediamo come rappresentare il valore. La scelta naturale è di rappresentarlo come margine, che scomponiamo nei suoi costituenti così:

$$\text{valore} = \text{margine}$$

$$\text{margine} = \text{ricavo} - \text{costo}$$

$$\text{ricavo} = \text{quantità venduta} \times \text{prezzo unitario} - \text{sconto} - \text{resi}$$

Vediamo l'opportunità di inserire nella *fact table Vendite* le varie grandezze che compaiono nelle uguaglianze: si potrà così indagare sul valore in modo più analitico.

Fin qui non abbiamo indagato sui driver del valore, ma piuttosto sui suoi componenti logico-aritmetici. Ora portiamo il ragionamento dal piano analitico al piano decisionale e ci chiediamo come possiamo intervenire sul valore per ottimizzare i risultati, il che implica chiedersi quali sono i determinanti, i driver del valore. La *forma mentis* che assumiamo è quella di risalire da un obiettivo alle leve manovrabili per raggiungerlo.

Come ottimizzare la quantità venduta? Questa grandezza non è decisa dall'impresa, non direttamente e non completamente: è una scelta del cliente. Quindi dobbiamo chiederci come influenzare la scelta del cliente. Poniamo il problema in termini di opzioni di ampliamento della relazione, che abbiamo schematizzato in opzioni di:

1. up selling
2. cross selling
3. trading up
4. business extension.

Abbiamo definito lo up selling come un aumento di volume degli acquisti, derivante da economie di scala del cliente nell'acquisto e nel consumo. La teoria ci dice dunque che nel datawarehouse e nelle applicazioni di data mining dovranno comparire questi concetti, connessi in qualche modo al fatto vendita. Questa considerazione non ci dice certamente quali tabelle di fatti o dimensioni dovremmo inserire nello schema a stella del datawarehouse, né quali attributi dovranno avere, né le aggregazioni che sceglieremo in un sistema Olap, né tanto meno come dovranno essere gli algoritmi di data mining che opereranno sulle tabelle. Piuttosto, serve metodologicamente per mettere a fuoco le problematiche da affrontare.

Il cliente (che, si noti, a questo punto dell'analisi è ancora del tutto sconosciuto) può aumentare le quantità acquistate per motivi esogeni all'impresa, come un aumento di reddito o un ampliamento della famiglia. Motivazioni di questo genere sono rilevanti per la customer intelligence in quanto opportunità per intervenire, pur non costituendo di per sé leve decisionali vere e proprie. Pertanto, è desiderabile avere nel datawarehouse informazioni di questo genere. Le due citate sono peraltro variabili che tradizionalmente sono usate nella segmentazione della clientela. Ciò non è casuale, poiché si tratta di variabili che determinano la funzione di risposta di un cliente alle azioni di marketing, che in precedenza abbiamo indicato come la vera essenza della segmentazione in quanto euristica di aggregazione per l'ottimizzazione del valore.

Dunque, il cliente dovrebbe realizzare economie di scala nell'acquisto e nel consumo, perciò dobbiamo individuare le informazioni che ci aiutano a comprendere e governare le variabili endogene, cioè sotto il controllo dell'impresa, che consentono o provocano tali economie.

La più ovvia è il prezzo unitario: diminuendo tale prezzo il cliente può essere motivato ad aumentare la quantità, e sotto opportune condizioni ciò può portare a maggior valore. Una decisione in merito richiede la conoscenza della funzione di risposta di un cliente alla diminuzione di prezzo, diciamo l'elasticità della domanda individuale rispetto al prezzo. Per conoscere tale funzione è utile estrarre dalla storia del cliente la correlazione tra prezzi e quantità acquistate *o non acquistate*. Ecco un concetto metodologicamente importante: il datawarehouse dovrebbe tenere traccia degli eventi ma anche dei non-eventi, delle occasioni in cui non accade qualcosa che sarebbe potuto accadere. Se un cliente rifiuta una offerta telefonica del nostro call center, o anche esce dal nostro negozio senza aver comprato, questo è comportamento porta informazione sulla funzione di risposta del cliente alle azioni di marketing e quindi ha un valore conoscitivo e decisionale.

Possiamo allora enunciare questo principio guida: *tenere traccia delle manifestazioni della funzione di risposta del cliente alle azioni dell'impresa*. Questo enunciato non è esattamente equivalente al principio di tenere traccia del comportamento del cliente. Qui si enfatizza nel comportamento l'aspetto della funzione di risposta, che nel problema di ottimizzazione del valore della clientela gioca un ruolo chiave.

Se pensiamo al comportamento del cliente come manifestazione della funzione di risposta, ne consegue che le vendite vere e proprie danno una visione parziale del comportamento del cliente. Il datawarehouse dovrebbe registrare le occasioni in cui un'azione dell'impresa ha raggiunto il cliente, e quindi dovrebbe contenere una *fact table* *Contatti*. Nella tabella dei contatti saranno registrati gli *esiti*, cioè le risposte del cliente, che sono osservazioni della sua funzione di risposta, e che possono essere eventi diversi dalla vendita, compresi i *non eventi*. Per esempio, un contatto telefonico per la proposta di un abbonamento può avere come esito la sottoscrizione, il rifiuto totale, il rinvio, la richiesta di riparlare nell'orario in cui ci sarà in casa il decisore della famiglia ...

Gli eventi Vendite possono dunque essere visti come una classe di risposte agli eventi Contatti. Non sempre questo modo di impostare l'analisi è in pratica fruttuoso, ma metodologicamente deve essere tenuto presente: in linea di principio, una informazione che non ha correlazioni utilizzabili con le azioni dell'impresa è una informazione inutile.

Gli eventi contatti e gli eventi vendite sono correlati tramite i clienti, e ci si attende che un datawarehouse per il marketing non possa non avere una *fact table* *Clienti*. Ciò non significa però che tutti i contatti siano individuali, relativi a un determinato cliente. Esistono contatti collettivi, per esempio inserzioni pubblicitarie, che raggiungono insieme di clienti e di regola non abbiamo traccia sul datawarehouse dei contatti individuali originati dal contatto collettivo. Possiamo però avere ipotesi probabilistiche di contatto individuale se sappiamo che un contatto collettivo ha raggiunto un insieme di clienti e che questo determinato cliente fa parte di quell'insieme.

Astraendo, in una situazione ideale il datawarehouse contiene la storia delle azioni dell'impresa che hanno raggiunto il cliente, compresa l'azione nulla, e la storia delle risposte del cliente a quelle azioni, compresa la risposta nulla. Applicando algoritmi di analisi a queste due storie correlate si cerca di acquisire informazioni sulla funzione di risposta del cliente. Un modello formale molto astratto potrebbe essere una sequenza di questo tipo

$$(a_1, r_1, e_1, t_1) \dots (a_n, r_n, e_n, t_n)$$

dove le a_i sono azioni dell'impresa, le r_i sono risposte del cliente e i t_i sono gli istanti di tempo in cui azioni e risposte avvengono, tendendo presente che le azioni e le risposte possono anche essere nulle, dei non eventi. Il fatto che ci siano azioni nulle indica che le risposte del cliente sono determinate anche da fattori diversi dalle azioni dell'impresa. Possiamo quindi introdurre nel modello anche degli eventi esogeni e_i che compaiono nel datawarehouse:

$$(a_1, r_1, e_1, t_1) \dots (a_n, r_n, e_n, t_n)$$

La conoscenza di questi eventi, per esempio l'acquisto di una casa, il matrimonio o la nascita di un figlio, serve a definire più realisticamente il comportamento del cliente come risposta alle azioni dell'impresa e ad altri fattori. La funzione di risposta del cliente può variare anche molto fortemente dopo eventi di questo genere.

Possiamo ancora arricchire il modello interpretandolo in modo probabilistico e quindi introducendo variabili aleatorie che rappresentano varianze non spiegate, o introdurre molti altri raffinamenti, peraltro di difficile messa in pratica. Uno di questi raffinamenti è però particolarmente interessante, ed è suggerito dalle considerazioni teoriche svolte in precedenti sezioni di questo lavoro. Abbiamo visto che il valore del cliente dipende fortemente da *stati interni* del cliente (si ricordino i concetti di fedeltà comportamentale e fedeltà mentale). Si arriva quindi a pensare al cliente non più come manifestazione di una *funzione* di risposta alle azioni dell'impresa ma come manifestazione di un *processo* che si svolge fra due *automi* dotati di stato interno, l'impresa e il cliente.

Il passaggio dal concetto di comportamento come funzione al concetto di comportamento come processo non è banale e permette di concettualizzare diversamente certi aspetti della business intelligence e del datawarehouse che la alimenta. Il datawarehouse dovrà avere come "cittadini di prima classe" i concetti di azione dell'impresa, risposta del cliente, evento esogeno, stato del cliente, stato dell'impresa. Questi concetti molto astratti si tradurranno di caso in caso in concetti più concreti, come i contatti promozionali o le vendite; riteniamo però che sia buona metodologia tenere presenti questi elementi di un quadro teorico della business intelligence da costruire sulla base delle teorie delle discipline manageriali ed economiche.

Le possibilità offerte dal quadro teorico delineato in precedenza sono numerose e meritevoli di indagine. La discussione precedente ha preso lo spunto dall'opzione di up sel-

ling del cliente. Si pensi a un'altra opzione di ampliamento della relazione, quella di cross selling, che abbiamo definito come aumento della varietà di prodotti acquistati, derivante da economie di scopo del cliente. Le opportunità di cross selling possono essere individuate con le tecniche di market basket analysis e delle regole di associazione, dei classici del data mining (si pensi alla proverbiale regola *pannolini* → *birra*). La considerazione di questa opzione suggerisce di introdurre nel datawarehouse le informazioni utili a scoprire queste opportunità e a muovere le opportune leve decisionali. La fact table Vendite dovrebbe permettere di effettuare la market basket analysis, se è strutturata in modo da fornire i collegamenti fra articoli venduti nella stessa transazione di vendita. Quanto alle leve decisionali, la domanda da porsi è: come facilitare al cliente le economie di scopo, aumentando nel contempo il valore per l'azienda? Non pretendiamo di dare qui una risposta, ma soltanto di mostrare come la teoria dell'impresa proattiva possa essere utilizzata per guidare la progettazione del datawarehouse, anche tenendo conto dei concetti interpretativi della business intelligence sviluppati nei capitoli precedenti.

2. Segmenti e stati del cliente

Da varie considerazioni esposte in questo capitolo come anche nei precedenti emerge l'interesse di una rappresentazione degli stati interni del cliente. Questo implica che nel datawarehouse non sarà sufficiente avere sul cliente informazioni sulle tradizionali variabili osservabili usate come basi di segmentazione, quali

- variabili geografiche
- variabili demografiche
- variabili socio-economiche
- variabili comportamentali
- ma anche variabili non osservabili quali
- personalità
- valori e stili di vita

- preferenze e benefici ricercati
- intenzioni e progetti di vita

Quando parliamo di interesse degli stati interni non ci limitiamo però a sostenere che sia desiderabile avere queste variabili nel datawarehouse, ma intendiamo che fra gli attributi del cliente dovrebbero esserci anche quelli relativi al suo profilo in qualche modello della relazione fra cliente e impresa.

Esemplificando con il modello dinamico della customer loyalty del capitolo 2, vediamo che la teoria suggerisce di assegnare un ruolo critico al concetto di customer satisfaction, in quanto questo dovrebbe essere un driver di fondamentale importanza e avere un forte potere predittivo, superiore a quello di molte variabili osservabili tradizionalmente usate.

Oggi molte imprese dispongono di soddisfacenti fonti informative riguardo al grado di soddisfazione dei clienti, sia rispetto all'impresa nel suo complesso alle singole componenti della sua offerta. Tali informazioni sono rilevate su base routinaria e quindi non richiedono costi addizionali di acquisizione.

I dati sulla customer satisfaction e la customer loyalty dovrebbero, nella misura del possibile, essere organizzati in modo da rispecchiare i costrutti teorici del modello di relazione prescelto dall'azienda, o anche di più modelli. Nel nostro esempio, dovrebbero quindi poter essere espressi in termini delle cinque dimensioni del costrutto customer loyalty, cioè le due dimensioni comportamentali di fedeltà comportamentale e comportamenti cooperativi e le tre dimensioni cognitive di fiducia, valore monadico e valore diadico. Dovrebbero poi essere temporalmente correlate alle cinque fasi canoniche del comportamento del consumatore, cioè percezione del bisogno, ricerca delle informazioni, valutazione delle alternative, decisione di acquisto, valutazione post-acquisto. Da tutto ciò dovrebbe risultare il posizionamento del cliente sul continuum relazionale a cinque fasi, cioè soddisfazione, fiducia, fedeltà comportamentale, fedeltà mentale e lealtà.

I dati, e gli algoritmi che li elaborano, dovranno allora rendere conto di concetti come valore atteso e valore percepito, e quindi di gap fra i due, e così via secondo modalità

che dipendono dalle peculiarità dell'attività dell'impresa, dei suoi clienti, del suo sistema di acquisizione e trattamento delle informazioni.

Tutto ciò può essere utile a vari scopi, non ultima la segmentazione dei clienti. Il concetto di cui qui ci occupiamo, però, ha un senso particolare nel contesto di un sistema di business intelligence che fa uso di data warehousing e data mining.

Lo stato del cliente e il segmento di clientela sono due concetti con forti analogie: come i segmenti si definiscono e si distinguono per le differenze nelle funzioni di risposta di clienti diversi, così gli stati si definiscono e si distinguono per le differenze nelle funzioni di risposta di uno stesso cliente in momenti diversi.

Possiamo allora trattare gli stati in modo analogo ai segmenti; possiamo richiedere che abbiano le stesse proprietà dei segmenti, cioè identificabilità, misurabilità, rilevanza, stabilità, accessibilità. E possiamo individuare gli stati usando le stesse variabili che si usano come basi di segmentazione.

L'analogia fra stati e segmenti non è completa, in primo luogo perché i segmenti coesistono nella clientela, mentre gli stati sono in successione temporale nel cliente. Tuttavia è utile tenere presente l'analogia nella progettazione del datawarehouse e degli algoritmi di data mining. I concetti centrali sono sempre la funzione di risposta, le azioni, l'ottimizzazione del valore atteso; segmenti, stati e altri concetti sono metodi di rappresentazione e calcolo strumentali a questi.

A nostra conoscenza, non ci sono esempi in letteratura di datawarehouse che rappresentino esplicitamente segmenti e stati dell'utente; occorre però dire che molti progetti interessanti non vengono pubblicati nella letteratura scientifica. È anche vero che la rappresentazione esplicita di insiemi può porre problemi non semplici quando le dimensioni dello schema a stella variano nel tempo, per motivi che qui non è necessario approfondire, e il segmento o lo stato dovrebbero verosimilmente avere proprio ruolo di dimensioni soggette a ridefinizioni nel tempo. Riteniamo però che segmenti e stati siano entità di estremo interesse per la customer intelligence, e che quindi valga la pena tenerle attentamente in considerazione in fase di progettazione tanto del datawarehouse quan-

to del data mining, anche se ciò comporta seri problemi tecnici di progettazione nel caso del datawarehouse.

3. Modelli multidimensionali segmentati

È noto il gradimento che il modello multidimensionale dei dati incontra presso gli analisti e i decisori aziendali per le sue caratteristiche di naturalezza, espressività e duttilità. In particolare, la possibilità di aggregare e disaggregare i dati è estremamente suggestiva e utile per le analisi e offre importanti spunti per le decisioni. Proprio riguardo a questo aspetto dei modelli multidimensionali vogliamo qui svolgere qualche considerazione di metodo.

Per concretezza, prendiamo in esame la dimensione *Tempo*, particolarmente significativa perché invariabilmente presente in ogni modello dimensionale di interesse pratico nella gestione di impresa. La progettazione la gerarchia sulla dimensione Tempo comincia in modo naturale fissando una scala “di calendario” del tipo

giorno → *mese* → *trimestre* → *anno*

o qualcosa di simile. Questa gerarchia fornisce la grana temporale dei fenomeni oggetto di analisi, in pratica fissa le possibili griglie degli ipercubi e dei report che si possono produrre in modo naturale dal sistema. Ci attendiamo allora che una cella di ipercubo o di report aggregi p.e. le vendite di un giorno, di un mese, di un trimestre o di un anno. In linea di principio niente impedisce di costruire un ipercubo sulle settimane o le quindicine, oppure su periodi irregolari definiti in un modo qualunque: se la grana minima è quella del giorno, si può sempre costruire qualunque aggregazione che *discretizza* l’asse temporale in sequenze di giorni. In pratica, una ridefinizione della discretizzazione del tempo è una operazione macchinosa per motivi tecnici e poco naturale per motivi cognitivi.

Supponiamo che le vendite di un articolo abbiano un andamento di questo tipo

Agosto	Settembre	Ottobre	Novembre	Dicembre	Gennaio	Febbraio
100	130	180	200	400	140	60

Evidentemente siamo in presenza di un articolo stagionale. Non si tratta però di un vero e proprio articolo “invernale”, perché in gennaio e febbraio le vendite precipitano. L’articolo si potrebbe definire meglio “natalizio”, tuttavia le vendite di ottobre e novembre non sono trascurabili e sono simili fra loro, quindi non si può dire che l’articolo si venda proprio solo a Natale. Poi c’è da considerare la coda di vendite in gennaio, forse per saldi. In questa situazione l’analista può essere curioso di scendere a un livello di dettaglio inferiore, ma questo dettaglio nel nostro esempio è il giorno, probabilmente una unità troppo piccola per consentire una lettura agevole. Potremmo allora decidere di introdurre nella gerarchia la settimana, in modo da poter avere prospetti su base settimanale in modo semplice (non è qui in discussione la possibilità di fare qualsiasi analisi, ma piuttosto la facilità e naturalezza). L’uso della settimana come unità di analisi non è però del tutto chiaro, dal momento che la sequenza delle settimane e quella dei mesi sono sfalsate e rispecchiano due logiche diverse; potremmo suddividere allora ciascun mese in tre decadi, ciascuna di 10-11 giorni. Tutto ciò è certamente sensato, ma si avverte la mancanza di una base teorica: la scelta dei “blocchi” di giorni avviene in modo molto empirico e un po’ confuso, con scelte ad hoc non facilmente generalizzabili. Aggiungiamo poi che nel progetto della dimensione Tempo si finisce per dover aggiungere degli attributi del tipo “giorno feriale o festivo” che finiscono per rendere molto irregolare la strutturazione dell’asse temporale, ma che sono in pratica necessari per comprendere meglio i fenomeni (c’è differenza rilevante nelle vendite dei giorni feriali e di quelli festivi?).

Proviamo ora a pensare che l’asse temporale non sia suddiviso in “blocchi” di giorni, come abbiamo detto poco prima, ma piuttosto in “segmenti” di giorni. Nel senso geometrico della parola segmento stiamo dicendo la stessa cosa, ma se intendiamo i segmento nel senso del marketing allora la riformulazione suggerisce una strada: suddividere i giorni in gruppi non in base alla contiguità temporale scandita da qualche forma di periodizzazione, ma piuttosto in base alla omogeneità dei giorni di uno stesso gruppo

e alla distinguibilità dei giorni di gruppi diversi. In pratica, si tratta di suddividere i giorni in cluster anziché in periodi.

Un cluster può anche comprendere giorni non consecutivi, accomunati non dalla contiguità temporale ma dalla similarità della funzione di risposta dei clienti o più in generale dalla *funzione di risposta del fenomeno in esame alle azioni dell'impresa*; una versione più debole consiste nel discretizzare il tempo in segmenti contigui ma di ampiezza ciascuno diversa, sempre in base al criterio di similarità della funzione di risposta.

Si pensi a una azienda di trasporti urbani che vuole rappresentare in un ipercubo i dati sul traffico cittadino. La dimensione temporale può essere vista in molti modi. Certamente un prospetto che rende conto del numero di passeggeri trasportati nei vari mesi dell'anno è interessante, ma è anche significativo suddividere l'anno in stagioni che non corrispondono a quelle del calendario ma a fasi diverse delle abitudini sociali: autunno mite, autunno inoltrato, festività, inverno, primavera incipiente, primavera avanzata, vacanze scolastiche, flussi turistici, ferie di agosto, rientro. Ancora, una suddivisione interessante è per tipologia di giorno: ha certamente senso nel trasporto urbano suddividere i giorni in 7 classi secondo la posizione nella settimana. E ancora, ha senso scendere di dettaglio al livello dell'ora e analizzare i flussi di viaggiatori per fascia oraria. Di nuovo, le fasce orarie non possono essere "tagliate" soltanto con scelte convenzionali, per esempio in tre blocchi da otto ore o sei da quattro e simili. Le fasce orarie davvero interessanti sono dettate dalle abitudini sociali di spostamento per lavoro, scuola, acquisti, spettacoli ed eventi. Quest'ultima riflessione suggerisce anche modi di suddividere il tempo dettati non dal calendario o dall'orologio, ma piuttosto dai fenomeni oggetto di analisi o da quelli che li determinano: l'autobus delle 7:30 e quello delle 16:30 sono anche gli autobus di "subito prima dell'apertura degli asili" e di "subito dopo la chiusura degli asili".

Le considerazioni precedenti valgono anche per dimensioni diverse da quella temporale, ad esempio per le dimensioni spaziali. Sempre nell'esempio dei trasporti urbani, due punti possono essere raggruppati perché prossimi in senso geometrico, oppure perché prossimi lungo una linea urbana, e le due classificazioni sono utili per analisi diverse. Altri esempi si possono trovare per altre topologie di dimensioni, seguendo la stessa logica

Le discretizzazioni delle dimensioni basate sulla funzione di risposta sono certamente significative per l'analista, ma in pratica non si può chiedergli di inventarsele ogni volta che deve effettuare un'analisi: si perde l'immediatezza d'uso, si crea sovraccarico cognitivo e per di più si mette in difficoltà il sistema per il calcolo continuo di nuove viste multidimensionali.

È invece desiderabile che la struttura delle gerarchie dimensionali sia proposta dal sistema stesso in base ai risultati dell'analisi dei dati, vale a dire che il data mining venga estensivamente usato proprio nella fase di progettazione delle viste nel datawarehouse. In questo approccio progettuale il disegno del datawarehouse viene effettuato con il concorso del data mining. Apparentemente ciò contraddice un principio metodologico fondamentale: l'indipendenza dei dati dalle applicazioni. In realtà la dipendenza che si introduce è limitata alle modalità di aggregazione dei dati, che nel loro massimo dettaglio restano comuni a tutte le applicazioni e da esse indipendenti. Possiamo dire che il data mining serve a progettare viste multidimensionali di alto valore, senza compromettere la generalità del datawarehouse.

4. Attributi non osservabili

Se si intraprende la fondazione dei sistemi di business intelligence su modelli teorici tratti dalle discipline economiche e manageriali, allora si arriva in modo naturale a pensare di inserire come dati anche degli elementi non osservabili presenti nel modello. Si può quindi pensare a un attributo *lealtà* che non è certo direttamente osservabile ma che ha senso ed è importante in un certo modello della relazione fra cliente e impresa.

Un valore non osservabile specificato in un modello deve essere determinato in base a valori osservabili con un algoritmo: quindi in un datawarehouse il valore di un attributo non osservabile deve essere correlato usando *metadati* a valori osservabili esplicitamente rappresentati e a un algoritmo. Con la rappresentazione esplicita dei valori osservabili di partenza ci si riserva di ricalcolare in futuro il valore non osservabile quando fossero disponibili altri dati o altri algoritmi, o forse anche un modello teorico diverso.

Riconsideriamo il modello presentato nel capitolo 2, che esprime il customer equity in funzione di 3 driver (value equity, brand equity, relationship equity) e questi in funzione di 11 subdriver: qualità, prezzo, convenienza, consapevolezza della marca, attitudine verso la marca, etica dell'impresa, programmi di fidelizzazione e costruzione della lealtà, affinità, costruzione di comunità, costruzione di conoscenza

$$CE = CE(VE, BE, RE)$$

$$VE = VE(Qua, Pri, Con)$$

$$BE = BE(Awe, Att, Eth)$$

$$RE = RE(Loy, Spe, Aff, Com, Kno)$$

Dovremo inserire nei metadati i descrittori degli algoritmi che dai subdriver risalgono ai driver e quindi al customer equity. Supponendo che si tratti di algoritmi di analisi delle componenti principali e regressioni logistiche, dovremo registrare i modelli usati, le ipotesi di partenza, i protocolli d'uso.

Poi dovremo inserire nel datawarehouse i valori delle grandezze non osservabili in gioco, 15 nel nostro esempio, e correlarle opportunamente fra loro e con gli algoritmi.

Infine dovremo inserire nel datawarehouse i dati osservabili di partenza. Questi verosimilmente saranno dei questionari compilati dai clienti. In un contesto di marketing *one-to-one* è possibile che ad ogni cliente sia somministrato il questionario; in altri contesti invece solo un campione di clienti avrà risposto al questionario. Nel secondo caso i dati del questionario saranno attribuiti ai segmenti di clienti, anziché ai singoli clienti (riproponendo le considerazioni già svolte sulle implicazioni dell'aggregazione dei dati come euristica).

Esemplifichiamo come le risposte di un questionario possano essere usate nel modello di customer equity presentato nel capitolo 2. Supponiamo di voler modellare il customer equity dei clienti di una compagnia di trasporto aereo.

Per costruire le probabilità di transizione da una compagnia all'altra (che entreranno nella switching matrix) chiediamo al cliente:

con quale compagnia ha fatto il suo ultimo volo;

con quali compagnie potrebbe fare il suo prossimo volo, assegnando a ognuna una probabilità.

Queste due domande non permettono di costruire la switching matrix, perché le risposte sono relative al momento dell'intervista e perché il cliente al momento si trova in uno stato e quindi può parlare solo delle probabilità di transizione a partire da quello stato (una riga della switching matrix). Però queste risposte contribuiscono evidentemente alla determinazione della switching matrix, unitamente ad altri dati dello stesso cliente e ai dati analoghi di altri clienti. L'ideale sarebbe farsi dettare la switching matrix dal cliente stesso, ma naturalmente questo non è realistico.

Poi chiederemo al cliente quali sono le sue frequenze di volo e quale è l'importo medio dei suoi biglietti. Le risposte servono a calcolare il suo customer equity una volta che sia nota la switching matrix.

Quindi dovremo fare al cliente delle domande sui driver del value equity, che sono la qualità, il prezzo e la convenienza. Chiederemo quindi al cliente di valutare il grado di qualità sia della nostra che delle compagnie concorrenti prese a benchmark. Quindi gli chiederemo come vede il rapporto fra prezzo pagato per il servizio e valore che egli attribuisce al servizio stesso, ancora sia riguardo a noi che ai nostri concorrenti. Una o più altre domande riguarderanno la convenienza, quindi la facilità di prenotazione, la disponibilità delle tratte desiderate e simili.

Le risposte a queste domande sono valori di attributi osservabili. Applicando gli opportuni protocolli statistici si arriverà a quantificare dei valori per i subdriver qualità, prezzo e convenienza e per il driver value equity. Analogamente si procederà per brand equity e relationship equity.

Capitolo 6

Autonomic customer intelligence

In questo capitolo descriviamo l'esperienza di un programma di ricerca industriale (denominato AKIM, per *Autonomic Knowledge Discovery for Interactive Marketing*) che mira a realizzare un sistema innovativo di customer intelligence. La peculiarità del sistema è la parziale automazione del processo di knowledge discovery, e più specificamente della fase di data mining. Il campo applicativo designato è il marketing one-to-one, sia nella forma del classico direct marketing per via postale e telefonica, sia nella forma del web marketing.

1. Obiettivi e motivazioni del progetto

L'esigenza di automatizzare almeno in parte il processo di knowledge discovery è motivata in parte dai costi elevati del personale specializzato, ma ancora più dall'esigenza di rendere continuo e rapido il processo di analisi e decisione. Nei settori del direct e del web marketing, che qui ci interessano, l'ottimizzazione delle campagne deve essere effettuata in tempo reale, sia esso dell'ordine dei decimi di secondo o dei giorni. In pratica, nelle organizzazioni è estremamente difficile, quasi impossibile, garantire la necessaria ininterrotta disponibilità del processo. Di conseguenza, si finisce in genere per usare le tecniche di knowledge discovery solo nella fase di pianificazione delle campagne e nella fase di controllo finale, o al più periodico, dei risultati.

Il programma di ricerca industriale qui descritto¹ si concentra su una fase del processo di knowledge discovery, il data mining inteso come applicazione di algoritmi a dataset

¹ Il programma di ricerca è condotto dalla Inventa S.r.l. di Livorno, in collaborazione con varie altre imprese ed enti di ricerca, usufruendo di parziali finanziamenti del Ministero dell'Istruzione, dell'Università e della Ricerca.

predisposti, senza rinunciare a indagini e sperimentazioni anche nella fase di preparazione dei dati.

La letteratura scientifica sull'automazione del knowledge discovery è in realtà assai limitata. La nostra interpretazione di questo fatto è che il problema appaia ad oggi prematuro e che nell'agenda della comunità della ricerca ad esso debbano essere anteposte altre problematiche, poiché le applicazioni di data mining sono ancora oggi in larga misura progetti nei quali la competenza e l'ingegnosità degli analisti hanno larghissimo spazio. Tuttavia, ci sembra che se il problema viene circoscritto opportunamente a certi aspetti, specifici ma di grande momento applicativo, esso possa trovare nel medio termine soluzioni utili e di complessità abbordabile.

L'approccio più convenzionale all'automazione del knowledge discovery è quello di [Livingston 2000, 2001a, 2001b].

Il processo è organizzato in *task* implementati da algoritmi predefiniti. Una *agenda* assegna ai task una *priorità* in base alla loro *plausibilità*, la quale è funzione della forza delle *motivazioni* per eseguire un task e dell'*interesse* degli *item* sui quali opera il task. Un item è un elemento dello spazio di ricerca, come una regola o un sottoinsieme dei dati. Le motivazioni sono una misura della coerenza e della rilevanza del task per il problema da risolvere. Un task viene eseguito usando delle *euristiche* e crea nuovi item e nuovi task che entrano in agenda, corredati da motivazioni, e anche nuove motivazioni per i task già in agenda.

Il funzionamento ad alto livello del sistema è il seguente (riportato da [Livingston 2001a]).

0. *Initialize the agenda*

1. *Compute the plausibilities of the tasks on the agenda using the strengths of the reasons given for performing them and the interestingness of the items involved in the tasks.*

2. *Select and remove from the agenda the task with the greatest plausibility.*

3. *Perform the task using heuristics, involving one or more of:*

· *Proposing new tasks*

- *Evaluating items*
 - *Creating new items*
 - *Examining relationships among the items*
4. *Check the stopping criterion, and if it is met, stop the discovery process.*

Otherwise, loop back to Step 1.

Per quanto interessante, questo lavoro non sembra avere avuto adeguati sviluppi, come dicevamo. Il programma di ricerca industriale al quale ci riferiamo ha tratto utili insegnamenti tecnici e metodologici da questo lavoro, ma ha intrapreso una strada alternativa. Il principio per così dire filosofico del programma è di lasciare spazio all'emergere di automatismi nel processo tramite evoluzione e apprendimento. Quindi il processo di data mining dovrebbe essere per quanto possibile *autoregolato*, senza seguire percorsi composti in modo esplicito. In realtà le due modalità, pianificata ed evolutiva, hanno ciascuna punti di forza e limiti e finiscono per essere usate entrambe. La peculiarità e l'originalità del programma di ricerca AKIM dal punto di vista della metodologia stanno comunque soprattutto nel processo evolutivo.

La metodologia potrebbe essere applicata a molti tipi di problemi, ma il programma AKIM è nato per applicazioni nel campo del marketing, come suggerito dal nome stesso. Il tipo di contesto applicativo di un sistema ispirato ad AKIM si può descrivere così: si vuole condurre una campagna di telemarketing, direct marketing o web marketing; i clienti (o i prospect) sono descritti in un dataset, e si suppone che fra i loro attributi abbiano un ruolo rilevante quelli comportamentali, cioè quelli che descrivono i contatti fra i clienti e l'impresa; i clienti sono molto numerosi, comunque tanto numerosi da giustificare che la campagna si estenda nel tempo e che sia importante modificare in itinere la sua pianificazione.

Si vuole condurre una campagna di telemarketing, direct marketing o web marketing. I clienti (o i prospect) sono descritti in un dataset, e si suppone che fra i loro attributi abbiano un ruolo rilevante quelli comportamentali, cioè quelli che descrivono i contatti fra

i clienti e l'impresa.. I clienti sono molto numerosi, comunque tanto numerosi da giustificare che la campagna si estenda nel tempo e che sia importante modificare in itinere la sua pianificazione.

Rifacendosi a un caso reale, si consideri una società di credito al consumo (in sostanza una banca) che attua una campagna di telemarketing diretta a un target di milioni di clienti e consistente nel proporre prodotti finanziari (polizze assicurative, prestiti personali). I clienti sono già ben noti e descritti in un grande database, dal quale viene estratto un dataset adatto all'elaborazione con strumenti di data mining. È possibile che dal database dei clienti vengano estratti più dataset strutturati in modi diversi per renderli trattabili da algoritmi di tipo diverso; in ogni caso i dataset sono congruenti nei contenuti.

L'operatrice di call center è collegata a un computer che riceve da un server i riferimenti ai clienti da contattare. L'operatrice viene automaticamente collegata al cliente e in contemporanea vede comparire sul monitor i dati del cliente stesso. Propone quindi i prodotti suggeriti dal sistema. L'offerta contiene un numero limitato di proposte, perché la proposizione di troppe alternative indurrebbe confusione e incertezza nel cliente e sarebbe controproducente,

L'operatrice registra quindi l'esito della conversazione. Gli esiti sono di due tipi:

- esiti commerciali: il cliente acquista, rifiuta, rinvia, chiede un nuovo appuntamento telefonico;
- esiti telefonici: il numero non risponde, dà occupato, non esiste.

Gli esiti vanno ad arricchire il database dei clienti e quindi anche il dataset a disposizione del sistema. Durante la conversazione l'operatrice pone anche domande al cliente e con le risposte aggiorna i suoi dati.

Ogni contatto telefonico può durare molti minuti, anche decine, e quindi ha un costo non irrilevante, di vari euro.

La campagna non può essere "a tappeto" su tutti i clienti, per due motivi. Il primo è che non tutti i clienti sono abbastanza promettenti da giustificare il costo del contatto; si noti che quali siano i clienti promettenti e quali no non è noto a priori. Il secondo motivo è

che, poiché una operatrice di call center può effettuare un numero limitato di chiamate, una campagna totale richiederebbe mesi o anni, e la durata della campagna supererebbe sia il budget sia il ciclo di vita del prodotto, che diventerebbe obsoleto in itinere. In pratica, i vincoli economico-finanziari e temporali impongono che solo una frazione piuttosto piccola dei clienti sia effettivamente raggiunta dalla campagna.

L'impresa deve quindi ottimizzare la scelta del target, tentando di selezionare la frazione di clienti migliori e di proporre a ciascuno il prodotto più verosimilmente per lui interessante, oppure un numero ristretto di prodotti.

La selezione dei clienti e quella dei prodotti a rigore sono due aspetti di uno stesso problema e dovrebbero essere effettuate simultaneamente. Posto in questa forma, il problema è un problema di segmentazione nel senso del capitolo 3, con un numero non prefissato di segmenti di clienti e uno spazio delle decisioni che nel caso reale era fortunatamente di piccole dimensioni (la gamma dei prodotti polizze e prestiti era solo di alcune unità), ma che in altri casi potrebbe essere invece molto grande.

In casi di questo genere non è pensabile una soluzione per ricerca esaustiva, e non è certo facile elaborare valide euristiche per abbattere la complessità computazionale. In questo contesto AKIM interviene a supportare la ri-pianificazione in tempo reale della campagna, cercando di fare progressivamente emergere mediante apprendimento evolutivo delle buone strategie di targeting, che dopo un certo numero di iterazioni riescano a concentrarsi su clienti promettenti con offerte ben centrate.

Un aspetto interessante di questa specifica applicazione è che la campagna dovrebbe ottimizzare due misure:

1. *l'efficacia* intesa come capacità di raggiungere buoni esiti commerciali, vendite innanzitutto, ma anche appuntamenti e interviste;
2. *l'efficienza* intesa come capacità di non sprecare risorse di tempo e denaro e quindi di raggiungere buoni esiti telefonici, cioè conversazioni effettive e non chiamate a vuoto per libero o occupato.

Un esempio reale di regola che migliora l'efficacia è “fra i clienti in questa area geografica gli uomini hanno una spiccata propensione per la polizza auto e le donne per la polizza casa”.

Un esempio reale di regola che migliora l'efficienza è “in questa area geografica i maschi singoli fra 30 e 40 anni rispondono al telefono con frequenza massima il sabato mattina fra le 10 e le 12”.

Queste due regole hanno una validità stabile nel tempo, almeno in orizzonti temporali rilevanti per la campagna. Una regola che migliora l'efficienza e che invece ha validità molto breve è “da un'ora la frequenza di risposta dei clienti in questa città è drasticamente scesa”. Questa regola è utilizzabile soltanto se trovata e implementata in tempo reale, a dimostrazione di caso in cui il data mining in tempo reale porta un concreto e inaspettato, seppur effimero, valore aggiunto. Si noti che la regola è puramente descrittiva, senza un fondamento teorico: la spiegazione può essere che in quella città è appena spuntato il sole dopo la pioggia, ma non è necessario che il sistema tenga conto di variabili meteorologiche. Il knowledge discovery autonomico ambisce fra l'altro a sfruttare questa forma di conoscenza accidentale ed effimera, che non è realistico pensare di introdurre in modelli formali ma che può emergere dall'apprendimento automatico in tempo reale.

2. Modello computazionale

Esponiamo ora il modello di funzionamento di AKIM nella sua forma più semplice ed essenziale.

Un *predittore* è un algoritmo che elabora un campione del dataset di clienti e crea una *ipotesi*. Una ipotesi contiene una regola per selezionare i clienti più promettenti. La forma concreta di una ipotesi può essere una query, un algoritmo di ranking (che induce un ordinamento parziale) o di scoring (che induce un ordinamento totale). Congiuntamente alla regola di selezione dei clienti ce n'è un'altra per la selezione del catalogo di prodotti da proporre a ciascun cliente selezionato. Nel quadro teorico del capitolo 3, una ipotesi è un generatore di soluzioni candidate di un segmentation problem.

La *sperimentazione* di una ipotesi consiste nell'usarla per guidare una parte di una campagna. Nell'esempio del telemarketing, si usa l'ipotesi per generare una serie di coppie

cliente-prodotto, ogni cliente viene chiamato automaticamente e “passato in cuffia” all’operatrice che sul monitor vede i dati del cliente e anche il prodotto da proporgli.

La sperimentazione viene applicata a un *campione* che sia:

1. abbastanza grande da essere statisticamente significativo e consentire sia test statistici affidabili che efficace apprendimento automatico;
2. abbastanza piccolo da non sprecare troppo tempo su ipotesi inefficaci e da non “bruciare” troppi clienti con approcci sbagliati.

Trovare il *trade-off* tra le due esigenze e quindi la giusta dimensione è di per sé un compito complesso, che ricorda le tecniche di pianificazione degli esperimenti nella statistica classica.

Sulla base degli esiti della sperimentazione si determina la *performance* del predittore. Il concetto di performance può essere secondo le applicazioni specifiche di tipo economico (profitti delle vendite, costi dei contatti) o conoscitivo (conferma che l’ipotesi formulata dal predittore era giusta).

Al predittore è dunque associata una performance relativa a quella sperimentazione. A questo punto entra in gioco il processo di *apprendimento* guidato da *evoluzione competitiva*.

Nel sistema coesistono coesistono vari tipi di predittori: algoritmi di regressione, classificatori bayesiani, reti neurali e così via. Per ogni tipo di predittore possono essere contemporaneamente presenti varie istanze, che si distinguono per la scelta di certi parametri, come il vettore di pesi da attribuire agli attributi del cliente oppure al matrice di incidenza degli archi sui nodi di una rete.

Un motore di evoluzione competitiva si comporta secondo la logica della *programmazione evolutiva* e degli *algoritmi genetici*:

3. sperimenta un certo insieme di predittori (intesi come istanze);
 - registra la performance di ciascuno nel risolvere il problema di formulare ipotesi corrette sul comportamento dei clienti;
 - con l’operatore di *selezione* sceglie i predittori con la migliore performance
 - con gli operatori di incrocio e mutazione genera una nuova popolazione di predittori.

Iterando il ciclo c'è la possibilità che le popolazioni di predittori siano sempre più abili nel risolvere il problema di segmentazione. Come per gli algoritmi genetici, la convergenza verso soluzioni ottimali o quasi ottimali non è affatto garantita, problema che AKIM condivide con tutti i modelli ispirati a queste logiche di ottimizzazione.

Questo modello è stato sperimentato dapprima in modo artigianale e “manuale” nel contesto della campagna di telemarketing di cui si è detto. In quel caso il ruolo del motore di apprendimento era svolto da esseri umani, per indagare le sue implicazioni teoriche, implementative e applicative.

Essendo i risultati promettenti, in un progetto di ricerca parzialmente finanziato si è realizzato un sistema prototipale che poi è stato sperimentato su dataset sintetici creati con appositi generatori. I risultati hanno mostrato una incoraggiante capacità di AKIM di apprendere pattern nascosti nei dati, introdotti volutamente dagli sperimentatori. L'interesse della sperimentazione non stava tanto nella raffinatezza o originalità delle scoperte, pure alcune non banali, quanto nel processo autonomo che riusciva a costruire predittori più abili in modo non supervisionato.

L'uso di dataset sintetici anziché reali è stato imposto dalla delicatezza di una campagna di telemarketing, poiché un comportamento errato o anomalo del sistema avrebbe potuto provocare danni economici enormi. È ora in corso un ulteriore progetto di ricerca industriale per raffinare il sistema e applicarlo a reali campagne di direct mailing, nelle quali i rischi della sperimentazione sono invece controllabili.

I metodi usati per realizzare il processo di apprendimento sono molto interessanti sotto il profilo scientifico e tecnologico ma che non riguardano direttamente questo lavoro.

Invece è interessante notare che la valutazione della performance in AKIM segue un approccio microeconomico come quello descritto nel capitolo 4 e che l'ottimizzazione dei predittori ha anche degli aspetti simili a quelli descritti nello stesso capitolo 4 discutendo dei modi per ottimizzare il processo analitico-decisionale. La performance richie-

sta a un predittore usato in campagne di vendita telefonica dipende dalla sua utilità in senso economico più che dalla sua correttezza conoscitiva, o per meglio dire la correttezza è strumentale alla utilità, che costituisce la vera funzione obiettivo da ottimizzare. Un aspetto interessante anche sotto il profilo teorico è che la funzione basilare di A-KIM non è la segmentazione in senso classico, perché non opera solo sui clienti, e non è neanche la *recommendation* in senso classico, perché non sceglie il prodotto migliore per un cliente assegnato. Invece questa funzione è l'esplorazione dello spazio di ricerca delle decisioni, intese come accoppiamenti di cataloghi a segmenti di clienti, per ottimizzare l'utilità mediante euristiche non predefinite ma apprese autonomamente.

Bibliografia

Adamson C., Venerable M. [1998]

Data Warehouse Design Solutions

John Wiley & Sons

Alon N., Sudakov B. [1999]

On Two Segmentation Problems

Journal of Algorithms 33:173--184

Amram M., Kulatilaka N. [2000]

Real options. Strategie di investimento in un mondo dominato dall'incertezza

Etas ed.

Barney J.B. [1991]

Firm Resources and Sustained Competitive Advantage

Journal of Management 17: 99-120

Blattberg R., Deighton J. [1996]

Manage Marketing by Customer Equity Test

Harvard Business Review July-August

Blattberg R., Getz G., Jacquelyn Thomas J. [2001]

Customer Equity

Harvard Business Press

Costabile M. [2001]

Il capitale relazionale

Mc Graw-Hill ed.

Costabile M. [2003]

Il capitale relazionale negli ambienti digitali
in W.G.Scott (a cura di) *Il management dell'impresa nell'economia digitale*
Il Sole 24 Ore ed.

Ford D. [1980]
The Development of Buyer-Seller Relationships
European Journal of Marketing, vol. 14

Ford D., Gadde L., Hakansson H., Lundgren A., Snehota I., Turnbull P., Wilson D.
[1998]
Managing Business Relationships
Wiley ed.

Han J.W., Kamber M. [2001]
Data Mining: Concepts and Techniques
Morgan Kaufmann Publishers

Kimball R., Ross M. [2002]
The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling
John Wiley & Sons ed.

Kleinberg J., Papadimitriou C., Raghavan P. [1998]
A Microeconomic View of Data Mining
Journal of Data Mining and Knowledge Discovery 2, 4, 311–324.

Kleinberg J., Papadimitriou C., Raghavan P. [2004]
Segmentation Problems
Journal of the ACM, 51: 2, 263-280

Livingston, G.R. [2000]
A Framework for Autonomously Performing Knowledge Discovery in Databases
PhD Thesis, University of Pittsburgh

Livingston, G. R., Rosenberg, J. M., Buchanan, B. G. [2001a]

Closing the Loop: An Agenda- and Justification-Based Framework for Selecting the Next Discovery Task to Perform

Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose, CA.
IEEE Computer Society Press.

Livingston, G. R., Rosenberg, J. M., Buchanan, B. G. [2001b]

Closing the Loop: Heuristics for Autonomous Discovery

Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose, CA.
IEEE Computer Society Press.

Marzo G., Turatto R. [2003]

Alle radici del valore

TPoint, vol. 3/2003

Pels J. [1992]

Identification and Management of Key Clients

European Journal of Marketing, 5: 5-20

Peteraf M. [1993]

The Cornerstones of Competitive Advantage: A Resource-Based View

Strategic Management Journal 14: 179-191.

Rust R., Lemon K., Zeithaml V. [2001]

Driving Customer Equity: Linking Customer Lifetime Value to Strategic marketing Decisions

Report 01-108 Marketing Science Institute Working Paper Series

Valdani E. [2000]

L'impresa pro-attiva

McGraw-Hill ed.

Wayland R., Cole P. [1997]

Customer Connections

Harvard Business School Press

Wilson D. [1995]

An Integrated Model of Buyer-Seller Relationship

Journal of the Academy of Marketing Science, vol. 4