



Il Prof. Hal R. Varian durante la Ottava Lezione «Angelo Costa» tenuta presso l'Università LUISS Guido Carli di Roma il 16 gennaio 2007.

L'economia politica dei motori di ricerca di *Internet*

Hal R. Varian*

University of California at Berkeley

Questa Lezione fornisce una introduzione all'aspetto economico dei motori di ricerca di Internet. Dopo un breve esame dello sviluppo storico della tecnologia e del settore, vengono descritte alcune delle caratteristiche economiche del sistema di aste adottato per presentare gli annunci pubblicitari. È emerso che alcuni modelli economici, relativamente semplici, consentono di comprendere a fondo il funzionamento di queste aste. In particolare, la teoria classica dell'abbinamento bilaterale dei mercati si rivela molto utile in questo contesto. [JEL Classification: L86, D83]

1. - Introduzione

I motori di ricerca sono una delle applicazioni di Internet più largamente usate. Secondo Fallows (2005) «I motori di ricerca sono molto amati dagli utenti di Internet. La ricerca in Internet è una delle prime attività che la gente prova quando inizia ad usare Internet, e la maggior parte degli utenti si sente presto a suo agio nell'atto della ricerca». Il rapporto 2005 indica che l'84% degli utenti di Internet hanno usato i motori di ricerca e, in un dato giorno, il 56% di quelli che si collegano on line usano un motore di ricerca.

I motori di ricerca non solo sono ampiamente usati, ma sono altamente redditizi. La loro principale fonte di ricavi deriva dagli annunci pubblicitari che sono connessi alle interrogazioni di

* <hal@ISchool.Berkeley.EDU>.

ricerca. Dal momento che gli utenti tendono a considerare questi spot molto pertinenti ai loro interessi, gli agenti pubblicitari pagano bene per la loro collocazione. Dato che i costi marginali sono molto bassi per i motori di ricerca, i margini di profitto tendono ad essere elevati.

La pubblicità *on line* è, per la sua stessa natura, un'attività su larga scala. Il tasso di accesso ad un buon annuncio potrebbe essere del 3%, e un normale tasso di conversione (acquisto) potrebbe essere anche esso attorno al 3%. Ciò comporta che meno dell'un per mille delle persone che vedono lo spot comprano effettivamente il prodotto pubblicizzato. Malgrado questo rendimento apparentemente basso, gli spot dei motori di ricerca sono una delle più efficaci forme di pubblicità. Gli spot in TV o le inserzioni sui giornali sono significativamente meno efficaci, dato che una frazione molto più piccola di quelli che vedono una pubblicità compra effettivamente il prodotto reclamizzato.

Dato che la probabilità dell'acquisto è bassa, perfino quando gli spot sono pertinenti, occorre raggiungere un pubblico vasto per avere una qualche speranza di vendere un prodotto. Di conseguenza, nuovi motori di ricerca che sperano di ottenere il successo economico devono sostenere elevati costi fissi per costruire la dimensione necessaria per poter presentare un numero di spot sufficiente per coprire tali costi di ingresso.

Sul lato della domanda, i costi del passaggio ad un altro motore di ricerca sono molto bassi, la concorrenza è ad un solo click di distanza. Fallows (2005) indica che il 56% degli utilizzatori dei motori di ricerca usa più di un motore di ricerca. Pertanto, si può prevedere una forte competizione per conquistare gli utenti tra i principali motori di ricerca.

Gli utenti non sono i soli a non essere legati esclusivamente ad un solo motore di ricerca; la cosa vale anche per gli agenti pubblicitari. Questi, di norma, seguono la corrente e fanno pubblicità ovunque ci siano sufficienti clienti potenziali tali da garantire l'investimento nel settore.

Queste caratteristiche — alti costi fissi, bassi costi marginali, la necessità di avere un mercato di massa, bassi costi di passaggio ad altro motore, e un modello commerciale supportato dagli

agenti pubblicitari — significano che la probabile struttura del mercato avrà pochi grossi concorrenti in un dato paese o gruppo linguistico.

Una struttura di mercato in equilibrio potrebbe essere simile a quella dei giornali o delle riviste nazionali; alcuni grandi fornitori, supportati principalmente dalle pubblicità con una continua competizione per conquistare nuovi lettori. Non ci sono effetti di rete significativi o economie di scala dal lato della domanda, tali da spingere il mercato verso un singolo fornitore.

Sosterrò più avanti che il processo economico più importante in grado di determinare il successo dei motori di ricerca consiste «nell'imparare facendo esperienza diretta» (Arrow, 1962). Considerati i bassi costi di passaggio ad un altro motore per l'utente, i motori di ricerca devono investire continuamente nel migliorare sia la loro ricerca che i loro ricavi. Questo aspetto può valere per quasi ogni prodotto, tuttavia il miglioramento continuo è particolarmente importante per i prodotti *on line* dato che i tempi della sperimentazione e della realizzazione sono particolarmente rapidi.

Pur essendo disponibili dozzine di motori di ricerca, i tre maggiori, in termini di quota di mercato, sono Google, Yahoo e MSN. Tratterò soprattutto di Google, dato che ho una grande familiarità con i suoi sistemi, ma gli altri motori di ricerca tendono ad usare modelli commerciali simili.

2. - Abbinamento bilaterale

Innanzitutto, cosa fa Google? La risposta, a mio parere, è che Google è uno “yenta”, vale a dire un termine *yiddish* tradizionale per “abbinatore”. Dal lato della ricerca, abbina le persone che cercano informazioni con quelle che forniscono informazioni. Dal lato dello spot, abbina le persone che vogliono comprare con quelle che intendono vendere.

Dal punto di vista economico, Google mette in atto un meccanismo di “abbinamento bilaterale”. Questo tema ha una lunga storia nella scienza economica, a partire dal classico problema di

attribuzione lineare, che cerca di individuare l'abbinamento tra controparti che massimizzi una qualche funzione di valore. Non sorprende quindi che la teoria matematica del problema di attribuzione appaia strettamente correlata all'asta degli spot di Google.

La necessità di un efficiente abbinamento degli utenti e dei contenuti è evidente, la crescita dei contenuti in Internet è stata fenomenale. Secondo *netcraft.com* ci sono circa 100 milioni di *web server*. Ovviamente, maggiori sono i contenuti sul web, e più è importante avere dei buoni motori di ricerca. Il web senza motori di ricerca sarebbe come la biblioteca universale di Borges senza un catalogo a schede.

In questo saggio, parlerò brevemente della storia dei sistemi di reperimento delle informazioni (*information retrieval*), mettendo in risalto alcuni dei punti di interesse dal punto di vista dell'economia. Quindi descriverò l'evoluzione del modello commerciale per supportare i motori di ricerca on line, e concluderò delineando alcuni degli aspetti economici delle aste degli spot di Google.

3. - *Information Retrieval* - Una breve storia

Subito dopo l'inizio dell'archiviazione dei testi nelle memorie dei computer, i ricercatori iniziarono a cercare di stabilire in che modo potessero essere agevolmente consultati. Un progresso significativo avvenne negli anni '60, e prima della fine degli anni '70 i sistemi operativi erano ampiamente disponibili. La disciplina raggiunse la maturazione negli anni '90, in cui gli utenti principali erano bibliotecari professionali e ricercatori¹.

Nei primi anni '90, la gran parte dei problemi più semplici era stata risolta, e gli utenti più impegnati nel campo della tecnologia dell'*information retrieval* temevano che il progresso tecnologico stesse per esaurirsi. Questa preoccupazione portò alla creazione nel 1992 della TREC (Conferenza sul Reperimento e l'Estrazione di testi) da parte di DARPA.

¹ Vedasi LESK M. (1995).

DARPA predispose dei dati preliminari consistenti in molte interrogazioni e molti documenti, assieme ad un indicatore 0-1 che indicava se il documento fosse o meno pertinente all'interrogazione. Questi indicatori di pertinenza erano stabiliti da esperti umani. I gruppi di ricerca quindi predisposero i loro sistemi sui dati della TREC. Successivamente, la TREC fornì un secondo insieme di dati per i quali i gruppi di ricerca cercarono di prevedere la pertinenza utilizzando i sistemi da loro predisposti.

Quindi TREC fornì una raccolta di test e il luogo di incontro per lo scambio delle idee, e molti dei gruppi che lavoravano nel campo dell'*information retrieval* parteciparono alla TREC. (Ved. TREC8, 2000). Il fatto di avere una base standard per confrontare diversi algoritmi era molto utile nel valutare i diversi criteri di approccio al lavoro.

I motori di ricerca usano diverse tecniche ma una di esse, che risulterà molto familiare agli economisti, è la regressione logistica. Si scelgono le caratteristiche del documento e dell'interrogazione, e quindi si cerca di prevedere la probabilità della pertinenza, mediante la regressione logistica semplice. Come esempio di questo metodo, Cooper *et al.* (1993). Cooper *et al.* (1994) usano le seguenti variabili:

- numero dei termini in comune tra il documento e l'interrogazione.
- logaritmo della frequenza assoluta delle presenze di un termine dell'interrogazione nel documento, mediata su tutti i termini che sono presenti sia nell'interrogazione che nel documento.
- radice quadrata della lunghezza dell'interrogazione.
- frequenza della presenza di un termine dell'interrogazione nella raccolta.
- radice quadrata dell'ampiezza della raccolta.
- frequenza inversa della raccolta, che costituisce la misura di quanto sia raro un termine nella raccolta.

Altri sistemi impiegano variabili differenti e forme differenti di previsione della pertinenza, ma questo elenco è rappresentativo.

A metà degli anni '90 si percepiva diffusamente che la ricerca era divenuta un bene economico. Esistevano diversi algoritmi

che avevano un rendimento più o meno simile, e i miglioramenti tendevano a sommarsi tra loro.

Con l'avvento del *web* nel 1995, l'esigenza di migliori motori di ricerca su Internet divenne evidente, e molti degli algoritmi sviluppati dalla comunità della TREC furono usati allo scopo di soddisfare questa esigenza. Tuttavia, il difficile compito di indicizzare il *web* non era così impellente per la comunità dell'IR come si sarebbe potuto pensare. Il problema era che il Web non era la TREC. La TREC aveva avuto tanto successo nel definire il problema dell'*information retrieval*, che la massima attenzione si concentrava su quel particolare compito della ricerca, ad esclusione di altre applicazioni.

Gli informatici, d'altra parte, consideravano il *web* come il problema del giorno. Il progetto della *Digital Library NSF*, e altre simili iniziative fornivano fondi per la ricerca sul reperimento delle informazioni su larga scala.

Il Dipartimento informatico di Stanford ricevette uno di questi sussidi della *Digital Library*, e due suoi studenti, Larry Page e Sergey Brin, si interessarono al problema della ricerca nel *web*. Essi svilupparono l'algoritmo *PageRank* — un metodo di reperimento delle informazioni che utilizzava la struttura dei link del *web*. L'idea di base (per semplificare alquanto) era che i siti che avevano molti link in siti importanti che ad essi indirizzavano² avevano una certa probabilità di contenere informazioni pertinenti.

PageRank costituiva un grosso miglioramento rispetto agli algoritmi esistenti, e Page e Brin abbandonarono la scuola nel 1998 per costruire un motore di ricerca commerciale, Google.

L'algoritmo che Google usa ora per la ricerca è brevettato, naturalmente. Inoltre è molto complesso. Il concetto di base combina il risultato di *PageRank* con un risultato dell'*information retrieval*. Il vero segreto del successo di Google consiste nella costante sperimentazione dell'algoritmo, che viene rettificato, messo a punto e ritoccato praticamente senza sosta.

² Vedasi LANGVILLE A.N. - MEYER C.D. (2006) per una descrizione dettagliata matematica sottostante a *PageRank*.

Uno dei principi del metodo del controllo qualità giapponese è il *kaizen*, che viene di norma tradotto come “miglioramento continuo”. Un motivo della rapidità del progresso tecnologico sul *web* è dato dal fatto che è molto facile sperimentare, usare un nuovo algoritmo di ricerca per una interrogazione su mille. Se il nuovo algoritmo supera in rendimento quello vecchio, può essere rapidamente messo all’opera. Mediante questo tipo di semplice sperimentazione, Google ha sofisticato il suo motore di ricerca nel corso degli anni per offrire un prodotto altamente avanzato con molte funzioni specializzate.

Google non è certamente l’unico servizio *on line* che impiega il metodo *kaizen*; Amazon, eBay, Yahoo e altri migliorano costantemente i loro siti *web*. Tali miglioramenti si basano di norma sulla sperimentazione sistematica e sull’analisi statistica, come nella tradizionale pratica del controllo della qualità.

4. - Sviluppo di un Modello Commerciale

Brin e Page, all’epoca del lancio di Google, non avevano in mente un modello commerciale. Ad un certo punto essi offrirono in vendita l’algoritmo *PageRank* da loro usato a Yahoo per 1 milione di dollari. Quando Yahoo rifiutò, essi pensarono di vendere dei servizi di ricerca intranet alle imprese.

Nel frattempo, una società di Pasadena, denominata *GoTo.com* stava mettendo all’asta i risultati di ricerca. Nel 1999 presentarono la domanda di brevetto USA 6.296.361 (accolta il 31 luglio 2001) che descriveva l’idea di mettere all’asta i risultati di ricerca³.

L’asta dei risultati della ricerca non andò molto bene, dato che la propensione a pagare per il collocamento non costituisce una indicazione molto efficace della pertinenza per gli utenti, cosicché *GoTo* alla fine adottò un nuovo modello commerciale, nel quale ve-

³ Mi dicono che forse sono stati stimolati da uno studente che seguì il corso di Economia sperimentale di Charlie Plott all’Università di Cal Tech. Se così fosse gli economisti hanno avuto un ruolo in questo disegno d’aste da uno stadio veramente iniziale!

nivano messi all'asta degli spot pubblicitari da unire a ciò che essi chiamavano i risultati della ricerca "algoritmica". Quasi contemporaneamente cambiarono la loro denominazione in *Overture*.

Due dipendenti di Google, Salar Kamangar e Eric Veach, osservarono quello che stava facendo *Overture*, e decisero che potevano aggiungere dei miglioramenti. Nell'autunno del 2001 svilupparono la *Google Ad Auction* (l'Asta degli spot pubblicitari in Google).

Nel loro modello, gli spot erano classificati con una combinazione delle offerte e del tasso di accesso stimato. Dato che le offerte sono espresse in unità di costo/click e il tasso di accesso è dato dal rapporto n. di click / n. di impressioni, questo significa che gli spot vengono classificati in base al costo per impressione. L'idea era di mettere gli spot che hanno il massimo volume di ricavi previsto nelle migliori posizioni, cioè le posizioni in cui essi avrebbero avuto la massima probabilità di ricevere dei click.

Così come una azienda si interessa al prezzo moltiplicato la quantità venduta (prezzo- n. vendite), un motore di ricerca deve badare al prezzo per click moltiplicato per il numero dei click che si prevede di ricevere, in quanto questo costituisce il totale dei ricavi derivante dalla presentazione dello spot. Naturalmente, ciò richiede un metodo per stimare la probabilità di un click, un compito non facile. Tratterò di questo metodo di seguito.

Google si rese presto conto che un'asta al primo prezzo (in cui le agenzie di pubblicità pagavano l'importo della loro offerta) non sarebbe stata attraente, in quanto le agenzie avrebbero voluto ridurre la loro offerta al minimo importo che consentisse di conservare la loro posizione. Questo costante monitoraggio del sistema avrebbe comportato un carico significativo sui *server*, così Google decise di fissare automaticamente il prezzo pagato, nella stessa misura della seconda massima offerta, dal momento che si trattava di ciò che le agenzie avrebbero voluto fare comunque. Questa scelta non aveva niente a che fare con le aste Vickrey, ma era soprattutto una decisione in fatto di ingegneristica⁴.

⁴ GoTo.com sperimentò con un'asta al primo prezzo e determinò che portava a comportamenti instabili. ZHANG X.M. - PRICE J.F. (2005) and ZHANG X.M. (2005) documentano e modellano questo comportamento.

All'inizio, l'asta degli spot pubblicitari di Google si applicava solo agli spot che apparivano sul lato destro della pagina, in cui gli spot in cima alla pagina (l'area con il migliore rendimento) erano riservati ai prezzi negoziati da una forza vendite. Alla fine divenne chiaro che i prezzi generati dall'asta erano più congrui di quelli generati dalla negoziazione, così Google passò ad usare un'asta per tutti gli spot presentati.

5. - L'asta Google degli spot

L'asta Google degli spot è probabilmente l'asta più vasta del mondo, con miliardi di aste eseguite ogni settimana. Inoltre l'asta risulta avere una struttura teorica molto interessante, come viene descritto in Edelman *et al.* (2005) e Varian (2006).

Esistono diversi slot in cui possono inserirsi gli spot, ma alcuni ricevono più click di altri. In condizioni di equilibrio, ciascun offerente deve preferire lo slot in cui si trova a qualsiasi altro slot. Ciò comporta una serie di rapporti di «preferenza rivelata», cosa che può essere risolta con regole di equilibrio della gara. Di converso, date alcune offerte osservate, si possono invertire le regole di offerta per individuare quali valori le agenzie pubblicitarie attribuiscono ai click.

Per vedere in che modo questo funziona, consideriamo un offerente che pensa di partecipare ad un'asta di parole chiave. Coloro che stanno già partecipando offrono, ciascuno, un certo importo. Perciò il nuovo offerente si trova di fronte ad una «curva dell'offerta dei click». Più offre, e più dei precedenti offerenti verranno superati, arrivando così ad una posizione più alta e a più click di accesso.

Nel determinare la sua offerta, l'inserzionista deve prendere in considerazione il *costo incrementale per click*: quanto denaro in più dovrà spendere per ottenere un maggior numero di click. Se il costo incrementale per click è inferiore al valore per click, deve aumentare la sua offerta. Se il costo incrementale per click è superiore al valore per click deve ridurre la sua offerta. In condizioni di equilibrio, il costo incrementale di salire di una posizio-

ne deve superare il valore per click dell'offerente, ma il risparmio incrementale di scendere di una posizione deve essere inferiore al valore per click dell'offerente.

Ciò implica che, in condizioni di equilibrio, il costo incrementale per click dovrebbe aumentare il tasso di accesso. Perché? Supponiamo che esso sia diminuito nel passare da una posizione a quella successiva. In tal caso, c'era qualche offerente che aveva acquistato dei click costosi, ma aveva rinunciato a quelli meno costosi, contraddicendo l'assunto dell'equilibrio.

Inoltre, dato che il valore per click deve essere delimitato dal costo incrementale per click in equilibrio, i costi incrementali osservati ci consentono di dedurre preziose informazioni in merito ai valori attribuiti dagli offerenti. In pratica, il costo incrementale per click sembra dare una stima plausibile del valore del click per l'offerente.

Tuttavia, è importante notare che c'è ancora una certa indeterminazione dell'equilibrio. Il requisito che ciascuno agente preferisce la sua posizione ad altre possibili posizioni non prefissa un risultato unico. Piuttosto, esso determina una gamma di offerte in equilibrio. Due equilibri particolarmente interessanti sono quelli che fruttano i ricavi massimi e minimi del motore di ricerca.

6. - Sistema VCG di determinazione del prezzo

L'asta degli spot di Google è un modo di vendere all'asta le posizioni degli spot, ma si possono prendere in considerazione anche altri modi. Un difetto dell'asta attuale è che ciascun inserzionista deve confrontare i suoi costi incrementali al suo valore, e tali costi incrementali dipendono dalle decisioni degli altri offerenti.

Ma esiste anche un altro meccanismo, simile alle aste, che non ha questo difetto: il sistema di Vickrey-Clarke-Groves (VCG). Nel sistema VCG: 1) ciascun agente dichiara un valore; 2) il motore di ricerca assegna gli agenti pubblicitari agli slot in modo tale da massimizzare il valore totale dell'assegnazione; 3) quindi ciascun agente «a» paga un prezzo pari al valore totale che matura

a favore degli altri agenti se “a” è presente, meno il valore totale che matura a favore degli altri agenti se “a” è assente. In questo modo, ciascun agente paga un importo pari al costo che esso impone agli altri agenti.

È possibile dimostrare che per questo meccanismo, ciascun agente deve dichiarare il suo vero valore, a prescindere dalle dichiarazioni degli altri agenti. Hermann (1983) è stato il primo ad applicare questo meccanismo al classico problema dell'assegnazione. Alcuni anni più tardi, Demange e Gale (1985) hanno mostrato che questo meccanismo determina gli stessi pagamenti dell'equilibrio Nash del ricavo minimo da un mercato equivalente all'asta degli spot di secondo-prezzo⁵.

Ci sono altre interessanti proprietà dell'asta VCG. Ad esempio, Krishna e Perry (1998) mostrano che il meccanismo VCG massimizza i ricavi del motore di ricerca tra tutti i meccanismi efficienti. Malgrado gli evidenti vantaggi di VCG, esso non è stato ancora utilizzato da alcuno dei principali motori di ricerca.

7. - L'importanza della competizione

È ampiamente riconosciuto che il ricavo realizzato in un'asta dipende, criticamente, da quanta competizione è presente in quell'asta. Klemperer (2002) descrive il caso delle aste del Giugno del 2000 per le licenze di telefonia mobile in Olanda, in cui vi erano 5 licenze e 6 offerenti. Un offerente minacciò un'azione legale contro un altro, se esso avesse continuato a fare offerte, costringendolo a rinunciare, lasciando così 5 offerenti per 5 licenze, cioè senza una competizione! Difatti l'asta raccolse meno del 30% di quanto aveva previsto il governo olandese.

Lo stesso principio vale per l'asta delle posizioni: i ricavi non decollano fino a che non c'è competizione.

Nell'asta Google ci sono 8 slot per gli spot sul lato destro della pagina, e fino a 3 slot per quelli in cima alla pagina. Come det-

⁵ Sto semplificando i risultati per facilità di esposizione. Vedasi VARIAN H.R. (2006) per i dettagli.

to in precedenza, l'ordine degli spot è determinato dalle offerte e dai tassi di accesso, ma gli spot che vengono "promossi" (spostati alla cima della pagina) devono soddisfare alcuni criteri addizionali, che riguardano anche la qualità degli spot.

Per semplificare, se un'asta ha meno offerenti degli slot disponibili, o un numero di offerenti appena sufficiente a riempire gli slot disponibili, noi diciamo che è "svenduta". Se ha più offerenti degli slot diciamo che è "venduta oltre la disponibilità". Se un'asta è svenduta, il prezzo pagato dall'ultimo offerente sulla pagina è il prezzo di riserva, che supponiamo sia di 5 centesimi⁶. Se la pagina è venduta oltre la disponibilità, il prezzo pagato dall'ultimo offerente sulla pagina è determinato dall'offerta del primo agente escluso, che può facilmente essere almeno 10 volte più alto del prezzo di riserva.

Consideriamo un esempio semplice in cui tutti gli offerenti hanno lo stesso valore v e il prezzo di riserva è r . Diciamo che p_s è il prezzo pagato per lo slot s e x_s è il numero di click che lo slot s riceve. Se la pagina è svenduta, ciascun offerente deve essere indifferente tra pagare p_s e ricevere x_s click rispetto a pagare r e ricevere x_m click, dove m è l'ultimo annuncio mostrato sulla pagina. Ciò implica

$$(v - p_s) x_s = (v - r) x_m$$

oppure

$$p_s x_s = v (x_s - x_m) + r x_m$$

Queste equazioni dicono che la spesa sullo slot s deve essere la spesa sull'ultimo slot più il valore incrementale dei click nella posizione s .

D'altra parte, supponiamo che la pagina sia venduta oltre la disponibilità, così che esiste almeno un offerente escluso con valore v . In tal caso, ciascun offerente deve essere indifferente tra ciò che sta pagando e il profitto derivante dal fatto di essere escluso — che è zero. Ciò ci dà

⁶ Il prezzo di riserva dipende anche dalla qualità dell'annuncio.

$$(v - p_s) x_s = 0$$

che implica $p_s = v$.

Da notare il grande salto nei ricavi, nel passare da una pagina parzialmente venduta ad una pagina venduta oltre la disponibilità. Nel primo caso, ognuno è indifferente tra l'essere nello slot in cui si trova e l'essere nello slot pessimo. Nel secondo caso, ognuno è indifferente tra l'essere mostrato e il non essere mostrato affatto, il che significa che i prezzi sono concorrenti fino al valore uguale.

Per meglio comprendere questo punto, consideriamo un esempio semplice.

Supponiamo che ci siano 2 slot. Quello in cima riceve 100 click al giorno, il secondo 80 click al giorno. Ci sono due agenti pubblicitari, ciascuno dei quali valuta un click a 50 centesimi.

In questo modello, un agente occupa lo slot 2 e riceve 80 click al giorno, per cui paga 5 centesimi ogni click = \$ 4 in totale. Il secondo agente occupa lo slot in cima ottenendo 20 click addizionali al giorno. La competizione lo costringe a pagare \$ 10 in più per quei click rispetto all'agente dello slot 2. In questo modo, egli spende \$14 = \$4 + \$10, in totale. I ricavi totali per il motore di ricerca sono \$18.

Ora supponiamo che ci siano 3 agenti che valutano i click 50 centesimi ciascuno, ma restano ancora solo 2 slot. L'offerta in equilibrio è ora di 50 centesimi ogni click, ci sono 180 click in totale, pertanto i ricavi totali ottenuti dai due agenti sono \$90. L'aggiunta di un altro agente fa aumentare i ricavi da \$18 a \$90!

Questo esempio illustra il punto importante che le pagine vendute oltre la disponibilità sono molto più redditizie delle pagine vendute parzialmente non solo perché ci sono più offerenti, ma anche perché le forze della competizione sono molto più forti.

Questo punto illustra anche l'importanza dell'algoritmo di abbinamento utilizzato per presentare gli annunci. L'utente immette una "interrogazione" e l'agente pubblicitario acquista delle "keyword". L'agente può specificare "abbinamento esatto", che significa che l'annuncio è mostrato solo se l'interrogazione dell'utente combacia esattamente con le keyword dell'agente pubbli-

tario. Ma è più frequente che gli agenti specifichino «abbinamento approssimativo» che significa che l'interrogazione si abbinerà con varie espansioni della Keyword quali i sinonimi e le sottostringhe.

Gli annunci addizionali per via dell'abbinamento approssimativo beneficiano l'utente e l'agente pubblicitario, dato che rendono più probabile che l'utente effettuerà il click. Ma accrescono anche la competizione dell'asta, facendo aumentare i prezzi.

8. - Qualità dell'annuncio pubblicitario

Ho indicato in precedenza che la classificazione utilizzata da Google e Yahoo si basa non solo sulle offerte, ma anche su una misura della qualità dell'annuncio. Nel caso più semplice, possiamo pensare alla qualità dell'annuncio, come al previsto tasso di accesso. Google classifica gli annunci mediante il tasso di accesso previsto, moltiplicato per il numero l'offerta, ma da dove viene la stima del tasso di accesso previsto?

Pensate ad un modello in cui l'effettivo tasso di accesso ricevuto da un annuncio dipende sia da un effetto specifico della posizione (x_p) che da un effetto specifico dell'annuncio (e_a). La specificazione più semplice del tasso di accesso di un annuncio a nella posizione p è data da $e_a x_p$.

Data questa forma moltiplicativa, è relativamente facile stimare i valori di pertinenza: basta mettere annunci a caso nella posizione p per stimare l'effetto specifico della posizione. Una volta che questo è noto, si può usare la storia dei click su un dato annuncio per stimare l'effetto specifico dell'annuncio. Si possono usare anche diversi altri predittori per integrare i dati storici.

La classificazione degli annunci si basa sugli effetti specifici dell'annuncio, moltiplicati per l'offerta: $b_a e_a$. L'offerta è in dollari per click, e l'effetto specifico dell'annuncio è il numero di click per impressione. Pertanto $b_a e_a$ è offerta per impressione: quanto è disposto a pagare l'agente perché il suo annuncio venga mostrato ad un utente. All'agente con il massimo valore per un'impressione viene data la posizione migliore: la posizione con la massima probabilità di ricevere un click. L'agente con il secondo mas-

simo valore per impressione ottiene la successiva miglior posizione, e così via.

Perciò un annuncio con un'alta offerta per click potrebbe essere superato da un annuncio con un'offerta inferiore se l'annuncio con un'alta offerta avesse un tasso di accesso basso. Assegnare gli annunci sulla base di b_a e e_a massimizza il valore delle impressioni sulla pagina, determinando un aumento dei ricavi previsti.

Così come è importante determinare quali annunci mostrare, è altrettanto importante determinare quali annunci *non* mostrare. Il motivo è che la probabilità che un utente clicchi su un annuncio dipende da quanto pertinente egli preveda che quell'annuncio sia. E questa previsione dipende, almeno in parte, da quella che è stata la precedente esperienza dell'utente.

Mostrando così che un «cattivo annuncio» può influire sulla propensione futura degli utenti a cliccare. Offrire un cattivo annuncio in una posizione particolarmente preminente può essere particolarmente costoso.

La decisione se, e dove, mostrare un annuncio dovrebbe dipendere non solo dagli attuali ricavi di un annuncio, ma da una stima di come la pertinenza dell'annuncio influirà sulla futura propensione a cliccare. È possibile modellare analiticamente queste decisioni. Il fatto di mostrare un annuncio oggi fa introitare un importo di ricavi noto, ma ha anche un effetto probabilistico sui futuri ricavi, influenzando la propensione a cliccare in futuro. Modellare questi effetti porta ad un problema di programmazione dinamica stocastica che offre una logica alle attuali pratiche e una guida a come esse debbano essere affinate.

9. - Conclusione

I motori di ricerca sono un esempio di un modello di abbinamento bilaterale supportato dalla pubblicità. Essi sono interessanti non solo di per sé, ma offrono un terreno fertile all'analisi economica.

Durante gli anni 1960 e 1970, lo studio scientifico dei mer-

cati finanziari fiorì a causa della disponibilità di massicce quantità di dati e dell'applicazione di metodi quantitativi. Io penso che il *marketing* sia nella stessa posizione in cui era la finanza nei primi anni 1960. Grandi quantità di dati informatizzati sulla *performance* del *marketing* stanno diventando disponibili solo ora attraverso i motori di ricerca, gli scanner dei supermercati, e altri tipi di tecnologia informatica. Tali dati forniscono il materiale grezzo per degli studi scientifici del comportamento dei consumatori, e prevedo che ci sarà un grande progresso in quest'area nel prossimo decennio.

BIBLIOGRAFIA

- ARROW K.J., «The Economic Implications of Learning by Doing», *Review of Economic Studies*, n. 29, 1962, pp. 155-73.
- COOPER W. - CHEN A. - GEY F.C., «Full Text Retrieval Based on Probabilistic Equations with Coefficients Fitted by Logistic Regression», *TREC93*, 1993.
- — — — —, «Experiments in the Probabilistic Retrieval of Full Text Documents», *TREC 1994*, 1994.
- DEMANGE G. - GALE D., «The Strategy Structure of Two-Sided Matching Markets», *Econometrica*, n. 53, 1985, pp. 873-8.
- EDELMAN B. - OSTROVSKY M. - SCHWARTZ M., «Internet Advertising and the Generalized Second Price Auction», *NBER, Working Paper*, n. 11765, November, 2005.
- FALLOWS D., «Search Engine Users», Pew Foundation, URL http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf, 2005.
- KLEMPERER P., «What Really Matters in Auction Design», *Journal of Economic Perspectives*, n. 16 (1), 2002, pp. 169-89. URL <http://www.nuff.ox.ac.uk/users/klempere/design3aweb.pdf>.
- KRISHNA V. - PERRY M., «Efficient Mechanism Design. Technical Report», SSRN, 1998, URL http://www.papers.ssrn.com/sol3/papers.cfm?abstract_id=64934.
- LANGVILLE A.N. - MEYER C.D., *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, NJ, 2006.
- HERMAN B.L., «Elicitation of Honest Preferences for the Assignment of Individuals to Positions», *Journal of Political Economy*, n. 91, 1983, pp. 461-79.
- LESK M., «The Seven Ages of Information Retrieval. Technical Report, International Federation of Library Associations and Institutions», 1995, URL <http://www.ifla.org/VI/5/op/udtop5/udtop5.htm>.
- TREC8, «The TREC Conferences: An Introduction», NIST, 2000, URL <http://www.trec.nist.gov/presentations/TREC8/intro/>.
- VARIAN H.R., «Position Auctions», *International Journal of Industrial Organization*, n. 24 (7), 2006, pp. 1-10, URL <http://www.sims.berkeley.edu/~hal>.
- ZHANG X.M., «Finding Edgeworth Cycles in Online Advertising. Technical report», MIT Sloan School, 2005, URL http://www.marketing.wharton.upenn.edu/news/colloquia/michael_zhang.pdf.
- ZHANG X.M. - PRICE J.F., «Cycles in Online Advertising Auctions», in *Proceedings of the 26th International Conference on Information Systems*, Las Vegas (NV), 2005, ICIS.

