

Sulla Classificazione delle Sedi di Pubblicazione nella Valutazione della Produzione Scientifica

Giansalvatore Mecca, Marcello Buoncristiano, and Donatello Santoro

Dipartimento di Matematica e Informatica
Università della Basilicata – Potenza – Italy
<http://db.unibas.it/valutazione>

1 Introduzione e Motivazioni

Questo lavoro discute alcune tecniche per la valutazione della produzione scientifica. Si tratta della versione ridotta di un lavoro più ampio (Mecca, 2012); rimandiamo alla versione estesa per tutti i dettagli di carattere tecnico che in questa sede sono stati omessi.

Il problema della valutazione della produzione scientifica è di grande attualità nelle Università italiane. E' di questi giorni l'avvio delle procedure di valutazione nazionale della qualità della ricerca relativa al quadriennio 2004-2010, e la pubblicazione del D.M. n. 19 del 27 gennaio 2012 (Ministero dell'Università e della Ricerca, 2012) che introduce il sistema nazionale di valutazione del sistema universitario e obbliga le università a dotarsi di un proprio sistema di autovalutazione basato su "parametri oggettivi, volti a misurare in ogni momento l'efficienza e l'efficacia della didattica e della ricerca messa in atto dai singoli atenei e a stimolare la competitività e la qualità degli stessi".

E' quindi ragionevole attendersi che nell'arco dei prossimi mesi tutte le università italiane dovranno dotarsi di modelli e procedure per la misurazione della "performance" dei docenti (professori e ricercatori). In questo articolo, ci concentriamo su uno degli aspetti che le Università sono chiamate a valutare, ovvero la ricerca, e più specificamente, sulla valutazione della produzione scientifica.

Le attività di valutazione di cui trattiamo hanno caratteristiche ben precise: (a) considerano la produzione scientifica dei docenti dell'ateneo in un periodo di riferimento, tipicamente tre o cinque anni; (b) vengono condotte periodicamente, per esempio una volta all'anno, e su larga scala, perché riguardano tutti i docenti dell'Ateneo; (c) richiedono di confrontare la produzione di docenti di aree anche molto diverse tra di loro – dalla matematica alla sociologia – per esempio a scopo di ripartizione di risorse; (d) devono essere concluse in tempi ragionevolmente brevi per servire allo scopo per cui sono adottate; (e) possono contare su budget limitati o addirittura assenti.

E' importante rimarcare le differenze tra valutazioni dei singoli su scala di Ateneo ed iniziative con scopi e finalità diverse. L'ANVUR, ad esempio, sta conducendo un'attività di valutazione che non riguarda i singoli, ma le strutture. Si tratta di un processo di lunga durata (durerà più di un anno), con un notevole budget economico. Non è possibile, di conseguenza, pensare di adottare esattamente le stesse tecniche per i nostri scopi.

In effetti, le caratteristiche del processo di cui ci occupiamo costringono a scelte sostanzialmente obbligate relativamente al metodo da adottare per condurre la valutazione. E' noto il fatto che esistono tre tecniche principali per la valutazione della produzione scientifica. (1) il "peer reviewing", ovvero la valutazione da parte di revisori anonimi; (2) l'uso di indicatori bibliometrici basati sul numero di citazioni ricevute dai lavori; (3) l'uso della qualità delle sedi di pubblicazione.

Nell'ambito delle procedure di valutazione di cui parliamo, il "peer reviewing"- tecnica, peraltro, non esente da limiti, come discusso ampiamente nella letteratura recente (Casati, 2009) – è chiaramente inapplicabile, per via dei vincoli di tempo e di risorse.

E' delicato anche l'utilizzo esclusivo di indici bibliometrici basati sul numero di citazioni, per varie ragioni. Per cominciare, questi indici sono più solidi se utilizzati per analizzare la produzione scientifica in un periodo medio-lungo; mostrano una certa fragilità se applicati a periodi più brevi e recenti. Inoltre, è dimostrato (Iglesias, 2007) che le diverse comunità scientifiche hanno stili e pratiche di citazione completamente diversi gli uni dagli altri, e di conseguenza che è estremamente delicato confrontare comunità lontane sulla base di questi indici.

Alla luce di queste considerazioni, nel seguito di questo lavoro ci concentreremo esclusivamente sul terzo metodo, ovvero la valutazione dei prodotti di ricerca attraverso la valutazione della qualità delle corrispondenti sedi di pubblicazione. E' possibile pensare che questa tecnica sia affiancata in modo più o meno significativo dall'utilizzo di indicatori basati sulle citazioni dei lavori, dei quali però non discuteremo ulteriormente.

In sintesi, nell'approccio di cui trattiamo, il valore di un lavoro scientifico – per esempio un articolo pubblicato su rivista o negli atti di un convegno – viene valutato sulla base della "classe di merito" della corrispondente rivista o del corrispondente convegno. La classe di merito assume normalmente un valore pari a A, B, C, D, per indicare le sedi di primo piano, quelle di secondo piano eccetera.

E' facile immaginare come, in questo approccio, l'aspetto centrale della valutazione sia l'adozione di classificazioni affidabili delle sedi di pubblicazione.

Il problema centrale affrontato in questo lavoro è quello della combinazione tra diversi ranking della stessa sede di pubblicazione. Vengono discusse tre diverse classificazioni, una per le riviste, una per i convegni, una per le case editrici. In tutti i casi, le classificazioni sono state generate combinando ranking esistenti di carattere internazionale e nazionale. A questo scopo, sono stati predisposti diversi strumenti informatici, a supporto di un processo complesso di estrazione, pulitura e integrazione dei dati.

ABDOMINAL IMAGING		Aggregate ranking B/C ?		
ISI	GASTROENTEROLOGY & HEPATOLOGY	ABDOMINAL IMAGING	0942-8925	C
ISI	RADIOLOGY, NUCLEAR MED. & MED. IMAGING	ABDOMINAL IMAGING	0942-8925	C
Scopus	Medicine	Abdominal Imaging	1432-0509	B
ERA		Abdominal Imaging	0942-8925	C

Figura 1 Gruppo di Valori di Ranking per la Rivista "Abdominal Imaging". Due provengono da categorie ISI, uno dall'Area Scopus "Medicine", l'ultimo dalla classificazione ERA

Al termine di questo passo, per ciascuna sede di pubblicazione è stato costruito un gruppo di valori di ranking, come mostrato in Figura 1 per la rivista “Abdominal Imaging”. Il passo successivo è quello cruciale, e consiste nell’assegnare a ciascuna sede di pubblicazione un valore di ranking, omogeneizzando opportunamente quelli presenti nelle classificazioni di partenza.

2 Funzioni di Voto e di Consenso

Il problema che intendiamo affrontare è quello della definizione di strategie per l’integrazione di classificazioni alternative della stessa sede di pubblicazione, ad esempio una rivista. Dall’unione delle diverse classificazioni, ciascuna sede di pubblicazione riceve un insieme di *valori di ranking*, come mostrato in Figura 1. I valori di ranking provengono da un dominio discreto, enumerabile ed ordinato. In prima istanza, supponiamo che i valori possibili siano, in ordine decrescente, $\{A+, A, B, C, D, NC\}$, dove NC è il “valore nullo” del dominio che indica una sede di pubblicazione non classificabile.

Si noti che l’insieme dei valori di ranking all’interno di una classificazione rappresenta una variabile aleatoria a valori discreti. Le diverse variabili aleatorie possono difficilmente essere considerate indipendenti, dal momento che tutte hanno l’obiettivo di misurare il valore scientifico delle sedi di pubblicazione, tipicamente attraverso una stima della relativa reputazione e del relativo impatto.

Una *funzione di ranking* associa ad un gruppo di valori di ranking un valore di ranking aggregato, e rappresenta una strategia per combinare i diversi ranking disponibili per una stessa sede di pubblicazione.

Nella letteratura sono documentate varie soluzioni al problema. Alcune di queste sono basate puramente sull’attribuzione di una priorità alle sorgenti di classificazione: in caso di valori multipli di ranking prevale quello della sorgente a maggiore priorità. Un esempio di questa strategia è applicato dall’Università di Bologna, che ha sviluppato una propria classificazione, ma in tutti i casi in cui una rivista è indicizzata dall’ISI dà priorità al ranking basato sull’impact factor.

Altre soluzioni adottano approcci più sofisticati. In particolare, è in linea di principio possibile adottare funzioni di carattere numerico che combinino i diversi valori del gruppo. Questo approccio, però, nasconde numerose insidie. Per cominciare, come discusso precedentemente, si tratta di combinare i valori di variabili aleatorie non indipendenti. E’ inoltre noto che funzioni di aggregazione semplici, come la media dei valori numerici corrispondenti ai valori di ranking, possono condurre ad errori grossolani (Thompson, 1993).

Queste considerazioni, che mettono in luce le insidie dell’operazione che stiamo conducendo, suggeriscono l’adozione di funzioni di ranking alternative. In effetti, è relativamente semplice immaginare funzioni di ranking che facciano al nostro scopo. Due esempi sono la funzione “*max*”, che associa alla sede di pubblicazione il valore più alto tra quelli del gruppo, e la funzione “*min*”, che restituisce il valore più basso diverso da NC.

Nel contesto della valutazione della ricerca, l'utilizzo esclusivo di queste due funzioni sembra inopportuno; l'esperienza mostra infatti che non è infrequente l'esistenza di gruppi di ranking ad alta distanza. La *distanza* di un gruppo di valori di ranking è il valore assoluto della differenza tra il valore più piccolo e quello più grande appartenenti al gruppo, escludendo il valore NC¹. Ad esempio, nel gruppo [A+, A, B, B, C] la distanza è 3. In un gruppo del genere, sembra del tutto arbitrario l'utilizzo tanto della funzione "max", che selezionerebbe il valore di ranking aggregato A+, che della funzione "min", che selezionerebbe il valore di ranking aggregato C.

Di conseguenza, in questo lavoro adotteremo un approccio diverso, basato sull'utilizzo di *funzioni di voto e di consenso*. Il modello sottostante al nostro approccio è quello dell'elezione. Ciascuna categoria esprime per una sede di pubblicazione un *voto* relativamente al suo ranking. Le funzioni di nostro interesse cercano i valori che, sulla base del ranking, raggiungono il massimo supporto.

La prima, naturale funzione che introduciamo è la funzione "majority". La funzione seleziona in un gruppo il valore di ranking che riceve un numero di voti pari alla *maggioranza assoluta*. Ad esempio, nel gruppo [A+, B, B, B, C], la funzione seleziona il valore B, che ha ricevuto 3 voti sui cinque espressi. Nel gruppo [A, A, B, NC], la funzione seleziona il valore A, che ha ricevuto 2 voti sui 3 espressi (in questo caso escludiamo dal computo della maggioranza il valore NC, che è considerato un "non voto"). Si noti che la funzione "majority" non è definita per tutti i gruppi. Ad esempio, per il gruppo [A, A, B, C], non restituisce alcun valore perché non c'è maggioranza assoluta. Lo stesso vale per il gruppo [A, A, B, B]. In questo caso assumeremo che il valore restituito dalla funzione sia NC.

Come è comune nell'utilizzo di funzioni di voto (Wikipedia - MF) per dirimere la maggioranza nel caso di voti in numero pari, è possibile introdurre un *pregiudizio* a favore o a sfavore della maggioranza. La ragione di questo fatto è che ottenere maggioranze assolute su gruppi di dimensione pari e di piccole dimensioni è più difficile: in un gruppo di 6 valori, la maggioranza assoluta richiede 4 voti, il che, in percentuale, è molto più dei 4 voti necessari in un gruppo di 7 valori.

Definiamo di conseguenza la variante "*mildMajority*" della funzione "majority", introducendo un pregiudizio a favore della maggioranza; più precisamente, diciamo che un valore ha la maggioranza assoluta se si verificano tutte le seguenti condizioni: (a) ha ricevuto la maggioranza relativa dei voti²; (b) il suo numero di voti aumentato di

¹ Per differenza tra due valori in un insieme discreto ordinato intendiamo la differenza tra la posizione nell'ordinamento del primo valore e quella del secondo valore.

² Un valore riceve la maggioranza relativa dei voti quando è il più votato in assoluto (non c'è un altro valore che ha riportato lo stesso numero di voti). In termini statistici, questo implica che il valore è la *moda* (Wikipedia - Moda) della distribuzione, e che la distribuzione è di tipo *unimodale*; aver ricevuto la maggioranza relativa non implica il raggiungimento della maggioranza assoluta; ad esempio, nel gruppo [A, A, B, C, D], il valore A ha ricevuto la maggioranza relativa dei voti, ma non la maggioranza assoluta, che richiederebbe 3 voti. Nel gruppo [A, A, B, B], come nel gruppo [A, B] nessun valore riceve la maggioranza relativa, e di conseguenza la maggioranza assoluta. La condizione imposta sulla funzione è particolarmente importante per questo tipo di distribuzioni, dette *bimodali*, per evitare che la funzione scelga arbitrariamente uno dei due valori più frequenti.

0,1 supera la metà del numero di voti espressi nel gruppo. Si noti che il pregiudizio è tale per cui, nel caso di numero di voti dispari non c'è nessuna differenza tra "majority" e "mildMajority"; viceversa, la differenza è significativa in caso di numero di voti pari. Nel caso discusso sopra, [A, A, B, C], "majority" vale NC, mentre "mildMajority" vale A. Resta, viceversa, NC il valore nel caso dei gruppi [A, A, B, B], [A, B, C] e [A, B].

Per risolvere opportunamente anche questi gruppi, è possibile ulteriormente ammorbidire la funzione, passando dal concetto di voto al concetto di *consenso*. Abbiamo finora assunto che una categoria di classificazione "vota" per un valore di ranking in tutti i casi in cui esprime esattamente quel valore. Diciamo ora che, votando per un valore, la categoria esprime anche implicitamente *consenso* per tutti i valori inferiori. In altri termini, votando per A, si esprime consenso per A, ma anche per B, C e D.

La funzione "*majorityConsensus*" (e la sua variante "*mildMajorityConsensus*") restituisce il valore più alto di un gruppo per cui c'è un consenso superiore alla metà dei voti (ovvero, il consenso aumentato di 0,1 supera la metà dei voti).

Le funzioni "*majorityConsensus*" e "*mildMajorityConsensus*" consentono di risolvere alcuni dei gruppi per cui non c'è maggioranza dei voti, ad esempio [A, B, C], con il valore B.

Ulteriori dettagli relativi alle funzioni utilizzate sono descritte nella versione estesa di questo lavoro (Mecca, 2012).

3 La Classificazione delle Riviste

Utilizzando le tecniche discusse, è stata generata un'ampia classificazione di riviste. La classificazione discende dall'integrazione di 4 classificazioni internazionali ed 8 classificazioni nazionali:

- la classificazione basata sugli IF della ISI Wos (ISI);
- la classificazione basata sull'indice SJR di Scopus (Scopus);
- la classificazione australiana ERA (ERA);
- la classificazione ERIH della European Science Foundation (ESF ERIH);
- le classificazioni prodotte nell'ambito della VQR dei GEV delle Aree 01, 08, 09, 10, 11, 12, 13, 14.

La classificazione risultante è disponibile all'indirizzo <http://db.unibas.it/valutazione>.

4 La Classificazione dei Convegni

Il presupposto alla base della classificazione dei convegni è quello di valutare, nei casi in cui questo è opportuno, una pubblicazione apparsa negli atti di un convegno alla pari di una pubblicazione su rivista. Si tratta di una pratica non comune: in molte aree disciplinari le pubblicazioni in atti di convegno vengono considerati lavori di secondo piano rispetto alle riviste. In effetti, nei criteri fissati per la VQR (VQR), vari dei GEV dicono chiaramente che un lavoro in atti di convegno non può essere valuta-

to oltre la classe C o addirittura D. Più specificamente, questo è detto esplicitamente per l'Area 02, 03, 04, 05, 06 e 07. Relativamente all'Area 01 – che accomuna matematici ed informatici – è certamente vero per i matematici.

Di conseguenza, il problema della classificazione si pone principalmente per l'area informatica, le aree dell'ingegneria, e le aree umanistiche. La disponibilità di ranking nazionali ed internazionali in queste tre aree è molto disomogenea. Per cominciare, non esistono ranking di convegni di area umanistica. L'unico ranking noto di convegni di area ingegneristica è il ranking australiano (ERA); nel ranking, sono classificati circa 300 convegni delle varie aree dell'ingegneria. Esistono, viceversa, numerosi ranking di convegni di ambito informatico, per i quali un'ampia letteratura (CSTB, 1994) (Patterson, 1999) (Meyer, 2009) (Rahm, 2005) (Fortnow, 2009) sottolinea la necessità di considerare gli atti dei convegni quali sedi pubblicazione di primo piano; non a caso, oltre 1500 dei convegni classificati dagli australiani sono di area informatica. Lo stesso GEV dell'Area 01, nelle Frequently Asked Questions, risponde a questo proposito: “Siamo consapevoli dell'importanza delle conferenze per la ricerca in informatica, in alcuni casi non inferiore a quella di riviste in classe di merito 1.”

In questo contesto, dovendoci necessariamente dotare di una classificazione dei convegni per tenere adeguatamente in considerazione un aspetto centrale delle pratiche di pubblicazione di alcuni settori, pare opportuno da una parte limitare il ranking alle fasce più alte, quelle da cui è più ragionevole attendersi un livello scientifico comparabile a quello delle riviste; d'altra parte, è necessario adottare una strategia in cui i valori di ranking risultino sufficientemente autorevoli. In particolare, per l'area informatica, abbiamo considerato 6 diversi ranking di convegni informatici:

- il ranking australiano ERA (ERA), integrato con i convegni di fascia A+ del ranking Core (CORE);
- il ranking del sito Microsoft Academic Search (Microsoft Academic Search), limitandosi ai primi 1000 convegni;
- il ranking del sito Arnetminer (ArnetMiner, 2008), limitandosi ai primi 1000 convegni;
- il ranking dell'Università dell'Alberta (Zaiane);
- il ranking della Technical University of Singapore (NTU);
- il ranking del GRIN.

5 La Classificazione delle Case Editrici

Tra le attività di classificazione delle sedi di pubblicazione, quella delle case editrici è decisamente la meno consolidata. A livello internazionale esistono pochissimi ranking di ampio respiro. Per i nostri scopi, abbiamo utilizzato il ranking dell'organizzazione olandese SENSE (SENSE) e quello del Finnish Publication Forum (Finnish Publication Forum). Più in dettaglio:

- il ranking SENSE classifica circa 1200 case editrici nelle fasce A, B, C, D;
- il ranking finlandese, classifica anch'esso circa 1200 case editrici; per la classificazione parallela delle riviste, il ranking finlandese utilizza 3 fasce, 1, 2, e 3, secondo

la convenzione per cui 3 è la fascia migliore (cioè la fascia A) ed 1 la peggiore (cioè la fascia C); per le case editrici, vengono utilizzate esclusivamente le fasce 1 e 2; di conseguenza, abbiamo adottato la convenzione di associare alle case editrici di fascia 1 il valore di ranking C, ed a quelle di fascia 2 il valore di ranking B.

Nel complesso, dall'integrazione dei due ranking è possibile ottenere dati di classificazione per circa 2100 case editrici. Questo dato, però, non è confortante per due ragioni. Per cominciare, la sovrapposizione tra le due classificazioni è bassissima – solo 200 case editrici su 1200; questo significa che la classificazione integrata è ampia, ma relativamente poco solida, dal momento che la stragrande maggioranza dei valori di ranking si basa su un'unica sorgente. Inoltre, molte delle case editrici italiane di rilievo sono del tutto assenti.

Bibliografia

- ArnetMiner. (2008). *Computer Science Conference Ranking*. Tratto da <http://v1.arnetminer.org/page/conference-rank/html/All-in-one.html>
- Casati, F. a. (2009). *Is Peer Review any Good? A Quantitative Analysis of Peer Review*. LiquidPub Project Technical Report. <http://eprints.biblio.unitn.it/1654/>.
- CORE. (s.d.). *Computing Research and Education Association of Australia - Conference Ranking*. Tratto da <http://core.edu.au/index.php/categories/conference%20rankings/1>
- CSTB. (1994). *Computer Science and Telecommunications Board - Academic Careers for Experimental Computer Scientists and Engineers*. CSTB Report. http://www.nap.edu/openbook.php?record_id=2236.
- ERA. (s.d.). *Excellence of Research Australia*. Tratto da <http://www.arc.gov.au/era>
- ESF ERIH. (s.d.). *European Science Foundation - European Reference Index for the Humanities*. Tratto da <https://www2.esf.org/asp/ERIH/Foreword/search.asp>
- Finnish Publication Forum. (s.d.). *Ranking delle Case Editrici*. Tratto da <http://www.tsv.fi/julkaisufoorum/engish.html>
- Fortnow, L. (2009). Time for Computer Science to Grow-Up. *Communications of the ACM*, 58(8), 33-35.
- GRIN. (2000). *Gruppo Nazionale di Informatica - Classificazione delle Conferenze*. Tratto da http://www.grin-informatica.it/opencms/opencms/grin/ricerca/valutazione/classificazione_grin.html
- Iglesias, J. E. (2007). Scaling the h-Index for Different Scientific ISI Fields. *Scientometrics*, 73(3), 303-320.
- ISI. (s.d.). *Thomson Reuters - ISI Web of Science*. Tratto da <http://www.isiknowledge.com/WOS>
- Mecca, G. a. (2012). *Sulla Classificazione delle Sedi di Pubblicazione nella Valutazione della Produzione Scientifica*. Università della Basilicata, Dipartimento di Matematica e Informatica, <http://www.db.unibas.it/valutazione>.

- Meyer, B. a. (2009). Research Evaluation for Computer Science. *Communications of the ACM*, 52(4), 31-34.
- Microsoft Academic Search. (s.d.). *Computer Science Conference Ranking*. Tratto da <http://academic.research.microsoft.com/?SearchDomain=2>
- Ministero dell'Università e della Ricerca. (2012). D.M. n. 19 del 27 gennaio 1012 - Valorizzazione dell'efficienza delle universita' e conseguente introduzione di meccanismi premiali nella distribuzione di risorse pubbliche sulla base di criteri definiti ex ante anche mediante la previsione di un sistema d.
- NTU. (s.d.). *Nanyang Technical University of Singapore - Computer Science Conference Ranking*. Tratto da <http://www.ntu.edu.sg/home/assourav/crank.htm>
- Patterson, D. a. (1999). Computing Research Association - Best Practices Memo: Evaluating Computer Scientists and Engineers for Promotion and Tenure. *Computing Research News*, A-B.
- Rahm, E. a. (2005). Citation Analysis of Database Publications. *SIGMOD Record*, 34(4), 48-53.
- Scopus. (s.d.). *Elsevier - Scopus*. Tratto da <http://www.info.sciverse.com/scopus>
- SENSE. (s.d.). *SENSE Publisher Ranking*. Tratto da <http://www.sense.nl/qualityassessment>
- SIDEA. (s.d.). *Società Italiana di Economia Agraria - Relazione di Sintesi del Gruppo di Valutazione della Ricerca*. Tratto da https://ilo.unimol.it/sidea/images/upload/comitati/relazione_di_sintesi_valutazione.doc
- SIE. (s.d.). *Società Italiana degli Economisti - Classificazione delle Riviste*. Tratto da <http://www.siecon.org/online/anvur-cepr-cun-documenti/classificazione-per-fasce-delle-riviste-di-economia/>
- Thompson, B. (1993). *GRE Percentile Ranks Cannot Be Added or Averaged: A Position Paper Exploring the Scaling Characteristics of Percentile Ranks, and the Ethical and Legal Culpabilities Created by Adding Percentile Ranks in Making "High-Stakes" Admission Decisions*. <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED363637>: Education Resources Information Center Report ED363637.
- VQR. (s.d.). *ANVUR - Valutazione della Qualità della Ricerca 2004-2010*. Tratto da <http://www.anvur.org/?q=schema-dm-vqr-definitivo>
- Wikipedia - M. (s.d.). *Mediana*. Tratto da [http://it.wikipedia.org/wiki/Mediana_\(statistica\)](http://it.wikipedia.org/wiki/Mediana_(statistica))
- Wikipedia - MF. (s.d.). *Wikipedia - Majority Function*. Tratto da http://en.wikipedia.org/wiki/Majority_function
- Wikipedia - Moda. (s.d.). *Moda*. Tratto da Wikipedia: [http://it.wikipedia.org/wiki/Mediana_\(statistica\)](http://it.wikipedia.org/wiki/Mediana_(statistica))
- Zaiane, O. (s.d.). Tratto da Computer Science Conference Ranking, University of Alberta: <http://webdocs.cs.ualberta.ca/~zaiane/htmldocs/ConfRanking.html>