

Verso un'incertezza di classificazione – La cultura metrologica nella valutazione delle proprietà classificatorie

Luca Mari, Claudio Narduzzi

Tutto_Misure, 2, 2019

[6.5.19]

Il *Joint Committee for Guides in Metrology* (JCGM) ha recentemente invitato il suo gruppo di lavoro 2 a finalizzare la bozza di una nuova edizione del *Vocabolario Internazionale di Metrologia* (VIM) (si vedano le “News from JCGM/WG2” all’indirizzo www.bipm.org/utls/en/pdf/news_jcgm-wg2.pdf). La principale novità attesa per il VIM4 è un nuovo capitolo, interamente dedicato alle proprietà classificatorie e alla loro valutazione. Ricordiamo prima di tutto di cosa si tratta. Una proprietà classificatoria (“nominal property” nell’edizione inglese e “propriété qualitative” nell’edizione francese del VIM) è una proprietà che consente solo di classificare le entità che la possiedono, ma non di stabilire un ordine e nemmeno una distanza tra esse. Un esempio è la forma delle macchie di inchiostro su una pagina, interpretate come caratteri alfanumerici da riconoscere (il compito dei sistemi software di OCR, *Optical Character Recognition*). Alla conclusione del processo di riconoscimento ogni forma è riconosciuta come un certo carattere, ma questo riconoscimento è appunto solo una classificazione: si stabilisce se la forma delle macchie è tale che due certe forme sono nella stessa classe o no – cioè se il carattere scritto è lo stesso o no – ma non se una forma sia maggiore di un'altra, o quanto due forme distino tra loro. Un altro caso di proprietà classificatoria è il gruppo sanguigno degli esseri umani, classificato per esempio mediante il sistema AB0 nelle quattro classi identificate dai valori 0, A, B e AB.

Astraendo a sufficienza dai dettagli, emerge l’analogia con la misurazione. Per esempio:

Nella	misurazione di masse	classificazione di forme di caratteri
si parte da	oggetti dotati di massa	oggetti dotati di forma
e se ne considera una proprietà, e cioè	la massa.	la forma.
Sono disponibili oggetti campione per la proprietà, e cioè	un insieme di masse campione,	un insieme di forme campione,
a ognuno dei quali è associato un valore per la proprietà, e cioè	a ogni massa campione è associato un valore di massa (1 kg, 2 kg, ...),	a ogni forma campione è associato un valore di forma (“a”, “b”, ...),
in cui il valore è un elemento di un insieme predefinito, e cioè	un insieme di multipli e sottomultipli del kilogrammo.	un insieme di caratteri, {“a”, “b”, ...}.
Si allora confronta la proprietà dell’oggetto in esame con le proprietà dei campioni, e dunque	si confronta la massa del corpo con le masse campione	si confronta la forma della macchia con le forme campione
fino a trovare il campione indistinguibile dall’oggetto per la proprietà, e dunque	si trova una certa massa campione.	si trova una certa forma campione.
Si attribuisce infine all’oggetto il valore della proprietà del campione	si attribuisce alla massa del corpo un certo valore di massa (1 kg,	si attribuisce alla forma della macchia un certo valore di forma

trovato, e dunque	oppure 2 kg, ...),	("a", oppure "b", ...),
così che il risultato del processo si può scrivere per esempio come	massa(corpo) = 1,234 kg	forma(macchia) = "b"

Questo pare sufficiente per spiegare l'invito del JCGM: nonostante le loro differenze, le analogie tra la misurazione – per come intesa tradizionalmente – e la classificazione sono tali da giustificare un lavoro finalizzato a rendere possibile l'adozione della cultura metrologica nel contesto della valutazione delle proprietà classificatorie, e ciò a partire dal vocabolario, dunque. Sull'importanza di questa adozione non ci dovrebbero essere dubbi: la metrologia è, fondamentalmente, la scienza della qualità dei dati acquisiti per via empirica, e poter stabilire la qualità dei risultati di classificazioni come, per esempio, le diagnosi mediche è evidentemente importante. Parafrasiamo al proposito il paragrafo di apertura della *Guida all'espressione dell'incertezza di misura* (GUM), con sottolineate le parti in cui i riferimenti alla misurazione sono diventati riferimenti alla valutazione di proprietà classificatorie, qui chiamate "classificazioni" per brevità: "Nel riportare il risultato di una classificazione, è obbligatorio fornire una qualche indicazione quantitativa della qualità del risultato, cosicché gli utenti ne possano accertare l'attendibilità. Senza tale indicazione i risultati della classificazione non possono essere confrontati né tra di loro, né con valori di riferimento assegnati da specifiche o norme. È pertanto necessario che esista una procedura, di agevole comprensione e applicazione, nonché generalmente accettata, per caratterizzare la qualità del risultato di una classificazione". Come si vede, lo spirito di questa dichiarazione è perfettamente trasferibile dalla misurazione alla valutazione di proprietà classificatorie: nel riconoscimento della "obbligatorietà" rimarcata dalla GUM nulla è specifico alla condizione che le valutazioni siano quantitative invece che qualitative.

Uno degli insegnamenti di base della metrologia è che è la struttura dell'intero sistema metrologico a garantire la qualità dei risultati di misura: non solo un appropriato strumento di misura, ma l'intera catena di riferibilità che, a partire dalla definizione dell'unità e fino alla taratura dello strumento, rende riferibili i risultati all'unità. Certo, il sistema metrologico non è utilizzabile come tale nel caso delle proprietà classificatorie, per le quali è necessario in particolare definire separatamente ogni proprietà di riferimento invece che solo l'unità (nell'esempio dell'OCR, occorre dunque definire separatamente ogni carattere). Ma la struttura del sistema non dipende dalla struttura delle proprietà considerate, e dunque anche a proposito della possibilità di trattare di campioni, taratura, riferibilità, ecc per proprietà classificatorie l'invito del JCGM pare avere molto senso.

In tutto ciò, la frase finale del paragrafo di apertura della GUM assume una rilevanza particolare, una volta riportata nella sua completezza: "È pertanto necessario che esista una procedura, di agevole comprensione e applicazione, nonché generalmente accettata, per caratterizzare la qualità del risultato di una misurazione / classificazione, vale a dire, per valutarne ed esprimerne l'incertezza.". L'idea è, di principio, ovvia:

– primo, la forma generale del risultato di una classificazione è una distribuzione di probabilità sull'insieme dei valori possibili, così che, per esempio, la forma di una certa macchia di inchiostro su una pagina potrebbe essere riconosciuta non come un (solo) carattere, ma come una distribuzione di probabilità sull'insieme dei caratteri, dunque un insieme di coppie (carattere, probabilità) come {"a", 0,80}, {"d", 0,15}, {"o", 0,05}, (ogni altro carattere, 0,00)}, con il significato che il carattere esaminato è molto probabilmente una "a" ma non si esclude che possa essere una "d" oppure, benché ancora meno probabilmente, una "o";

– secondo, ha senso sintetizzare il risultato della classificazione in un singolo valore, e dunque l'intera distribuzione in un solo carattere, ma in tal caso è opportuno fornire anche un'informazione sull'affidabilità che si attribuisce a questo valore; si sceglie tipicamente la moda della distribuzione, cioè il suo valore più probabile, ma mentre nel caso precedente la scelta della "a" non lascia troppi dubbi, se il risultato della classificazione fosse {"a", 0,55}, {"d", 0,45}, (ogni altro carattere, 0,00)} la sintesi sarebbe più controversa, benché formalmente anche in questo caso la distribuzione sia unimodale: l'alternativa tra "a" e "d" sarebbe infatti ben più problematica di prima.

Potremmo dunque dotarci di un'*incertezza di classificazione*, che, in analogia con l'incertezza di misura, fornisca un indice di qualità dell'informazione che si riporta come risultato della classificazione stessa. Il concetto assume un rilievo particolare se riferito ai sistemi ad apprendimento automatico (*machine learning*), in considerazione del ruolo crescente che essi rivestono in molte attività che richiedono decisioni basate sulla classificazione. Questo mette ulteriormente in risalto l'attualità del tema, che si estende oltre l'ambito della comunità metrologica tradizionale.

Nelle indagini preliminari che, in quanto componenti del JCGM/WG2, abbiamo svolto intorno alla possibilità di introdurre nel VIM un concetto di incertezza di classificazione ci sono state rivolte in particolare due obiezioni, che commentiamo qui.

La prima obiezione può essere formulata più o meno così: dato che la misura di probabilità ha una struttura ordinale, e in effetti additiva su sottoinsiemi / eventi disgiunti, come può essere impiegata per proprietà classificatorie, che invece non sono nemmeno ordinate? L'esempio proposto sopra a riguardo dell'OCR dovrebbe essere sufficiente a mostrare l'infondatezza di questa obiezione: semplicemente, il fatto di riconoscere la "a" come più probabile della "z" non ha nulla a che fare con le relazioni tra i caratteri da riconoscere. E infatti è ben noto che di una qualsiasi distribuzione di probabilità $\{p_i\}$ è calcolabile l'entropia à la Shannon, $-\sum_i p_i \log(p_i)$, che misura la quantità di informazione che si ottiene scegliendo un valore dalla distribuzione.

La seconda obiezione è più sottile. Ci è stato fatto notare che da tanto tempo si realizzano classificazioni e che, pur con eccezioni, generalmente non emerge la necessità di riportare un indice della qualità dei loro risultati (insomma, il risultato dell'esame del gruppo sanguigno è un gruppo, non una distribuzione di probabilità): perché ce ne dovremmo preoccupare ora, dunque? Le risposte possibili, non alternative, sono varie.

Ci si potrebbe rifare alla storia della misurazione, praticata per secoli in assenza di una "teoria degli errori", che infatti fu sviluppata solo all'inizio del XIX secolo: non siamo costretti a trattare di incertezza di misura (anche se – come abbiamo ricordato – la GUM parla di "obbligatorietà"...), ma dubitiamo che con ciò qualcuno rimpianga un passato di illusorie certezze.

Si potrebbe considerare che nella vita quotidiana anche i risultati della gran parte delle misurazioni sono riportati come singoli valori e senza alcuna indicazione esplicita circa la loro incertezza, solo affidandosi all'informazione implicitamente presente nel numero di cifre significative del valore riportato (lo ammette perfino il VIM, in una nota alla definizione di 'risultato di misura': "Qualora l'incertezza di misura sia considerata trascurabile per un determinato scopo, il risultato di misura può essere espresso indicando il solo valore misurato. In molti campi, questo è un modo comune per esprimere il risultato di misura."). Per analogia, nel caso delle proprietà classificatorie sarebbe sufficiente scegliere un insieme di classi non troppo fine per trovarsi generalmente nella condizione di poter produrre un solo valore, in pratica certo, come risultato della classificazione. Ciò ha senso, ma non pare sufficiente: come trattare i casi in cui si volessero invece adottare classificazioni più fini, in cui l'informazione acquista sperimentalmente potrebbe non essere sufficiente a identificare una sola classe con sufficiente certezza?

Più interessante è un'ulteriore posizione: introdurre un'incertezza di classificazione non sarebbe necessario perché nella pratica si scelgono metodi di classificazione che conducono a risultati non equivoci, e quando un metodo non è sufficiente per discriminare tra due o più classi se ne adotta un altro, fino a ottenere la classificazione in una singola classe con un tale grado di certezza da rendere inutile un'indicazione dell'incertezza di tale risultato. Una tale procedura iterativa ha, evidentemente, molto senso ma, altrettanto evidentemente, ha dei costi, che l'introduzione di un'incertezza di classificazione consentirebbe di gestire attraverso un nuovo grado di libertà: si potrebbe ridurre tali costi accettando di produrre risultati la cui incertezza è dichiarata in modo esplicito, ed è minore di una data incertezza obiettivo.

Questo è l'humus culturale in cui, a proposito delle proprietà classificatorie e della loro valutazione, si sta sviluppando la bozza della prossima edizione del VIM. Si tratta però ancora di "lavori in corso". I pareri che i lettori ci volessero far giungere sarebbero preziosi.