

Un algoritmo contro le adulterazioni

Una nuova tecnica potrebbe semplificare i controlli

di **Emanuela Giorgi**

Coordinamento redazionale "Alimenti&Bevande"

Due ricercatori dell'Università di Milano-Bicocca, coinvolti in un team di ricerca internazionale, hanno messo a punto una tecnica che potrebbe diminuire il numero di controlli da effettuare per rilevare alimenti adulterati, riducendo tempi e costi. Possibili vantaggi anche per il mondo delle certificazioni

18

Rendere più semplici i controlli per garantire che sulle tavole dei consumatori arrivino alimenti che corrispondono, per qualità e origine, a quanto indicato in etichetta. È l'obiettivo di un lavoro di ricerca condotto da due ricercatori del Dipartimento di Statistica e Metodi quantitativi dell'Università di Milano-Bicocca, Francesca Greselin e Andrea Cappozzo, in collaborazione con i colleghi Ludovic Duponchel dell'Università di Lille (Francia) e Brendan Murphy dell'University College Dublin (Irlanda). I primi risultati, descritti in uno studio dal titolo



Francesca Greselin, professore associato dell'Università di Milano-Bicocca

"Robust variable selection in the framework of classification with label noise and outliers: Applications to spectroscopic data in agri-food", pubblicato da "Analytica Chimica Acta", rivista di chimica analitica e spettroscopia, sono promettenti.

"Negli ultimi decenni – si legge in un comunicato dell'Università di Milano-Bicocca – l'utilizzo della spettroscopia negli studi di "Food Authenticity" ha consentito di analizzare le sostanze senza danneggiare il campione sottoposto a verifica. Grazie all'utilizzo di sistemi di "Machine Learning", poi, è stato possibile semplificare l'analisi della grande mole di dati raccolti. Un ulteriore passo in avanti è quello frutto dello studio condotto dal team internazionale di ricercatori che hanno "testato"



Andrea Cappozzo, assegnista di ricerca dell'Università di Milano-Bicocca

la metodologia su tre diverse tipologie di prodotti: lieviti, carne e olio. La tecnica messa a punto, infatti, consente di ridurre dall'ordine delle migliaia a quello delle decine il numero di misurazioni da acquisire dal segnale spettrometrico per un'accurata verifica che escluda adulterazioni delle sostanze. Tutto ciò con evidenti vantaggi sia in ordine di tempo che di costo delle operazioni di controllo. L'impiego di moderne tecniche di spettroscopia e "Machine Learning" nel settore agroalimentare

aiuterà ad automatizzare i controlli dei cibi che entrano nelle nostre case, per assicurare maggiore qualità e sicurezza per consumatori. Tali metodologie, infatti, potranno trovare applicazione sia nell'ambito delle verifiche condotte dalle autorità governative, sia nelle procedure di certificazione di qualità dei prodotti".

Per saperne di più, abbiamo rivolto alcune domande ai due ricercatori del Dipartimento di Statistica e Metodi quantitativi dell'Università di Milano-Bicocca: Francesca Greselin (professore associato) e Andrea Cappozzo (assegnista di ricerca).

• **Com'è nata l'idea di condurre una ricerca su questa tematica?**

Cappozzo: Sin dall'inizio del mio percorso di dottorato, la mia relatrice, la professoressa Francesca Greselin, mi ha indirizzato nello studio di alcuni articoli di metodologie statistiche che risultavano efficacemente applicate all'ambito della *Food Authenticity*. In particolare, molti *papers* rilevanti in questo ambito sono stati scritti dal professor Brendan Murphy dell'University College Dublin (UCD), che da tempo si occupa di queste problematiche sia dal punto di vista metodologico che applicato. Vedendo che il mio interesse per questo





Ludovic Duponchel, Università di Lille (Francia)

tipo di argomento cresceva, la mia relatrice ha preso contatti con il professor Murphy di modo che potessi trascorrere un anno come "visiting PhD student" presso lo UCD. Grazie a questa opportunità si è instaurato un solido rapporto di ricerca, tant'è che il professor Murphy è diventato correlatore della mia tesi di dottorato ed assieme alla professoressa Greselin abbiamo pubblicato quattro articoli su riviste scientifiche di alta rilevanza.

• **Che tipo di collaborazione si è instaurata con il professor Murphy e con il professor Duponchel?**

Cappozzo: Come già menzionato, la collaborazione con il professor Murphy è nata dalla sua diretta supervisione durante il mio percorso di dottorato. Quella con il professor Duponchel, al contrario, ha una storia molto più bizzarra, che dimostra come ogni occasione sia buona per instaurare delle efficaci relazioni scientifiche. Durante una conferenza stavo presentando gli ultimi sviluppi metodologici della mia tesi quando un professore dell'Università di Nizza mi si avvicina dicendo: *"Conosco un collega chemiometrico dell'Università di Lille che sarebbe interessato ad un metodo di classificazione che possa funzionare anche a partire da informazione parzialmente errata: lui si occupa di dati spettrometrici e sicuramente sarebbe un ottimo contesto applicativo per la tua metodologia"*. Un'e-mail ed un successivo colloquio via Skype sono stati sufficienti per iniziare un'attiva collaborazione.

• **In cosa consiste la tecnica della spettroscopia? E quella del *Machine Learning*?**

Greselin: La spettroscopia è quella scienza che studia lo spettro elettromagnetico di una sostanza. In dettaglio, attraverso uno strumento chiamato spettrometro è possibile misurare, dato un campione, il livello di intensità di assorbimento della luce in funzione della lunghezza d'onda. In questo modo, per ogni *sample* analizzato si ottiene uno spettro, ovvero un segnale formato da migliaia di misurazioni, al variare – appunto – della lunghezza d'onda.

L'obiettivo del *machine learning* è quello di sviluppare una metodologia (o algoritmo) che sia in grado di "comprendere" in modo autonomo l'informazione contenuta in questi segnali eterogenei ed usarla, ad esempio, per identificare strutture simili (*clustering* o classificazione) o scoprire segnali inconsueti (anomali o *novelty detection*).

• **Quali sono i risultati della ricerca?**

Greselin: Abbiamo sviluppato un algoritmo robusto di selezione delle variabili per problemi di classificazione ad alta dimensionalità. Il fondamento logico che è alla base del metodo è il seguente: lo spettrometro raccoglie migliaia di misurazioni per ogni campione; tuttavia, al fine dell'obiettivo che ci si propone (in questo caso classificare tipologie di alimenti provenienti dallo stesso gruppo), la maggior parte dell'informazione raccolta dallo strumento è ridondante.

Il nostro modello è riuscito a presentare risultati di classificazione superiori a moderni metodi di *Machine Learning*

Ci si prefigge quindi di selezionare le sole misurazioni che meglio distinguono i vari tipi di alimenti. Questo processo, ben noto nella



Brendan Murphy, University College Dublin (Irlanda)

ricerca statistica, è chiamato *variable o feature selection*. La novità da noi introdotta è stata quella di proporre un algoritmo che non solo facesse selezione delle variabili, ma che fosse anche efficace in quelle situazioni (finora non affrontate in chemiometria) di etichette errate e/o segnali anomali, che possono seriamente compromettere la qualità della classificazione. Questa situazione è particolarmente rilevante in contesti delicati come la *Food Authenticity*, dove per l'appunto è lecito aspettarsi che ci siano dei campioni contraffatti o erroneamente classificati. Tre applicazioni su dati agroalimentari (discriminazione di lieviti, carni e olio) hanno dimostrato l'efficacia della nostra proposta. In particolare, nelle situazioni in cui i dati presentavano delle contaminazioni, il modello è riuscito non solo a selezionare lunghezze d'onda rilevanti senza risultare distorto dai valori anomali, ma anche a presentare risultati di classificazione superiori a moderni metodi di *Machine Learning* come

Support Vector Machine e *Partial Least Squares*.

• **Le vostre scoperte come potrebbero impattare, in concreto, sui controlli delle autorità? E su quelli condotti nelle procedure di certificazione di qualità dei prodotti?**

Cappozzo: Ad oggi, la metodologia è stata applicata solo a dati di laboratorio, di cui noi ricercatori avevamo il pieno controllo, e dove conoscevamo la tipologia e la percentuale di contaminazione presente nei dati. Sebbene questi primi risultati siano molto promettenti, ulteriori studi di affidabilità e generalizzazione sono necessari prima di poterne proporre un impiego industriale su larga scala.

Tuttavia, il trend tecnologico a cui stiamo assistendo è chiaramente indirizzato all'automazione. È quindi verosimile aspettarsi che la nostra (speriamo!) o altre metodologie vengano in un futuro non troppo remoto impiegate nel controllo e nella certificazione della qualità dei prodotti che finiscono sulle nostre tavole.

• **La ricerca avrà dunque un seguito?**

Greselin: Come affermato precedentemente, ora il problema principale è valutare quanto tale metodologia sia generalizzabile a situazioni diverse e più eterogenee rispetto a quanto studiato finora. In particolare, stiamo al momento cercando di capire se, perturbando l'insieme di dati del *training* (quelli usati dal modello per identificare le lunghezze d'onda più rilevanti), il nostro algoritmo sia congruente nell'identificare il set di informazioni trattenute nella classificazione che ne deriva sugli alimenti e nell'identificazione dei campioni manomessi. Rimanete sintonizzati per ulteriori sviluppi!