**Multiple instance learning**

Nel machine learning, multiple-instance learning (MIL) è un tipo di apprendimento supervisionato (supervised learning).

Le istanze di formazione sono organizzate in set, chiamati bags, e viene fornita un'etichetta per l'intera bag. La label riguarda quindi l’intero bag.

Invece di ricevere un set di istanze etichettate individualmente, il learner riceve una serie di **bags** ognuna delle quali contiene più istanze. Quindi attraverso questa tecnica si possono sfruttare dati debolmente etichettati. Imparare dai bags è una sfida unica per il MIL.

Es. Nel caso di multiple-instance binarie, una bag può essere etichettata negativamente se tutte le istanze in essa contenute sono negative. E’ sufficiente invece che una sola delle sue istanze sia positiva per etichettarla (labeled) come bag positiva.

Da una collezione di labeled bags il learner cerca di (una delle due):

* Indurre un concetto che etichetti correttamente le singole istanze
* Imparare come etichettare le bags senza indurre il concetto

Questa formulazione sta guadagnando interesse perché si adatta naturalmente a vari problemi e consente di sfruttare dati con etichettatura debole.

Babenko (2008) fornisce un semplice esempio per MIL.

Immagina diverse persone e ognuna di loro ha un portachiavi che contiene poche chiavi. Alcune di queste persone sono in grado di entrare in una certa stanza e altre no. Il compito è quindi quello di prevedere se una determinata chiave o un determinato portachiavi può farti entrare in quella stanza. Per risolvere questo problema dobbiamo trovare la chiave esatta che è comune a tutte le catene chiave "positive". Se siamo in grado di identificare correttamente questa chiave, possiamo anche classificare correttamente un'intera catena di chiavi: positiva se contiene la chiave richiesta o negativa se non contiene.

**L'apprendimento a più istanze (MIL)** rientra nel quadro di apprendimento supervisionato, in cui ogni istanza di formazione ha un'etichetta, discreta o di valore reale. MIL si occupa di problemi con una conoscenza incompleta delle etichette nei set di addestramento. Più precisamente, nell'apprendimento a più istanze, il set di formazione è costituito da "bags" etichettati, ognuno dei quali è una raccolta di istanze senza etichetta. Una borsa è etichettata positivamente se almeno un'istanza in essa è positiva ed è etichettata negativamente se tutte le istanze in essa sono negative.

L'obiettivo del MIL è quello di prevedere le etichette di borse nuove e mai viste.

Il supervisionamento riguarda solo interi set, e le singole label delle istanze delle bag non ci sono.   
Si sta sviluppando negli ultimi anni, perché si è registrato un aumento della quantità di dati necessari per affrontare grandi problemi. Grandi quantità di dati richiedono uno sforzo crescente di etichettatura e con il MIL si allevia questo onere.

In generale, l'apprendimento a istanze multiple può affrontare problemi di classificazione, problemi di regressione, problemi di classificazione e problemi di raggruppamento.

Quando le istanze sono raggruppate in bags, la predizione può essere eseguita a due livelli: bag-levels oppure instace-level. Gli algoritmi sono spesso ottimizzati per solo uno di questi due tipi di task. La composizione delle bag come la porporzione delle istanze si riperquote sulle performance dei metodi MIL.

Nell'apprendimento a istanze multiple, la rappresentazione della borsa è la tecnica che consiste nell'ottenere un unico vettore che rappresenta tutta la borsa.

**Storia**

Keeler et al., [[2]](https://en.wikipedia.org/wiki/Multiple_instance_learning#cite_note-Keeler-2) nel suo lavoro nei primi anni '90 è stato il primo ad esplorare l'area di MIL. Il termine vero e proprio apprendimento multiistanza è stato introdotto a metà degli anni '90 da Dietterich et al. mentre studiavano il problema della previsione dell'attività dei farmaci. [[3]](https://en.wikipedia.org/wiki/Multiple_instance_learning#cite_note-Dietterich-3) Hanno cercato di creare un sistema di apprendimento in grado di prevedere se una nuova molecola fosse qualificata per produrre qualche farmaco o meno, analizzando una raccolta di molecole note. Le molecole possono avere molti stati alternativi di bassa energia, ma solo uno, o alcuni di essi, sono qualificati per produrre un farmaco. Il problema è sorto perché gli scienziati potevano solo determinare se la molecola è qualificata o meno, ma non potevano dire esattamente quale delle sue forme a bassa energia ne fosse responsabile.

Uno dei modi proposti per risolvere questo problema era utilizzare l'apprendimento supervisionato e considerare tutte le forme a bassa energia della molecola qualificata come istanze di allenamento positivo, mentre tutte le forme a bassa energia delle molecole non qualificate come istanze negative. Dietterich et al. ha dimostrato che tale metodo avrebbe un alto rumore di falsi positivi, da tutte le forme a bassa energia che sono etichettate erroneamente come positive, e quindi non erano realmente utili. [[3]](https://en.wikipedia.org/wiki/Multiple_instance_learning#cite_note-Dietterich-3) Il loro approccio consisteva nel considerare ogni molecola come un sacchetto etichettato e tutte le forme alternative a bassa energia di quella molecola come istanze nel sacchetto, senza etichette individuali. Formulando così l'apprendimento a più istanze.

Soluzione al problema dell'apprendimento a istanze multiple che Dietterich et al. proposto è l'algoritmo del rettangolo parallelo all'asse (APR). [[3]](https://en.wikipedia.org/wiki/Multiple_instance_learning#cite_note-Dietterich-3) Tenta di cercare rettangoli paralleli all'asse appropriati costruiti dalla congiunzione degli elementi. Hanno testato l'algoritmo sul set di dati Musk, [[4]](https://en.wikipedia.org/wiki/Multiple_instance_learning#cite_note-Musk-4) che è un test concreto di dati sulla previsione dell'attività dei farmaci e il benchmark più comunemente utilizzato nell'apprendimento a più istanze. L'algoritmo APR ha ottenuto il miglior risultato, ma APR è stato progettato tenendo conto dei dati di Musk.

Il problema dell'apprendimento multiistanza non riguarda solo la ricerca di farmaci. Nel 1998, Maron e Ratan hanno trovato un'altra applicazione dell'apprendimento a istanze multiple per la classificazione delle scene nella visione artificiale e hanno ideato il framework Diverse Density. [[5]](https://en.wikipedia.org/wiki/Multiple_instance_learning#cite_note-Maron-5) Data un'immagine, un'istanza viene considerata come una o più sottoimmagini di dimensioni fisse e la borsa delle istanze viene considerata l'intera immagine. Un'immagine viene etichettata come positiva se contiene la scena di destinazione, ad esempio una cascata, e negativa in caso contrario. L'apprendimento di più istanze può essere utilizzato per apprendere le proprietà delle immagini secondarie che caratterizzano la scena di destinazione. Da lì in poi, questi framework sono stati applicati a un ampio spettro di applicazioni, che vanno dall'apprendimento del concetto di immagine e dalla categorizzazione del testo, alla previsione del mercato azionario.

**Esempi**

Prendiamo ad esempio la classificazione delle immagini. Amores (2013).

Data un'immagine, vogliamo conoscere la sua classe di destinazione in base al suo contenuto visivo. Ad esempio, la classe di destinazione potrebbe essere "spiaggia", dove l'immagine contiene sia "sabbia" che "acqua". In termini MIL, l'immagine è descritta come una borsa X = {X1 , .., Xn} dove ciascuno Xi è il vettore di caratteristiche (chiamato istanza ) estratto dalla corrispondente regione i-esima nell'immagine e ‘N’ è il totale delle regioni (istanze) che partizionano l'immagine.

La borsa è etichettata come positiva ("spiaggia") se contiene sia istanze della regione "sabbia" che istanze della regione "acqua".

Esempi dove viene applicato MIL sono:

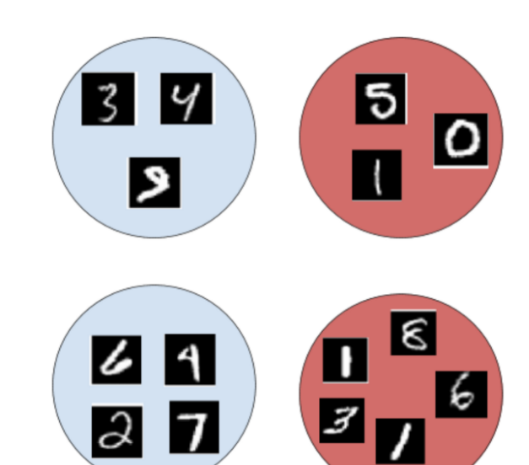
* Attività della molecola
* Previsione dei siti di legame delle proteine ​​leganti la calmodulina [6]
* Funzione di previsione per isoforme impiombate alternativamente Li, Menon & et al. (2014) , Eksi et al. (2013)
* Classificazione delle immagini Maron & Ratan (1998)
* Classificazione di testi o documenti Kotzias et al. (2015)
* Predire siti di legame funzionali di MicroRNA bersaglia Bandyopadhyay, Ghosh & et al. (2015)
* Classificazione delle immagini mediche Zhu et al. (2016), PJSudharshan et al. (2019)

Numerosi ricercatori hanno lavorato sull'adattamento delle tecniche di classificazione classiche, come le macchine a vettori di supporto o il potenziamento, per lavorare nel contesto dell'apprendimento a più istanze.

Es.2

Supponiamo di avere diverse diapositive e l'etichetta per ciascuna diapositiva. Poiché non possiamo addestrare il classificatore sull'intera diapositiva, dividiamo ogni diapositiva in tessere ed elaboriamo solo poche tessere alla volta. Tuttavia, non conosciamo le etichette per ogni riquadro, quindi abbiamo bisogno di apprendimento di istanze multiple. Nel framework MIL, le diapositive sono la "borsa" e le tessere sono le "istanze". Usandolo, siamo in grado di risparmiare lo sforzo di etichettatura e di sfruttare i dati etichettati in modo debole.

Es.3

MNIST dataset di cifre scritte a mano:

***Passaggio 1:****dividere il dataset MNIST originale nel set etichettato con borsa per un addestramento MIL appropriato e nel set etichettato come istanza.*

Per renderlo un problema MIL, dobbiamo prima costruire il dataset MIL-MNIST raggruppando diverse cifre (istanze) in un sacchetto. Nel nostro esempio MNIST, assegneremo l'etichetta bag come "1" se un'istanza ha un'etichetta "1"; se tutte le etichette di istanza sono "0–9" diverse da "1", allora assegniamo l'etichetta bag come "0".  La bag riempita di rosso ha un'etichetta "1" e la bag riempita di blu ha un'etichetta "0".

Le immagini in una bag sono inseirte casualmente ed ogni bag contiene da 3 a 7 istanze.

***Passaggio 2:****pre-training su 2 parti del set di dati MNIST. Scelta modello, addestramento modello e salva il modello.*

***Passaggio 3:****carica il modello pre-addestrato ed estrai le funzionalità dall'ultimo livello.*

1. Suddividere il resto del dataset per il training, la validation e il test
2. Get features for training, validation and test set
3. Get bag\_indices and bag\_labels
4. Map bag\_indices with features based on indices and create bag\_features

***Passaggio 4 :*** *addestrare il modello per MIL su bag\_features e bag\_labels e valutare sul set di prova.*

**Model for Multiple Instance Learning:**

L'algoritmo esegue tre passaggi. Possono essere funzioni fisse o funzioni ottimizzabili (reti neurali):

1. Trasforma le istanze in un incorporamento di bassa dimensionalità. (fisso)
2. Passa l'incorporamento attraverso una funzione di aggregazione invariante di permutazione. (ottimizzabile)
3. Si trasforma nella probabilità della borsa. (ottimizzabile)

Es. MUSK-188 è un esempio. È composto da più di una \* istanza \*.  
L'esempio nel suo insieme ha una classe. Non sappiamo quale delle  
istanze nell'esempio sia responsabile della determinazione della classe  
dell'esempio nel suo insieme. Questo è il problema di apprendimento "multiistanza".

Es.5   
Classificaione del desereto, mare e spiaggia. L’immagine del deserto contiene immagini di sabbia mentre quella del mare contiene acqua. L’immagine della spiaggia ocntiene però entrambi i segmenti. Per classificare una immagine come spiaggia il modello deve verificare la presenza di entrambi le caratteristice, quidni se si lavorasse in standars MIL assumption in questo caso si fallirebbe. Per alcuyni problemi sono necessarie più istanze positive per assegnare una label positiva alla bag.

Es.6

Nella rilevazione di un ingorgo stradale dall’immagine della strada, una macchina sarebbe una positive label per la bag. Tuttavia, servirebbero più auto per creare un ingorgo. In questa indagine, l’ipotesi collettiva designa tutte le ipotesi in cui più di un’istanza definisce le etichette delle bags.

**Definizione**

Se lo spazio delle istanze è X, quindi l'insieme delle bags è l'insieme delle funzioni Nx = {B : X -> N} che è isomorfo all'insieme dei multi-sottoinsiemi di X.  
Per ogni borsa B ∈ Nx ed ogni istanza x ∈ X, B(x) è visto come il numero di volte X che si verifica B.

Sia Y lo spazio delle etichette, quindi un "concetto di istanze multiple" è una mappa c : Nx -> Y L'obiettivo di MIL è apprendere un tale concetto.

**Presupposti**

La maggior parte del lavoro sul Multiple instance learning sta nel creare le relazioni tra le istanze in una bag e l’etichetta della classe della bag. Proprio per la sua importanza, l’assunzione e anche chiamata “standard MI assumption”

* **Standard assumption – Presupposti standard**

Il presupposto standard prende ogni istanza x ∈ X per avere un'etichetta associata y ∈ {0,1} che è nascosto al learning.

La coppia (x, y) è chiamato "concetto a livello di istanza". Una bag è ora vista come un insieme di concetti a livello di istanza ed è etichettata come positiva se almeno una delle sue istanze ha un'etichetta positiva e negativa se tutte le sue istanze hanno etichette negative.

Formalmente, se B = {(x1 , y1) … (Xn , Yn)} è una bag.   
L’etichetta di B è .

L'ipotesi MI standard è asimmetrica, il che significa che se le etichette positive e negative sono invertite, l'assunzione ha un significato diverso. Per questo motivo, quando usiamo questa ipotesi, dobbiamo essere chiari su quale etichetta dovrebbe essere quella positiva.

Tutte le bag negative contengono solo istanze negative e le positive contengono almeno una istanza negativa. Questa assunzione può essere ottenuta per affrontare problemi dove bag positive non possono essere identificate da una singola istanza, ma dall’interazione di diverse istanze.

**Algoritmi per l’apprendimento MIL**

Esistono due tipi principali di algoritmi per l'apprendimento di più istanze: algoritmi basati su istanza e basati su metadati (instance-based and metadata-based) o basati su incorporamento ( embedding-based algorithms). Il termine "basato sull'istanza" indica che l'algoritmo tenta di trovare un insieme di istanze rappresentative basate su un'ipotesi MI e di classificare le borse future da questi rappresentanti. Al contrario, gli algoritmi basati sui metadati non fanno supposizioni sulla relazione tra istanze ed etichette dei bagagli e cercano invece di estrarre informazioni (o metadati) indipendenti dall'istanza sui bagagli per apprendere il concetto. Per una rassegna di alcuni dei moderni algoritmi di MI vedi Foulds e Frank. (continua su wiki se necessario)

**Task**

**Classificazione**

La classificazione può essere eseguita a due livelli: bag e instance.   
La **bag classification** è il task più comune per gli algoritmi MIL. Consiste nell’assegnare una label di classe ad un set di istanze.

**Instance classification** è differente dalla bag classification perché mentre il training avvine usando dati organizzati in set l’obiettivo è di classificare le istanze individualmente.

Quando l'obiettivo è la bag classification, l'errata classificazione di un'istanza non influisce necessariamente sulla perdita a livello di bagaglio. Per esempio, in una borsa positiva, pochi casi veri negativi possono essere erroneamente classificati come positivi e l'etichetta della borsa rimarrà invariata.

Così la struttura del problema ed il numero di istanze di una bag giocano un ruolo fondamentale. Di conseguenza le prestazioni ottenute per la bag-classification non sono rappresentative per la prestazione ottenuta per una instance-classification. Inoltre spesso non è possibile eseguire la classificazione delle istanze

La classificazione MIL non si limita all'assegnazione di una singola etichetta a istanze o bags. L’assegnazione di più le etichette alle borse sono particolarmente rilevanti considerando che possono contenere istanze che rappresentano concetti diversi. Questa idea è stata oggetto di numerose pubblicazioni [35]. La classificazione multi-etichetta è soggetto alle stesse caratteristiche problematiche della classificazione con etichetta singola.

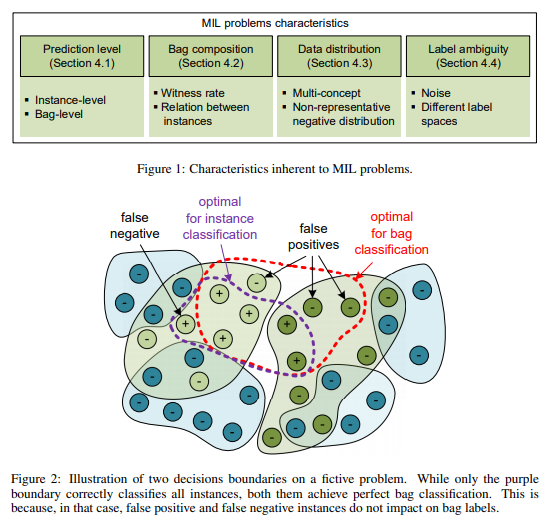
La classificazione è il task più studiato.

**Regression**

La regressione MIL consiste nell’andare ad assegnare un valore reale ad una bag (o istanza) anziché una class label. Il problema è stato affrontato in modi diversi. Alcuni metodi assegnano l'estensione etichetta della borsa basata su una singola istanza. Questo esempio può essere il più vicino a un concetto di obiettivo [36], o il miglior adattamento in un modello di regressione [37]. Altri metodi funzionano sotto l'assunzione collettiva e utilizzare la media o una combinazione ponderata delle istanze per rappresentare i sacchetti come una singola caratteristica vettore [38–40].   
In alternativa, si può semplicemente sostituire un classificatore a livello di borsa con un regressore.

**Ranking**: alcuni metodi sono stati proposti per classificare borse o istanze invece di assegnare una class label o un punteggio (ranking). Il problema è diverso dalla regressione perché l'obiettivo non è ottenere un reale esatto etichetta di valore, ma per confrontare l'entità dei punteggi per eseguire l'ordinamento. La classifica può essere eseguita a livello di bag [42] o a livello di istanza [43].

**Clustering**: questo compito consiste nel trovare gruppi o una struttura tra un insieme di borse senza etichetta. La letteratura sull'argomento è limitata. In alcuni casi, il raggruppamento viene eseguito nello spazio della borsa. In alternativa, il clustering può essere eseguito a livello di istanza. Ad esempio, in [45], l'algoritmo identifica l'istanza più rilevante di ciascuna borsa ed esegue il raggruppamento del margine massimo su queste istanze.

Abbiamo identificato quattro ampie categorie di caratteristiche chiave associate ai problemi MIL che hanno un impatto diretto sul comportamento degli algoritmi MIL: task, bag composition, data distributions and label ambiguity (come mostrato in Fig.1). Ogni caratteristica pone sfide diverse che devono essere affrontato specificamente.

**Prediction: Instance level vs Bag level**

La principale differenza tra le due attività è il costo di classificazione errata delle istanze. Sotto il presupposto MIL standard, non appena un testimone viene identificato in una borsa, viene etichettato come positivo e tutte le altre etichette di istanza possono essere ignorate. In tal caso, i falsi positivi (FP) e i falsi negativi (FN) hanno

nessun impatto sull'accuratezza della classificazione del sacchetto, ma viene comunque considerato come errore di classificazione a livello di istanza. La principale differenza tra le due attività è il costo di classificazione errata delle istanze. Inoltre, quando si considerano le bag negative, un singolo FP causa una classificazione errata di un sacchetto.   
Questo significa che se l'1% delle istanze in ciascuna borsa negativa è stata classificata in modo errato, l'accuratezza delle negative bag sarebbe 0%, anche se l'accuratezza sulle istanze negative sarebbe del 99%. Questo è illustrato in Fig. 2. Gli insiemi verdi rappresentano i sacchetti positivi, mentre i sacchetti negativi corrispondono al blu. Le singole etichette delle istanze vengono identificate su ciascuna istanza. In questa figura, entrambi i confini decisionali (linee tratteggiate) sono ottimali per la classificazione delle bag perché includono almeno una istanza da tutte le borse positive, escludendo tutte le istanze dalle borse negative. Tuttavia, solo uno dei due confini raggiunge la classificazione di istanza perfetta (viola). Questo è il motivo per cui gli algoritmi MIL utilizzando l'accuratezza della sacca come criterio di ottimizzazione (ad esempio APR [3], MI-SVM [6], MIL-Boost [54], EMDD [33], MILD [55]) è possibile apprendere un confine decisionale non ottimale per la classificazione dell'istanza.

**Bag Composition**

Il **tasso di testimonianza (witnessrate)** (WR) è la proporzione di istanze positive in sacchetti positivi. Quando il WR è molto alto i sacchetti positivi contengono solo pochi casi negativi.  
In tal caso, l'etichetta delle istanze si presume sia la stessa dell'etichetta della loro borsa. Il problema quindi ritorna a un problema supervisionato con rumore unilaterale che può essere risolto in un quadro controllato regolare [59]. Tuttavia, in alcune applicazioni, WR può essere arbitrariamente piccolo e ostacolare le prestazioni di molti algoritmi. Per esempio, in metodi come Diverse Density (DD) [29], Citation-kNN [33] e APR [3] sono considerato avere la stessa etichetta della loro borsa. Quando il WR è basso, questo non è più ragionevole e porta a prestazioni inferiori. Alcuni metodi rappresentano i sacchetti in base alla media delle istanze che contengono. Infine, inproblemi di classificazione delle istanze, WR inferiori significano gravi problemi di squilibrio di classe, che portano a cattive prestazioni per molti metodi. TODO pag 2 Relations

https://arxiv.org/pdf/1612.03365.pdf