



# Métodos de clasificación multi-etiqueta



# Métodos de clasificación multi-etiqueta

UNIVERSIDAD DE CÓRDOBA

## 1. Introducción

Como ya se introdujo en la lección anterior, existen varios enfoques para resolver problemas multi-etiqueta: transformación de problemas, adaptación de algoritmos, y ensembles de clasificadores multi-etiqueta. En la lección anterior también se introdujeron dos métodos básicos para clasificación multi-etiqueta: Binary Relevance (BR) y Label Powerset (LP). A lo largo de esta sección se presentan varios métodos para clasificación multi-etiqueta.

## 2. Classifier chains (CC)

Classifier Chains (CC) [Rea11] es un método de transformación de problemas basado en BR, donde se generan  $q$  clasificadores binarios, uno por cada etiqueta. Sin embargo, CC trata de compensar la asunción de independencia de BR encadenando los clasificadores binarios de tal forma que cada uno de ellos incluye las predicciones de los modelos anteriores como atributos de entrada adicionales (Figura 1). De este modo, al contrario que BR, CC es capaz de modelar algunas de las dependencias entre etiquetas sin introducir una complejidad demasiado alta.

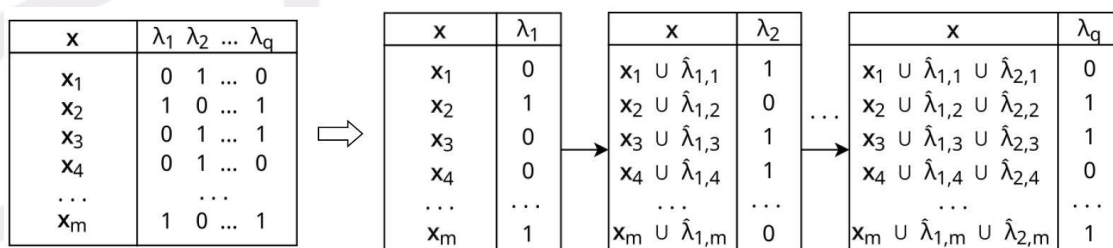


Figura 1. Ejemplificación del método Classifier Chains.

La elección del orden en que se encadenan los clasificadores es clave en el rendimiento final del clasificador. Por lo general, esta selección se realiza de modo aleatorio, aunque hay otros métodos que tratan de obtener un encadenamiento óptimo. Por otro lado, el método Ensemble of Classifier Chains (ECC) genera un ensemble donde cada clasificador base es un clasificador CC con un orden aleatorio para el encadenamiento (y por lo general, cada CC se entrena con un muestreo de los datos, con o sin reemplazamiento). La salida final de ECC se realiza por voto mayoritario para cada una de las etiquetas. Cabe destacar que ECC se ha demostrado que es uno de los métodos que mejor rendimiento obtienen en clasificación multi-etiqueta.

## 3. Pruned Sets (PS)

Pruned Sets (PS) [Bou04] es otro método de transformación de problemas, basado en LP. Al igual que LP, en un primer paso, PS transforma el problema multi-etiqueta en un problema multi-clase.

Dado que uno de los principales inconvenientes de LP es que el problema multi-clase podría tener un número excesivamente alto de clases, llegando a ser también un problema muy desbalanceado. Para tratar de solventar este problema, PS poda (es decir, elimina) aquellos patrones que pertenecen a clases muy infrecuentes. Para compensar esta pérdida de información, PS vuelve a reintroducir estos patrones en el conjunto de datos, asociados con una clase cuyo *labelset* sea un subconjunto más frecuente del original (Figura 2). Aunque el espacio de salida pueda seguir siendo bastante grande, permite reducir notablemente su complejidad.

Para PS, también existe la versión en ensemble, llamada Ensemble of Pruned Sets (EPS). En este caso, EPS sigue un enfoque de bagging, donde cada PS se entrena sobre un subconjunto de los datos de entrenamiento, y la predicción final se realiza por voto mayoritario para cada etiqueta.

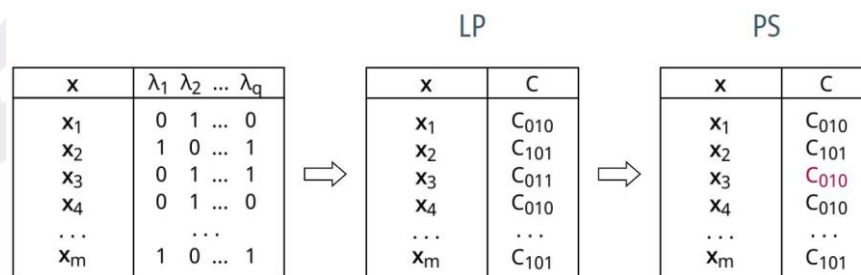


Figura 2. Ejemplificación del método Pruned Sets (y Label Powerset). En el ejemplo, consideramos que la clase  $C_{011}$  es muy poco frecuente en la transformación LP.

## 4. Random k-labelsets (RAkEL)

El método Random k-labelsets (RAkEL) [Tso10] es un método encuadrado dentro del grupo de ensembles de clasificadores multi-etiqueta. Nótese que pese a que otros métodos anteriores (como BR o CC) combinaban varios clasificadores, no se llegan a considerar métodos de ensemble en clasificación multi-etiqueta ya que los clasificadores base son para clasificación tradicional. Sin embargo, otros métodos como RAkEL (o ECC y EPS) sí que combinan varios clasificadores multi-etiqueta, de ahí que se consideren como métodos de ensemble dentro de este paradigma de aprendizaje.

RAkEL también se basa en el paradigma de transformación LP pero buscando problemas mucho menos complejos. En este caso, cada clasificador base de RAkEL selecciona un subconjunto de  $k$  etiquetas. Una vez filtrado el conjunto de datos con únicamente esas etiquetas, RAkEL entrena un LP sobre dicho conjunto. Como se observa en la Figura 3, la salida final del ensemble se obtiene por voto mayoritario.

	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$	$\lambda_7$	$\lambda_8$
MLC1	·	0	1	·	·	1	·	·
MLC2	1	·	·	·	1	1	·	·
MLC3	·	·	1	0	·	·	·	1
MLC4	·	0	·	·	·	0	0	·
MLC5	1	·	·	0	·	·	1	·
MLC6	·	·	·	0	0	1	·	·
MLC7	·	0	·	1	·	·	·	1
MLC8	1	0	0	·	·	·	·	·
MLC9	·	0	·	·	0	0	·	·
MLC10	·	1	0	·	·	·	0	·
$t=0.5$	$\frac{3}{3}$	$\frac{1}{6}$	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{3}{5}$	$\frac{1}{3}$	$\frac{2}{2}$
	1	0	1	0	0	1	0	1

Figura 3. Ejemplo de clasificación en RAKEL.

Analizando en detalle la Figura 3, podemos observar que en el ejemplo, el problema tiene un total de 8 etiquetas distintas, y el ensemble está compuesto por 10 clasificadores. Cada uno de ellos está entrenado para predecir 3 de las 8 etiquetas; por tanto, en la figura cada uno proporciona una salida binaria para las etiquetas que estaba considerando. Por ejemplo, el clasificador MLC<sub>1</sub> predice que, para un patrón dado, las etiquetas relevantes serían  $\lambda_3$  y  $\lambda_6$ . Una vez que todos los clasificadores base proporcionan salida para un patrón, se obtiene la salida final del ensemble. Por ejemplo, para la etiqueta  $\lambda_1$ , como 3 de los 3 clasificadores predicen que la etiqueta es relevante, la salida final es relevante. Sin embargo, para la etiqueta  $\lambda_2$ , como solo uno de los clasificadores de los 6 que la tienen en cuenta, predicen la etiqueta como relevante, la salida final es que  $\lambda_2$  no es relevante.

Existen principalmente dos configuraciones básicas para RAKEL: 1) 10 clasificadores en el ensemble y  $k = \frac{q}{2}$  (es decir, cada clasificador considerará un 50% de las etiquetas); 2)  $2q$  clasificadores, y  $k = 3$ .

Por lo general, RAKEL selecciona los subconjuntos de etiquetas de manera aleatoria, sin considerar en ningún modo las características de los datos; sin embargo, sí que existen diversos métodos en la literatura para seleccionar dichos subconjuntos de etiquetas teniendo en cuenta las características de los datos, como por ejemplo la relación entre las propias etiquetas.

## 5. Label specific features for multi-label learning (LIFT)

Pese a tener un subconjunto amplio de etiquetas de salida, hasta el momento todos los métodos estudiados trabajan con todo el conjunto de atributos de entrada por completo. Sin embargo, existen otros métodos que buscan los atributos o características concretas que contribuyen al correcto aprendizaje de cada una de las etiquetas.

Label specific FeaTures for multi-label learning (LIFT) [Zha14] se basa en esta idea de modelar cada etiqueta teniendo en cuenta únicamente sus atributos de entrada relacionados, es decir, sus “características específicas”. El objetivo es reducir el ruido incluido por las características que no están correladas con cada etiqueta. Para ello, LIFT usa un algoritmo de clustering en el espacio de características para seleccionar el subconjunto que mejor discrimina cada etiqueta independientemente; después, construye un clasificador binario (al estilo de BR) para cada una de las etiquetas. En la Figura 4 se puede observar el funcionamiento de LIFT, donde clasificador binario se construye también sobre un subconjunto de los atributos de entrada ( $\mathbf{x}^{\lambda_i}$ ).

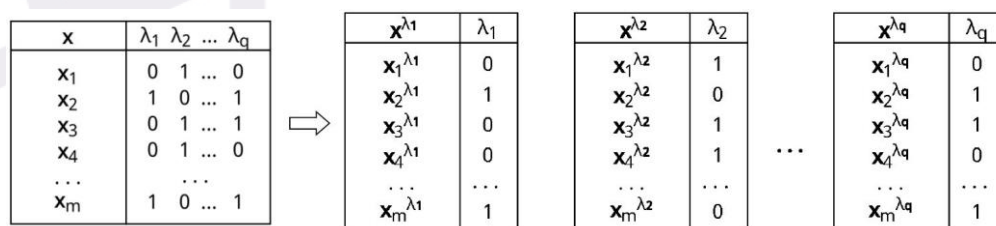


Figura 4. Ejemplo de funcionamiento de LIFT.

## 6. Otros métodos para clasificación multi-etiqueta

Además de los métodos introducidos a lo largo de esta lección, existen otras muchas propuestas para la resolución de problemas de clasificación. En esta sección se nombran algunas de ellas, sin entrar en demasiado nivel de detalle. Para más detalle de estos métodos, puede consultar los estudios realizados en [Gib14, Gib15, Moy18].

- ChiDep (también llamado LPBR). Este método de transformación de problemas se basa en generar una partición de las etiquetas en distintos grupos según su correlación. Aquellas etiquetas que queden aisladas (es decir, que no tengan relación con otras etiquetas), se modelarán en clasificadores binarios independientes del resto. Por otro lado, los grupos que contengan más de una etiqueta, indicando relación entre las etiquetas de ese grupo, se modelarán siguiendo un enfoque LP. De este modo, se busca un compromiso entre modelos simples binarios, y modelos más complejos como LP teniendo en cuenta etiquetas que tienen relación entre sí.

- Multi-Label k-Nearest Neighbors (ML-kNN). Este algoritmo adapta directamente el conocido método kNN a problemas multi-etiqueta, de modo que considera todas las etiquetas a la vez al clasificar un nuevo patrón.
- Instance-Based Logistic-Regression (IBLR-ML). Este algoritmo es una adaptación de la regresión logística a escenarios multi-etiqueta; sin embargo, también utiliza las etiquetas de los vecinos más cercanos como atributos de entrada extra en el esquema de regresión logística.
- Back-Propagation for Multi-Label Learning (BP-MLL). Aunque no se ha visto en mucho detalle las redes neuronales artificiales para clasificación en este curso, cabe destacar que el principal algoritmo existente para aprender dichas redes (el método de propagación del error), también ha sido adaptado a clasificación multi-etiqueta.
- Multi-Label Stacking (MLS). Este método de ensemble implementa un esquema de stacking. Para ello, en la primera fase entrena  $q$  modelos binarios (como BR); mientras que en la segunda fase, vuelve a entrenar  $q$  modelos binarios pero en este caso las predicciones de los  $q$  modelos anteriores se incluyen como atributos de entrada extra. Así, aplica un esquema similar a BR pero capaz de considerar las relaciones entre etiqueta.
- Hierarchy Of Multi-label ClassifiERs (HOMER). Este método de ensemble crea una estructura jerárquica de modelos. En cada nodo, el espacio de salida se divide utilizando un algoritmo de clustering, agrupando etiquetas similares, y de tal modo que en los nodos hoja de dicha estructura quedan etiquetas independientes.
- Chi-Dep Ensemble (CDE). Este método se basa en el Chi-Dep explicado anteriormente. En lugar de buscar la partición de etiquetas óptima, CDE genera un número muy grande de particiones aleatorias del espacio de salida, las evalúa teniendo en cuenta la correlación de los grupos de etiquetas generados, y se queda con un subconjunto de los que considera mejores. Posteriormente, entrena clasificadores Chi-Dep sobre cada una de las particiones seleccionadas.

## Referencias

- [Bou04] Boutell, M. R., Luo, J., Shen, X., & Brown, C. M. (2004). Learning multi-label scene classification. *Pattern recognition*, 37(9), 1757-1771.
- [Gib14] Gibaja, E., & Ventura, S. (2014). Multi-label learning: a review of the state of the art and ongoing research. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(6), 411-444.
- [Gib15] Gibaja, E., & Ventura, S. (2015). A tutorial on multilabel learning. *ACM Computing Surveys (CSUR)*, 47(3), 1-38.

- [Moy18] Moyano, J. M., Gibaja, E. L., Cios, K. J., & Ventura, S. (2018). Review of ensembles of multi-label classifiers: models, experimental study and prospects. *Information Fusion*, 44, 33-45.
- [Moy20] Moyano, J. M. (2020). Multi-label classification models for heterogeneous data: an ensemble-based approach. [Tesis doctoral, Universidad de Córdoba y Virginia Commonwealth University].
- [Rea11] Read, J., Pfahringer, B., Holmes, G., & Frank, E. (2011). Classifier chains for multi-label classification. *Machine learning*, 85(3), 333-359.
- [Tso07] Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, 3(3), 1-13.
- [Tso10] Tsoumakas, G., Katakis, I., & Vlahavas, I. (2010). Random k-labelsets for multilabel classification. *IEEE transactions on knowledge and data engineering*, 23(7), 1079-1089.
- [Zha14] Zhang, M. L., & Wu, L. (2014). Lift: Multi-label learning with label-specific features. *IEEE transactions on pattern analysis and machine intelligence*, 37(1), 107-120.