



Introducción a la clasificación multi-etiqueta



Introducción a la clasificación multi-etiqueta

UNIVERSIDAD DE CÓRDOBA

1. Introducción

Hasta el momento, a lo largo del curso hemos definido un problema de clasificación como un problema donde “dado un conjunto de instancias, cada una de ellas descrita por un conjunto de atributos de entrada, el objetivo es construir un modelo capaz de predecir la clase asociada a cada uno de los patrones”. En los problemas de clasificación vistos hasta ahora, cada instancia tiene asociada un único valor de clase de entre un conjunto de dos (problemas binarios) o más (multi-clase). Por ejemplo, tal y como se observa en la Figura 1, una planta de tipo Iris se puede clasificar en una (y solo una) de las clases disponibles.



- ☐ Setosa
- ☐ Virginica
- ☒ Versicolor

Figura 1. Ejemplificación de un problema de clasificación tradicional.

Sin embargo, existen muchos problemas del mundo real, que no se ajustan a dicha definición. Supongamos por ejemplo un problema de clasificación de imágenes, como se puede observar en la Figura 2. De entre el conjunto de clases (o etiquetas) disponibles, esta imagen podría etiquetarse como “Animal”, “Mapache”, o “Madera”; mientras que no podría etiquetarse como “León” o “Nieve”. En este caso, una imagen podría estar asociada con una o más etiquetas (incluso podría no tener ninguna etiqueta asociada), de entre un conjunto de etiquetas predefinido. Si convirtiésemos este problema a un problema de clasificación tradicional, la imagen podría clasificarse quizá solo como “Mapache”, lo que conllevaría una pérdida de información.



- ☒ Animal
- ☒ Raccoon
- ☐ Lion
- ☒ Wood
- ☐ Snow

Figura 2. Ejemplificación de un problema de clasificación multi-etiqueta.

Dado que los métodos de resolución de problemas de clasificación tradicional no pueden dar solución a estos problemas, aparece el paradigma de clasificación multi-etiqueta, cuyo objetivo es entrenar modelos capaces de asignar, para un patrón dado, su conjunto de etiquetas asociadas. Nótese que el concepto de clase y etiqueta es prácticamente el mismo, aunque se suele utilizar *clase* en clasificación tradicional, y *etiqueta* en multi-etiqueta, donde la salida es un conjunto de varias clases o etiquetas. Sí que hay que tener en cuenta, que en el escenario multi-etiqueta, todas las etiquetas son binarias (es decir, un patrón está o no, asociado con cada una de las etiquetas).

Existen otros problemas del mundo real que también pueden resolverse gracias al paradigma de clasificación multi-etiqueta. Por ejemplo, en sistemas de recomendación, un usuario podría recibir varias recomendaciones al mismo tiempo; o en marketing, un mismo cliente podría recibir publicidad centrada en varios de sus intereses. También, en diversos problemas médicos, un mismo paciente puede estar infectado o estar sufriendo más de una enfermedad al mismo tiempo, lo que lo convierte en un problema inherentemente multi-etiqueta.

Formalmente, la clasificación multi-etiqueta puede definirse como a continuación. Sea \mathcal{D} un conjunto de datos multi-etiqueta, compuesto por m instancias, cada una de ellas definida como $\mathcal{D} = \{(\mathbf{x}_i, Y_i) \mid 1 \leq i \leq m\}$. Sea $\mathcal{X} = X_1 \times \dots \times X_d$ el espacio de entrada d -dimensional, e $\mathcal{Y} = \{\lambda_1, \dots, \lambda_q\}$ el espacio de salida compuesto por $q > 1$ etiquetas. Cada instancia multi-etiqueta se compone por un vector de entrada \mathbf{x}_i y un conjunto de etiquetas relevantes asociadas con ella $Y_i \subseteq \mathcal{Y}$ (donde el conjunto de etiquetas predicho para un patrón dado se indica con \hat{Y}_i). Además, cada combinación distinta de etiquetas se suele llamar *labelset*. En la Tabla 1 se muestra un ejemplo de un conjunto de datos multi-etiqueta. Como puede observarse, cada patrón está descrito por un conjunto de atributos de entrada (como en clasificación tradicional), y está etiquetado con una o varias de las etiquetas de manera simultánea.

Tabla 1. Ejemplo de conjunto de datos multi-etiqueta.

	Atributos de entrada	Etiquetas				
		λ_1	λ_2	λ_3	...	λ_q
#1	\mathbf{x}_1	0	1	0	...	0
#2	\mathbf{x}_2	1	0	1	...	1
#3	\mathbf{x}_3	0	1	0	...	1
#4	\mathbf{x}_4	0	1	0	...	0
#5	\mathbf{x}_5	0	1	1	...	0
...				
#m	\mathbf{x}_m	1	0	1	...	1

2. Principales retos en clasificación multi-etiqueta

El hecho de tener que lidiar al mismo tiempo con varias etiquetas de salida, supone nuevos problemas o retos que los métodos para clasificación multi-etiqueta deben resolver, como modelar la relación entre etiquetas, el desbalanceo del espacio de salida, y su alta dimensionalidad.

En primer lugar, las etiquetas suelen tener cierta relación entre ellas, por lo que, si los métodos de clasificación son capaces de modelar dichas dependencias, previsiblemente podrán obtener un mejor rendimiento predictivo. Supongamos, por ejemplo, un problema de categorización de imágenes. Previsiblemente, las etiquetas “bebé” y “juguete” estarán bastante relacionadas entre ellas, mientras que las etiquetas “bebé” y “cocodrilo” también podrían tener una alta correlación inversa (es decir, por lo general, si aparece la etiqueta bebé, no va a aparecer cocodrilo). Sin embargo, también puede haber otras etiquetas, como “bebé” y “nubes” que no presenten ninguna correlación, pues una de ellas no parece que pueda influenciar el modelado de la otra.

Por otro lado, las etiquetas por lo general no aparecen con la misma frecuencia en el conjunto de datos. Por ejemplo, es muy posible que una etiqueta “animal” aparezca con más frecuencia que las etiquetas “mapache” o “cocodrilo”; o en el caso de diagnóstico de enfermedades, puede haber enfermedades que sean más comunes (y aparezcan con más frecuencia en el conjunto de datos) y otras que sean mucho menos frecuentes, apareciendo en muy pocos patrones. Esto, puede llevar a que los métodos se centren en modelar únicamente (o les den más peso a) las etiquetas que aparecen con mayor frecuencia en el conjunto de datos, o que aquellas cuya frecuencia de aparición es muy baja no sean correctamente modeladas.

Por último, el inherente crecimiento del espacio de salida (con respecto a problemas de clasificación tradicional), donde se debe proporcionar salida (positiva o negativa) para todas las posibles etiquetas, puede hacer que los problemas sean intratables desde ciertos enfoques. Por ejemplo, sería mucho más sencillo aprender la dependencia o relación entre etiquetas en casos donde existen 6 etiquetas, que en problemas donde tenemos cientos de etiquetas.

3. Primeros enfoques para resolución de problemas de clasificación multi-etiqueta

Los métodos de clasificación multi-etiqueta suelen categorizarse dentro de 3 grupos principales: transformación de problemas, adaptación de algoritmos, o ensembles de clasificadores multi-etiqueta. Los primeros, resuelven el problema multi-etiqueta gracias a convertirlo en uno o varios problemas de clasificación tradicional, utilizando posteriormente algoritmos de resolución clásicos.

Los segundos, adaptan directamente algoritmos de clasificación tradicional para que sean capaces de trabajar directamente con datos multi-etiqueta, sin necesidad de transformar el problema. Por último, los ensembles de clasificadores multi-etiqueta combinan las salidas de varios clasificadores multi-etiqueta, independientemente del tipo que sean.

En esta sección se presentarán dos de los primeros enfoques que se propusieron para la resolución de problemas multi-etiqueta: *Binary Relevance* (BR), y *Label Powerset* (LP), ambos encuadrados dentro del grupo de transformación de problemas.

3.1. Binary Relevance (BR)

El método de Binary Relevance es uno de los más básicos para resolver problemas de clasificación multi-etiqueta. Para resolver el problema multi-etiqueta, BR descompone el problema en q problemas de clasificación binarios independientes, uno por cada etiqueta, como se observa en la Figura 3.

Para cada uno de los nuevos problemas binarios, BR utiliza un método de clasificación tradicional para resolverlo, generando q modelos. La predicción multi-etiqueta final viene dada por la unión de las salidas de cada uno de los clasificadores binarios. El hecho de que BR trate las etiquetas por separado, hace que sea simple y altamente paralelizable; sin embargo, no tiene en cuenta las combinaciones de etiquetas, lo que le hace incapaz de modelar las dependencias entre ellas.

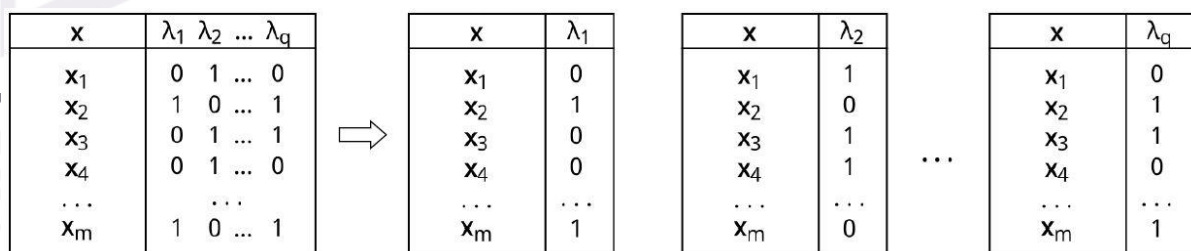


Figura 3. Transformación Binary Relevance.

3.2. Label Powerset (LP)

El método de Label Powerset transforma el problema multi-etiqueta en un único problema multi-clase, donde cada posible combinación de etiquetas (o *labelset*) en el conjunto de entrenamiento se considera como una clase distinta, como se observa en la Figura 4.

Dado el conjunto de datos transformado, LP utiliza cualquier método de clasificación tradicional para problemas multi-clase. La predicción final se obtiene descomponiendo la clase predicha en la combinación de etiquetas correspondiente. LP considera las relaciones entre todas las etiquetas, pero su complejidad crece exponencialmente con el número de etiquetas. Además, no es capaz de predecir combinaciones de etiquetas que no aparecían en entrenamiento, y dado que es muy posible que algunos (o muchos) *labelset* tengan asociados un pequeño número de patrones de entrenamiento, llevaría a problemas muy desbalanceados.

x	λ_1	λ_2	...	λ_q		x	C
x_1	0	1	...	0		x_1	C_{010}
x_2	1	0	...	1		x_2	C_{101}
x_3	0	1	...	1		x_3	C_{011}
x_4	0	1	...	0		x_4	C_{010}
...		
x_m	1	0	...	1		x_m	C_{101}

Figura 4. Transformación Label Powerset.

4. Métricas de evaluación

Dado que en clasificación multi-etiqueta un patrón puede estar asociada con varias etiquetas simultáneamente, las predicciones pueden considerarse como totalmente correctas (la predicción para todas las etiquetas es correcta), totalmente incorrecta (todas las predicciones son erróneas), o parcialmente correctas (solo algunas de las etiquetas relevantes han sido predichas, y/o algunas etiquetas no relevantes han sido predichas). Como consecuencia, es necesario definir métricas de evaluación que se ajusten a las predicciones multi-etiqueta. Estas métricas se agrupan comúnmente en métricas basadas en ejemplos y en etiquetas.

4.1. Métricas basadas en ejemplos

Las métricas basadas en ejemplos calculan la métrica para cada patrón de los datos, y después, promedia los resultados. De este modo, el cálculo da el mismo peso a todas las instancias, por lo que los resultados no están sesgados por el desbalanceo de los datos. A continuación, se detallan algunas de las métricas basadas en ejemplos más importantes.

Hamming loss (HL) calcula el promedio de veces que cada etiqueta se predice incorrectamente, incluyendo tanto errores de predicción (etiquetas irrelevantes se predijeron como relevantes), como de omisión (una etiqueta relevante no fue predicha). Al medir el error en predicción, es una métrica a minimizar. Pese a ser utilizada con bastante frecuencia, HL tiene ciertos inconvenientes,

principalmente en casos donde existen muchas etiquetas, pero muy pocas de ellas suelen ser relevantes, haciendo que la métrica tienda a 0. A continuación se define la métrica, donde Δ indica la diferencia simétrica entre dos conjuntos binarios.

$$\downarrow \text{HL} = \frac{1}{m} \sum_{i=1}^m \frac{1}{q} |Y_i \Delta \hat{Y}_i|$$

Subset Accuracy (SA) es una métrica estricta, que mide el ratio de instancias cuya predicción coincide exactamente con el conjunto de etiquetas relevantes reales. Dado que SA tiene en cuenta las predicciones que son completamente correctas y no las parcialmente correctas, debe utilizarse con cuidado y por lo general combinada con la obtención de otras métricas que nos ofrezcan una visión más global del rendimiento del clasificador. A continuación, se define la ecuación de la métrica, donde $[[\pi]]$ devuelve 1 si el predicado o condición π es correcto, y 0 en caso contrario.

$$\downarrow \text{SA} = \frac{1}{m} \sum_{i=1}^m [[Y_i = \hat{Y}_i]]$$

Por otro lado, también se han adaptado las métricas de clasificación tradicional al escenario multi-etiqueta en el enfoque basado en ejemplos. A continuación, se detallan las ecuaciones para el cálculo de precision (ExP), recall (ExR), F1-Score (ExF) y specificity (ExS). Nótese que \bar{Y} indica el conjunto de etiquetas irrelevantes (o negativas).

$$\uparrow \text{ExP} = \frac{1}{m} \sum_{i=1}^m \frac{|Y_i \cap \hat{Y}_i|}{|\hat{Y}_i|}$$

$$\uparrow \text{ExR} = \frac{1}{m} \sum_{i=1}^m \frac{|Y_i \cap \hat{Y}_i|}{|Y_i|}$$

$$\uparrow \text{ExF} = \frac{1}{m} \sum_{i=1}^m \frac{2 \cdot |Y_i \cap \hat{Y}_i|}{|Y_i| + |\hat{Y}_i|}$$

$$\uparrow \text{ExS} = \frac{1}{m} \sum_{i=1}^m \frac{|\bar{Y}_i \cap \bar{\hat{Y}}_i|}{|\bar{Y}_i|}$$

4.2. Métricas basadas en etiquetas

Las métricas basadas en etiquetas calculan las métricas de acuerdo a las etiquetas en lugar de a las instancias, y dado que se basan en las matrices de confusión, permiten calcular cualquier métrica definida para clasificación binaria. A su vez, estas métricas pueden calcularse siguiendo dos enfoques: enfoque micro y enfoque macro.

Supongamos que B es una métrica de evaluación para clasificación binaria basada en los verdaderos positivos (TP), falsos positivos (FP), falsos negativos (FN), y verdaderos negativos (TN) de la matriz de confusión. El enfoque micro, primero combina las matrices de confusión de todas las etiquetas, y después calcula la métrica correspondiente. Por otro lado, el enfoque macro calcula la métrica basándose en la matriz de confusión de cada una de las etiquetas, y posteriormente promedia los valores de la métrica entre todas las etiquetas. Como consecuencia, las métricas en el enfoque micro estarán más sesgadas por las etiquetas más frecuentes (tendrán más peso en la matriz de confusión combinada, y por tanto en la métrica final), mientras que el enfoque macro les da el mismo peso a todas las etiquetas, independientemente de su frecuencia de aparición en el conjunto de datos. Este segundo enfoque puede ser más útil en problemas donde las etiquetas poco frecuentes sean interesantes (como por ejemplo en diagnóstico de enfermedades raras o poco frecuentes), ya que la métrica calculada las tendrá en cuenta y no las despreciará (como sí puede hacer el enfoque micro).

Referencias

- [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.
- [Moy20] Moyano, J. M. (2020). Multi-label classification models for heterogeneous data: an ensemble-based approach. [PhD Thesis, University of Córdoba and Virginia Commonwealth University].
- [Tso07] Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, 3(3), 1-13.