

Aux 10. Introducción a la Minería de Datos

Gastón L'Huillier^{1,2}, Richard Weber²
glhuilli@dcc.uchile.cl

¹Departamento de Ciencias de la Computación
Universidad de Chile

²Departamento de Ingeniería Industrial
Universidad de Chile

2010



Múltiples Clasificadores

Problema

Tenemos **muchas formas de Clasificar** y sabemos que cada una puede entregar **información importante** para distintas aplicaciones (o instancias de estas aplicaciones).

En general, la idea es **no quedarnos con una sola técnica**, aunque sea muy buena (que nos entregue los mejores resultados de acuerdo a una instancia del problema), o la más efectiva (presentando la oportunidad de tener un modelo robusto ante nuevos elementos a clasificar).

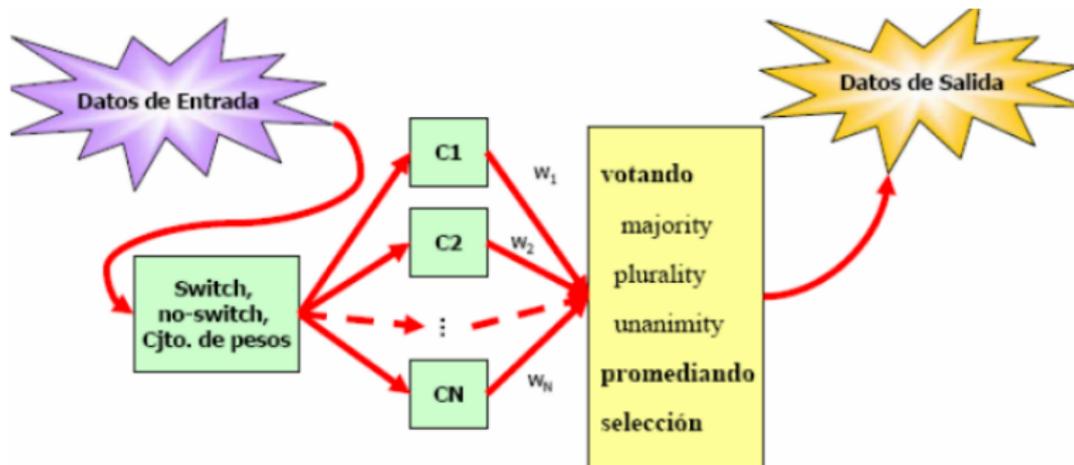
- 1 Idea 1: **Explotar las características** de cada método de manera **independiente** para obtener mejores resultados.
- 2 Idea 2: **Combinar** los resultados de cada método de la manera adecuada.



Múltiples Clasificadores [2]

Problema

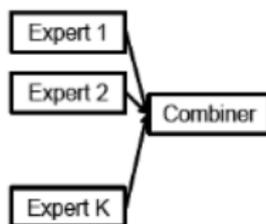
Tenemos **muchas formas de Clasificar** y sabemos que cada una puede entregar **información importante** para distintas aplicaciones (o instancias de estas aplicaciones).



Estrategias

- 1 **Subsampling de las observaciones** de entrenamiento mediante técnicas de resampling (*Boosting, Bagging*).
- 2 Manipulación de la **selección de atributos para entrenar distintos modelos** con conjuntos de atributos distintos.
- 3 Manipulación de la variable dependiente de manera de buscar **solución a problemas representados por distintos valores objetivos**.
- 4 Modificación de los **parámetros del clasificador de manera de obtener distintos modelos** asociados a un conjunto de entrenamiento dado.
- 5 **Diversificación de modelos** a utilizar para determinar los valores asociados a la solución del problema objetivo (*Majority voting, etc.*)

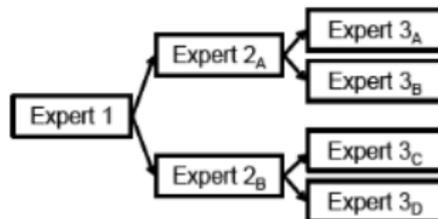
Paralelos



En cascada



Jerárquicos



Estructura Múltiples Clasificadores [2]

Paralelos

Todos los clasificadores son **invocados independientemente** y sus resultados son combinados a través de algún criterio adecuado.

En cascada

Clasificadores son llamados de manera **secuencial**.

Jerárquicos

Clasificadores son llamados a través de una **estructura “de árbol”** definida por la jerarquía.



Estructura Múltiples Clasificadores [2]

Estáticos

- **No entrenables:** La votación se realiza de manera independiente y no paramétrica c/r al desempeño de cada clasificador. (*Majority voting, promedio de resultados, etc.*).
- **Entrenables o Ajustables:** El “combinador” inicia una nueva fase de entrenamiento de manera de mejorar el desempeño general de los modelos generados. (*nuevos modelos, MIP, etc.*)

Adaptativos

La función que define al “combinador” depende de los atributos iniciales considerados para los distintos modelos



Subsampling del Conjunto de Entrenamiento

Bagging: Bootstrap AGgregation [Breiman, 1996]

Crea un **multclasificador entrenando modelos individuales** en conjuntos de muestras que provienen de la técnica de **resampling "bootstrap"** como conjunto de entrenamiento.

Bagging: La idea... [Breiman, 1996]

- Dado un conjunto de entrenamiento con N observaciones, Bagging genera m conjuntos $C_j, j = \{1, \dots, m\}$, donde $|C_j| = n \leq N$. Estos conjuntos se generan utilizando **selección uniforme con reemplazo**.
- N observaciones tienen la **probabilidad $1 - (1 - \frac{1}{N})^N$** de ser seleccionadas al menos una vez en N muestras. Si $N \rightarrow \infty$, converge a $(1 - \frac{1}{e}) = 0,632$.
- Es decir, con N muy grande, **cada conjunto C_j tiene una esperanza de tener un 63.2% de los datos de entrenamiento una sola vez**, y el resto duplicados. Esto indica que la evaluación en el límite, **permite evaluar gran parte de la base de datos**.

Bagging: La idea... [Breiman, 1996]

- Con cada uno de los conjuntos C_j , se entrena un modelo, obteniendo de esta manera m modelos ajustados con los conjuntos $C_j, j \in \{1, \dots, m\}$
- Los resultados considerados para Bagging se pueden promediar (para el caso de las regresiones) o se puede utilizar un Majority Voting (para el caso de clasificación).



Boosting

- Utiliza una técnica de resampling distinta a Bagging, la cual **mantiene una probabilidad constante de $1/N$ para la selección** de cada observación.
- Se va **actualizando la probabilidad** utilizada en resampling en el tiempo, **basado en el desempeño** del modelo.
- Basado en el concepto de “**clasificador débil**”, donde basta un modelo que entregue resultados un poco mejor que al azar. (mayor a una clasificación de un 50 %)
- Existe una gran variante de técnicas de boosting, uno de los ejemplos **más utilizados es el algoritmo Adaboost** [Schapire, 2001].



AdaBoost (Adaptive boosting): La idea...

Permite al experimentador ir **agregando nuevas componentes de clasificación** al modelo a medida que se va **logrando un error más pequeño**. Opera de la siguiente manera:

- 1 En la iteración n -ésima se provee al “clasificador débil” con una **distribución $D_n(i)$** , que representa la **probabilidad de seleccionar la observación i -ésima**.
- 2 Se **genera el modelo** en base a los datos de entrenamiento formada por la distribución $D_n(i)$, **midiendo la tasa de error** en base a un test de hipótesis H_n con respecto a $D_n(i)$
- 3 Se **genera $D_{n+1}(i)$** , **bajando la probabilidad** de seleccionar aquellas observaciones que fueron bien clasificadas en la iteración n , e **incrementando la probabilidad de aquellas mal clasificadas**.



Breiman, L. (1996).
Bagging predictors.
Machine Learning, 24(2):123–140.



Schapire, R. (2001).
The boosting approach to machine learning: An overview.