



## Aprendizaje de supervisado de funciones de distancia para el reconocimiento de patrones

**ENTIDADES EJECUTORAS PRINCIPALES:** Universidad Central Marta Abreu de Las Villas, Cuba; Universidad de Gantes, Bélgica

**Autores:** Prof. Dr. C. Carlos Morell Pérez<sup>1</sup>, MSc. Bac Nguyen Cong<sup>2</sup>, Prof. Dr. Bernard De Baets<sup>2</sup>

**Colaboradores:** Prof. Dr. C. Francesc J. Ferri<sup>3</sup>

**Filiación:** <sup>1</sup>Centro de Investigaciones de la Informática. Facultad de matemática, Física y Computación de la Universidad Central Marta Abreu de Las Villas, Cuba. <sup>2</sup>KERMIT, Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Belgium. <sup>3</sup>Universidad de Valencia, España

### Palabras clave

*aprendizaje automático; funciones de distancia; reconocimiento; patrones*

### RESUMEN

El presente trabajo tiene como objetivo diseñar nuevas estrategias para el aprendizaje de funciones de distancia de modo que el método k-NN pueda ser aplicado con eficacia y eficiencia en diferentes contextos del aprendizaje automático. Ante cada contexto particular se define un problema de optimización adecuado y se plantea un método de solución eficiente. Cada uno de los métodos ha sido implementado computacionalmente y posteriormente se ha evaluado de forma empírica utilizando conjuntos de datos tomados de repositorios internacionalmente reconocidos. Esta validación incluye la comparación con otros algoritmos actualizados de modo que pueden destacarse las fortalezas y debilidades de las propuestas propias. Cada una de las estrategias propuestas ha sido publicada en 6 artículos en revistas de la corriente principal y presentada en un evento internacional de prestigio.

Uno de los métodos más simples y eficaz para la predicción automática a partir de datos es el de los k vecinos más cercanos (del inglés *k-nearest neighbour* o simplemente k-NN). A pesar del tiempo transcurrido desde su creación, el método k-NN continúa siendo utilizado con frecuencia en diversas tareas de clasificación y regresión en gran número de aplicaciones prácticas.

En el caso de la clasificación, el método asignará a un nuevo ejemplo la clase mayoritaria entre los vecinos más cercanos que se encuentran en el conjunto de entrenamiento, utilizando para ello una función de distancia o similitud. Por la propia naturaleza de su regla de decisión, la calidad de la clasificación depende de la forma en que se calculan las distancias entre los ejemplos previos. Cuando no hay ningún cono-

cimiento disponible sobre el dominio de aplicación, o incluso en presencia de tal conocimiento, la mayoría de las implementaciones de k-NN utilizan la función de distancia euclidiana (suponiendo que los ejemplos se representan como vectores de número reales). Sin embargo, la utilización de una función de distancia adecuada es fundamental para un buen comportamiento de cualquiera de los algoritmos de clasificación basados en ejemplos, tales como k-Means, el prototipo más cercano, y otros. Las funciones de distancia como la euclidiana ignoran cualquier regularidad estadística que existe entre los atributos de los ejemplos del conjunto de entrenamiento.

Se puede aprender la función de distancia adecuada en diversos campos, tales como clasificación, visión por computadora, recuperación de información o bioinformática, según el problema que se quiere resolver. Es por ello que el presente trabajo propone diversas estrategias para aprender funciones de distancia en diferentes contextos en los que puede ser aplicable el método de aprendizaje supervisado k-NN, a saber, clasificación, clasificación ordinal, regresión y problemas de alta dimensionalidad. Ante cada situación específica, se ha modelado el problema de aprendizaje como un problema de optimización y se ha propuesto un método de solución específico que aproveche las especificidades del problema para encontrar una solución con eficiencia. A continuación se describen estas problemáticas, así como la estrategia seguida para darles una solución eficaz y eficiente.

## **Aprendizaje de funciones de distancia para problemas clasificación**

En el contexto de la clasificación supervisada el presente trabajo incluye tres propuestas. La primera de ellas consiste en utilizar una heurística, de modo que el problema de optimización derivado del aprendizaje de distancias se convierte en otro de más fácil solución. En esencia, la propuesta se basa en la observación de que los objetos más cercanos, de la misma clase, a un ejemplo de aprendizaje determinado se distribuyen de forma esencialmente distinta a los objetos con clase diferente. Desde esta perspectiva, se puede hacer una transformación lineal del espacio de entrada de modo que se maximice la distancia o separación entre ambas distribuciones. Por ello se propone un marco de optimización para el aprendizaje de funciones distancia a través de transformaciones lineales maximizando la divergencia de Jeffrey entre dos distribuciones gaussianas multivariadas obtenidas a partir de las restricciones locales apareadas.

El problema de optimización planteado tiene una solución cerrada, de modo que se evade la utilización de tediosos procedimientos de optimización. La solución es fácil de implementar y manejable para problemas a gran escala. Los resul-

tados experimentales obtenidos tanto para la versión lineal como para la versión "kernelizada" del método propuesto son buenos en tanto en varios casos de prueba se obtienen precisiones de clasificación superiores a los métodos de aprendizaje de funciones de distancia de vanguardia mientras que es competitivo en otros [4].

La segunda propuesta se basa en la observación de que los objetos de una misma clase no necesariamente se encuentran localizados en una misma región del espacio de características. En tales condiciones, la utilización de una única función de distancia global puede no resultar apropiada. Por ello se propone un enfoque basado en múltiples funciones de distancia que dependan de la localidad de los datos. Para lograr ese objetivo, se agrupan los ejemplos de aprendizaje, de forma previa, en clústeres disjuntos y posteriormente se plantea un problema de optimización global cuya solución determina una función de distancia para cada grupo. Adicionalmente se introduce en el problema de optimización un regularizador que permite conservar las escalas de las diferentes funciones aprendidas a la vez que garantiza aprender características comunes a los diferentes subgrupos. La solución al problema de optimización se obtiene mediante un proceso iterativo que converge rápidamente, de modo que esta propuesta garantiza una eficacia superior a otros métodos de la competencia que aprenden una única función de distancia mientras resulta más eficiente que otros métodos similares basados en múltiples funciones de distancia [1].

La tercera propuesta se basa en explotar interioridades del problema de optimización derivado del problema de aprendizaje de distancias clásico para proponer vías de solución más eficientes. En primer lugar, se propone un método de solución en dos etapas que primero produce una transformación ortogonal del espacio de entradas y luego una transformación en la dirección de los nuevos ejes de coordenadas. Este enfoque se puede resumir informalmente en la frase: "primero rotar y luego estirar". Al dividir el problema original en dos problemas más sencillos, se logra un nuevo método de aprendizaje de funciones de distancia que resulta altamente eficiente [7]. El método para determinar la transformación ortogonal óptima mediante una solución analítica es conocido de la literatura relacionada con el tema mientras que el método para transformar el espacio de característica en la dirección de los ejes de coordenada es una propuesta propia publicada con anterioridad por los autores de este trabajo.

## **Aprendizaje de funciones de distancia para problemas de regresión**

El aprendizaje de funciones de distancia en el contexto de la regresión está menos extendido que en el caso de la clasifi-

cación. Sin embargo, recientemente han aparecido enfoques que muestran la validez de esta aproximación a los problemas de regresión supervisados. El presente trabajo incluye una propuesta de solución al problema de optimización que se define al aprender una función de distancia en este contexto. Es decir, la modelación del problema resulta en un programa cuadrático (convexo) cuya solución iterativa es costosa en términos de complejidad computacional, aunque resulta poco práctica para problemas de grandes dimensiones.

La propuesta propia incluye un método dual de descenso por coordenada que permite obtener resultados competitivos en cuanto a eficacia cuando se compara con los métodos más actuales, a la vez que resulta mucho más rápido [5].

## **Aprendizaje de funciones de distancia para problemas de clasificación ordinal**

La clasificación ordinal es una problemática que incluye elementos de la clasificación y de la regresión donde la meta consiste en predecir la clase asociada a un ejemplo desconocido teniendo en cuenta que existe una relación de orden entre las distintas etiquetas. A pesar de este ordenamiento, la distancia entre estas etiquetas es usualmente desconocida, contrario a lo que sucede en la regresión. Desechar la información contenida en este ordenamiento y tratar los problemas de clasificación ordinal como problemas de clasificación estándar puede causar una pérdida importante en la eficacia de los algoritmos de predicción. Es por ello que en el presente trabajo se propone un nuevo método de aprendizaje de funciones de distancia para este contexto.

La propuesta incluye la formulación de un problema de optimización con restricciones, donde se incluye la información que provee la naturaleza ordinal de las clases. Al igual que en otras propuestas de este trabajo, también se propone un método de solución para el problema de optimización formulado y se hace una validación empírica comparando la eficacia de la propuesta con otras alternativas disponibles. Los resultados obtenidos muestran la competitividad de la propuesta realizada [2].

## **Aprendizaje de funciones de distancia en problemas de clasificación de alta dimensionalidad**

En varios dominios de aplicación se pueden encontrar ejemplos caracterizados por un gran número de atributos. Cuando la cantidad de atributos es muchos mayor que la

cantidad de ejemplos de aprendizaje, los métodos comunes del aprendizaje automático pueden fallar. Esto se conoce como “la maldición de la dimensionalidad”, y sucede en muchas áreas de aplicación como la genómica, la quimiointormática, el análisis de imágenes médicas y otras. En el presente trabajo se proponen dos formas de lidiar con esta problemática desde la perspectiva del aprendizaje de funciones de distancia.

La primera estrategia consiste en aprender una matriz diagonal en lugar de una matriz completa (matriz de Mahalanobis). Con esto se reduce la complejidad espacial (de  $n^2$  a  $n$ ) a la vez que se hace una selección de atributos implícita. La formulación de este conduce a un problema de optimización con restricciones y para ello se ha propuesto una vía de solución específica que hace uso de particularidades del problema para encontrar una solución eficiente en cuanto a complejidad computacional. El modelo resultante ha sido sometido a una sólida validación empírica que ha demostrado la validez de la propuesta [6].

La segunda estrategia para lidiar con la maldición de la dimensionalidad se basa en la observación de que en varias áreas del conocimiento es posible encontrar nuevos ejemplos del dominio de aplicación, pero que no pertenezcan a ninguna de las clases bajo análisis. Por ejemplo, si se quiere clasificar de forma automática dígitos manuscritos, es posible crear varios clasificadores que distingan correctamente entre dos dígitos.

En el caso hipotético del clasificador que debe distinguir entre los dígitos 5 y 8, se puede utilizar la información proveniente del resto de los dígitos en forma de ejemplos adicionales (conocido en la literatura como *universum*). Existe evidencia empírica de que un clasificador binario tendrá mayor poder de generalización en la medida en que produzca más contradicciones clasificando los ejemplos del universo.

El presente trabajo incluye una propuesta de extensión de uno de los mejores algoritmos más actuales (conocido como LMNN) para incorporar la información adicional proveída por los ejemplos del universo. La validación realizada muestra que el algoritmo resulta competitivo con propuestas similares actuales [3].

## **Referencias bibliográficas**

- [1] Nguyen, B.; Ferri, F., Morell, C.; De Baets, B. (2019). An efficient method for clustered multi-metric learning. *Information Sciences*. 471, 149-163.

- [2] Nguyen, B.; Morell, C.; De Baets, B. (2018). Distance metric learning for ordinal classification based on triplet constraints. *Knowledge-Based Systems*. 142, 17-28.
- [3] Nguyen, B.; Morell, C.; De Baets, B. (2017). Distance metric learning with the Universum. *Pattern Recognition Letters*, 100, 37-43.
- [4] Nguyen, B., Morell, C. y De Baets, B. Supervised distance metric learning through maximization of the Jeffrey divergence. *Pattern Recognition* 64 (2017), 215-225.
- [5] Nguyen, B., Morell, C. y De Baets, B. Large-scale distance metric learning for k-nearest neighbors regression. *Neurocomputing* 214 (2016), 805-814.
- [6] Nguyen, B., Morell, C. y De Baets, B. Un método eficiente para resolver un problema de programación cuadrática derivado del aprendizaje de funciones de distancia. *Investigación Operacional* 37.2 (2016), 124-137.
- [7] Nguyen Cong, B.; Morell, C.; De Baets, B. Distance metric learning: A two-phase approach. In ESANN 2017: 25-th European Symposium on Artificial Neural Networks. *Computational Intelligence and Machine Learning*, 123-128.

#### AUTOR PARA LA CORRESPONDENCIA

**Carlos Morell Pérez.** Centro de Investigaciones Informáticas. Facultad de Matemática, Física y Computación Universidad Central Marta Abreu de Las Villas. Carretera a Camajuaní. km 5½. Santa Clara, Villa Clara, Cuba. CP 54830. Correo electrónico: [cmorellp@uclv.edu.cu](mailto:cmorellp@uclv.edu.cu)