

Optimización de métodos de selección de características en precisión reducida mixta para dispositivos con recursos limitados*

Simón Noya Domínguez, Samuel Suárez-Marcote, Verónica Bolón-Canedo y Laura Morán-Fernández

CITIC, Universidade da Coruña, A Coruña, España
simon.noyad@udc.es

Resumen

El crecimiento de los dispositivos de Internet de las Cosas (IoT) genera grandes volúmenes de datos que deben procesarse en dispositivos con recursos limitados. Este artículo analiza el uso de la precisión reducida mixta en métodos de selección de características basados en Información Mutua, en comparación con la precisión doble. Se evaluaron 12 conjuntos de datos tanto reales como sintéticos, con diferente número de características, recurriendo a la validación cruzada y pruebas estadísticas. Los resultados muestran que, en baja dimensionalidad, algunas configuraciones mixtas de punto fijo mantienen un rendimiento similar al de la precisión doble. En alta dimensionalidad, se observa mayor variabilidad, aunque ciertas configuraciones siguen siendo competitivas. En conclusión, la propuesta permite reducir el uso de recursos, lo que resulta atractivo para dispositivos con recursos limitados.

1. Introducción

Con el crecimiento de los dispositivos de IoT en áreas tan diversas como la salud, la investigación o el entretenimiento, en los últimos años la cantidad de datos disponibles ha aumentado de forma considerable. Sin embargo, estos equipos cuentan con limitaciones tanto por su consumo energético, como por su capacidad de almacenamiento y cómputo. Dichas limitaciones obligan a que el cálculo de operaciones de alto coste deba ser realizado en la nube. Esto lleva a la aparición del paradigma de *Edge Computing* [1], que busca realizar las operaciones antes mencionadas lo más cerca posible de la fuente de los datos, es decir, en el propio dispositivo.

Para poder trabajar con las cantidades de datos disponibles, es necesario un paso previo de selección de características para así poder disminuir la dimensión de los datos y favorecer la búsqueda de patrones en ellos [2].

2. Metodología

El objetivo principal de este trabajo es la evaluación del impacto del uso de la precisión reducida mixta en ciertos algoritmos de selección de características de tipo filtro basados en Información Mutua [3]. Los algoritmos reimplementados fueron *Mutual Information Maximization (MIM)*, *Joint Mutual Information Maximization (JMI)* y, por último, *minimum Redundancy Maximum Relevance (mRMR)*.

*Investigación financiada por el Ministerio de Transformación Digital y Función Pública y Next Generation (TSI-100925-2023-1).

Los métodos de selección de características en doble precisión fueron modificados para poder soportar precisión simple mixta, adaptando cada operación de los algoritmos para que puedan operar de la misma manera que lo hacían en doble precisión. La implementación de precisión mixta se aplicó en tres fases dentro de los métodos estudiados. En estas fases encontramos primero el cálculo de las probabilidades marginales, después tenemos las operaciones intermedias y por último la resolución de las operaciones aritméticas. En las probabilidades marginales se probó con 4 y 8 bits; en las operaciones intermedias usamos 8 y 16; y en las operaciones aritméticas se utilizaron 8, 16 y 32 bits.

La experimentación consistió en el uso de diferentes configuraciones de precisión reducida mixta en los métodos de selección de características, y luego comparar los resultados con las versiones en doble precisión. Para poder evaluar la similitud de los rankings obtenidos mediante el uso de la precisión reducida y los de doble precisión, utilizando la tasa de verdaderos positivos (TVR). Además, se analizó el impacto que las diferentes configuraciones en precisión reducida mixta tienen sobre la información contenida en los datos, empleando un paso posterior de clasificación, utilizando validación cruzada, y se aplicaron test estadísticos para determinar la existencia de diferencias significativas entre las configuraciones evaluadas.

3. Resultados y conclusión

Para realizar el estudio, se usaron 12 conjuntos de datos, tanto reales como sintéticos. Algunos de estos conjuntos son de tipo microarray, que se caracterizan por tener un gran número de características, pero un número reducido de muestras [4]. Los resultados obtenidos muestran que, para todos los métodos y conjuntos de datos, una configuración con 8 bits para las probabilidades marginales, 16 para las operaciones intermedias y 16 para las operaciones aritméticas mantiene un rendimiento muy similar y, en muchos casos, prácticamente idéntico al de doble precisión. También se destaca que configuraciones más reducidas como 4 bits para las probabilidades marginales y 8 bits para las operaciones intermedias y aritméticas pueden alcanzar resultados similares a la doble precisión con ciertos métodos y conjuntos de datos de baja dimensionalidad.

En conclusión, los resultados demuestran que el uso de la precisión reducida mixta, no implica una pérdida de la calidad en los resultados de los algoritmos de selección de características basados en información mutua, siendo así una estrategia viable su uso con la configuración adecuada para garantizar su efectividad.

Referencias

- [1] Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li, and Lanyu Xu. Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5):637–646, 2016.
- [2] A. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1–2):245–271, 1997.
- [3] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.
- [4] Verónica Bolón-Canedo, Noelia Sánchez-Marono, Amparo Alonso-Betanzos, José Manuel Benítez, and Francisco Herrera. A review of microarray datasets and applied feature selection methods. *Information Sciences*, 282:111–135, 2014.