Sistema federado de regresión basado en reglas difusas

Markel Crespo-Lamelas[1], Humberto Bustince[1,2][0000−0002−1279−6195], and Asier Urio-Larrea[1,2][0000−0001−7667−5766]

[1] Universidad Pública de Navarra, Pamplona 31006, Navarra
crespo.153999@e.unavarra.es
[2] Institute for Smart Cities (ISC), Universidad Pública de Navarra, Pamplona 31006, Navarra
{bustince,asier.urio}@unavarra.es

Abstract. El aumento de la disponibilidad digital de datos ha provocado una mayor y más efectiva utilización de sistemas de aprendizaje automático en multitud de sectores. Este uso conlleva varios problemas, entre ellos la privacidad de los datos que se utilizan y la capacidad de comprender el funcionamiento interno del sistema. El aprendizaje federado trata el problema de la privacidad mediante la capacidad de combinar modelos aprendidos en distintos nodos sin llegar a compartir los datos, mientras que sistemas como los basados en lógica difusa producen modelos fácilmente comprensibles para el ser humano. Este trabajo propone un método para combinar las bases de reglas de los distintos clientes de un sistema federado basado en reglas difusas para tratar problemas de regresión. Este sistema evita compartir los datos y produce un sistema explicable basado en reglas difusas.

Keywords: Inteligencia Artificial Explicable · Aprendizaje Federado · Lógica Difusa.

1 Introducción

El aumento de la disponibilidad digital de datos ha provocado una mayor y más efectiva utilización de sistemas de aprendizaje automático en multitud de sectores. Este uso conlleva varios problemas, entre ellos la privacidad de los datos que se utilizan[6] y la capacidad de comprender el funcionamiento interno del sistema[3].

El aprendizaje federado propuesto por McMahan [5] busca abordar diferentes problemas, como son la transmisión de grandes volúmenes de datos, la necesidad de requerir gran capacidad de cómputo o la privacidad de los datos. Mediante el aprendizaje de modelos parciales en cada uno de los clientes que participan en la federación, en lugar de enviar todos los datos al punto central solamente se envían los parámetros que aprende el modelo, de esta manera se reduce el volumen de datos a transmitir y se aumenta la privacidad del sistema. Así mismo, cada cliente necesita de una menor capacidad de cómputo al tratar solamente sus datos.

Gran parte del crecimiento de los sistemas de aprendizaje automático se ha dado gracias a las redes neuronales profundas. Estas redes, a pesar de su capacidad de producir buenos resultados, no permiten conocer con facilidad el proceso por el que producen un resultado, por esto se llaman sistemas de caja negra[1]. Estos sistemas no son

adecuados en muchos sectores, como el sanitario o el financiero, sectores en los que se requiere conocer las razones por las que se toma una decisión o se llega a un resultado. Entre los sistemas explicables se encuentran los basados en lógica difusa los cuales producen modelos basados en reglas fácilmente comprensibles para el ser humano[2].

Este trabajo combina los dos conceptos anteriores con el objetivo de producir un sistema federado explicable basado en reglas difusas para tratar problemas de regresión.

2 Sistemas federados difusos

En este trabajo utilizamos el modelo propuesto por Mamdani y Assilian [4]. En estos sistemas, las reglas presentan los antecedentes y consecuentes como conjuntos difusos, de tal forma que siguen la siguiente estructura:

$$R_j: \text{ Si } x_1 \text{ es } A_{j1} \text{ y } \ldots \text{ y } x_n \text{ es } A_{jn} \text{ entonces } y_p \text{ es } B_j$$

Donde R_j es la etiqueta de la regla j-ésima, $A_{j1},\ldots,A_{jn},B_j$ son los conjuntos difusos y $X=(x_1,\ldots,x_n)$ son las variables de entrada.

El aprendizaje federado aplicado a sistemas de lógica difusa presenta una gran diferencia respecto a los enfoques tradicionales basados en redes neuronales. Mientras que en las redes neuronales la agregación consiste en la combinación numérica de pesos (parámetros), en los sistemas basados en reglas difusas el desafío radica en la fusión semántica de bases de conocimientos heterogéneas.

Dado que cada cliente desarrolla sus reglas de forma independiente, el modelo global debe integrar estos conocimientos preservando tanto la coherencia lógica como la interpretabilidad del sistema. En este contexto, la agregación no constituye una simple operación aritmética, sino una fase crítica donde deben resolverse redundancias y contradicciones latentes que, de no ser gestionadas, comprometerían la capacidad explicativa y la fiabilidad del modelo.

3 Metodología

Técnicas de agregación de reglas Para poder tener criterio en las estrategias de agregación, hemos asignado a cada regla un grado de certeza (CF_j) que cuantifica su fiabilidad basada en la evidencia de los datos. Este valor se calcula determinando el grado de activación mínimo de los antecedentes de la regla para cada ejemplo de entrenamiento y seleccionando, posteriormente, el máximo de dichos valores: $CF_j = \max_k(\min_i \mu_{ji}(x_{ik}))$. Este parámetro puede utilizarse en las estrategias de filtrado, ya que permite al servidor tomar decisiones priorizando la regla que cuenta con un mayor respaldo en los datos locales de los clientes.

Para tratar la combinación de bases de reglas, se han implementado y comparado tres estrategias de agregación, diseñadas para depurar y optimizar la base de reglas global:

- **Sin filtrado:** Se mantiene la totalidad de las reglas generadas, sirviendo como escenario de referencia.

– **Eliminación de duplicados y contradicciones:** Se eliminan las reglas duplicadas y las que son contradictorias.
– **Eliminación de duplicados y contradicciones con grado de certeza:** Se eliminan las reglas duplicadas y las que son contradictorias que tienen un menor grado de certeza, conservando las de mayor valor.

Estrategia de evaluación del modelo Evaluaremos el modelo bajo dos perspectivas:

– **Personalizada (FP):** Evalúa el modelo en cada cliente utilizando los datos locales de cada cliente.
– **Global (FG):** Evalúa el modelo frente al conjunto de test global, formado por los conjuntos de test de todos los clientes, midiendo la capacidad de generalización final de todo el sistema federado.

Dataset Para la validación del sistema, se ha utilizado el dataset *Combined Cycle Power Plant*[8,7], orientado a la predicción de la producción neta de energía eléctrica de una central térmica en función de variables ambientales. El conjunto de datos consta de 9.568 mediciones y cuatro variables de entrada.

Métricas de evaluación Para evaluar la calidad predictiva se emplean el coeficiente de determinación (R^2), que mide la bondad de ajuste del modelo, y el error absoluto medio (MAE), que cuantifica la magnitud del error en las unidades de la variable objetivo.

4 Resultados y Conclusiones

Table 1. Resultados por Método de Filtrado: R^2-SCORE y MAE por Modelo

	R^2-SCORE		MAE	
Filtrado	**FP**	**FG**	**FP**	**FG**
sin filtrado	0.3679	0.3707	10.3058	10.3058
Eliminación contradicciones	0.3675	0.3709	11.7311	11.7311
Agregación grado certeza	0.7955	0.7966	7.1201	7.1141

Tal y como se observa en la Tabla 1, la estrategia de agregación es determinante para el rendimiento del sistema. Podemos concluir:

Importancia del grado de certeza: Los resultados evidencian que el rendimiento es peor tanto en el escenario *sin filtrado* como en la *eliminación de duplicidades y contradicciones* simple, lo que sugiere que la depuración indiscriminada de reglas degrada la capacidad predictiva al desechar información valiosa. Frente a esto, la incorporación del grado de certeza permite una mejora en el rendimiento al cuantificar la fiabilidad de cada regla mediante la evidencia en los datos.

Evaluación personalizada y global: La ausencia de diferencias significativas entre los modelos evaluados de forma personalizada (FP) y los evaluados globalmente (FG) puede deberse a que la distribución de los datos de este dataset específico entre los clientes sea homogénea. En este caso, también puede observarse que el sistema no pierde capacidad al generalizar.

Líneas futuras: El trabajo tiene que continuar evaluando más alternativas de agregación y comparando con otros modelos equivalentes del estado del arte. También es necesario el análisis de más datasets con diferentes características de distribución de datos.

Acknowledgments. Este trabajo ha sido posible gracias al proyecto PID2022-136627NB-I00 financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE.

References

1. Adadi, A., Berrada, M.: Peeking inside the black-box: A survey on explainable artificial intelligence (xai). IEEE Access **6**, 52138–52160 (2018). https://doi.org/10.1109/ACCESS.2018.2870052
2. Alonso, J.M., Castiello, C., Mencar, C.: Interpretability of fuzzy systems: Current research trends and prospects. Springer handbook of computational intelligence pp. 219–237 (2015)
3. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D.: A survey of methods for explaining black box models. ACM computing surveys (CSUR) **51**(5), 1–42 (2018)
4. Mamdani, E.: Application of fuzzy algorithms for control of simple dynamic plant. Proceedings of the Institution of Electrical Engineers **121**(12), 1585 (1974). https://doi.org/10.1049/piee.1974.0328
5. McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.y.: Communication-Efficient Learning of Deep Networks from Decentralized Data. In: Singh, A., Zhu, J. (eds.) Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 54, pp. 1273–1282. PMLR (20–22 Apr 2017), https://proceedings.mlr.press/v54/mcmahan17a.html
6. Papernot, N., McDaniel, P., Sinha, A., Wellman, M.P.: Sok: Security and privacy in machine learning. In: 2018 IEEE European symposium on security and privacy (EuroS&P). pp. 399–414. IEEE (2018)
7. Tfekci, P., Kaya, H.: Combined Cycle Power Plant. UCI Machine Learning Repository (2014), DOI: https://doi.org/10.24432/C5002N
8. Tüfekci, P.: Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. International Journal of Electrical Power & Energy Systems **60**, 126–140 (2014), https://api.semanticscholar.org/CorpusID:111365542