

Predicción de heladas usando aprendizaje automático e internet de las cosas

Keywords: frost prediction, frost events, viticulture, agriculture, machine learning, SMOTE

1 Introducción

En este trabajo se comparten las experiencias de la construcción de un sistema predictivo para heladas agronómicas [2,9] usando R para la experimentación y análisis de resultados. Se utilizaron los enfoques actuales de experimentación para la predicción de temperaturas mínimas diarias que utilizan algoritmos de aprendizaje automático entrenados por lecturas pasadas de los sensores de temperatura y humedad para predecir temperaturas futuras. Sin embargo, al contrario de los enfoques actuales, asumimos que las condiciones termodinámicas circundantes son informativas para la predicción. Por esto, se desarrolló un modelo por cada ubicación, que incluye en su información de entrenamiento las lecturas de sensores de todas las demás ubicaciones que son más relevantes.

Evaluamos nuestro enfoque mediante el entrenamiento de modelos de regresión y clasificación, muchos ya propuestos en la literatura para el problema de predicción de heladas, sobre datos de cinco estaciones meteorológicas distribuidas a lo largo de la Provincia de Mendoza en Argentina. Usamos los paquetes de R *caret* [4] y *doParallel* [8] para construir los experimentos. Los algoritmos de aprendizaje automático utilizados fueron: regresión logística, árboles de decisión (C5.0 [3], particionamiento recursivo [7]), *random forest* [5] y redes bayesianas [6]. Dada la escasez de eventos de heladas, se procedió a balancear el dataset utilizando la técnica de sobremuestreo de minorías sintéticas (SMOTE) [1].

Las métricas utilizadas para comparar los modelos de clasificación fueron: *sensitivity* o *recall* (también denominado exhaustividad), precisión, *F-measure* y para los modelos de regresión *RMSE* y *MAE*. Los resultados experimentales muestran que seleccionar a los vecinos más relevantes y entrenar los modelos con SMOTE reduce los errores de predicción de ambos predictores de regresión (*random forest* y redes bayesianas) para las cinco ubicaciones, en algunos casos hasta un 10%. Además aumenta el rendimiento de los predictores de clasificación, en términos de *F-measure*, de *random forest* para cuatro ubicaciones mientras se mantiene sin cambios para la restante, y produce resultados no concluyentes para el predictor de regresión logística. Los valores de *F-measure* de los mejores modelos resultantes variaron entre 0.7 y 0.9 y los de *recall* entre 0.7 y 0.8. Otra observación esperable es en el aumento del *recall* en detrimento de la

precisión en comparar los experimentos con y sin aplicación de SMOTE respectivamente.

Estos resultados comprobaron la afirmación principal: que la información termodinámica de las ubicaciones vecinas puede ser informativa para mejorar las predicciones de regresión y clasificación, pero también es lo suficientemente buena como para sugerir que el enfoque actual es un recurso válido y útil para los productores agropecuarios o tomadores de decisiones, ya que disminuye los errores en la predicción. En un esquema de redes inalámbricas de sensores (IoT) en el campo no siempre los nodos tienen acceso a internet a un servidor central para recibir la predicción u otra información. Conocer los vecinos más relevantes ayudaría a tomar mejores decisiones de ruteo de la información, que permita alimentar a sus predictores de heladas locales.

2 Referencias

- [1] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- [2] Diedrichs, A. L., Bromberg, F., Dujovne, D., Brun-Laguna, K., & Watteyne, T. (2018). Prediction of frost events using machine learning and IoT sensing devices. *IEEE Internet of Things Journal*, 5(6), 4589-4597.
- [3] Kuhn, M., Weston, S., Coulter, N., & Quinlan, R. (2014). C50: C5. 0 decision trees and rule-based models. *R package version 0.1. 0-21*, URL <http://CRAN.R-project.org/package=C50>.
- [4] Kuhn, M. (2012). The caret package. *R Foundation for Statistical Computing, Vienna, Austria*. URL <https://cran.r-project.org/package=caret>.
- [5] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- [6] Scutari, M., & Ness, R. (2012). bnlearn: Bayesian network structure learning, parameter learning and inference. *R package version*, 3.
- [7] Therneau, T. M., Atkinson, B., & Ripley, B. (2010). rpart: Recursive partitioning. *R package version*, 3(3.8).
- [8] Weston, Steve, and Rich Calaway. "Getting Started with doParallel and foreach." (2018). URL <https://cran.r-project.org/web/packages/doParallel/vignettes/gettingstartedParallel.pdf>
- [9] Ana Laura Diedrichs. (2019, July 29). anadiedrichs/diedrichs2017prediction-frost-experiments: first release (Version v0.1). Zenodo. <http://doi.org/10.5281/zenodo.3354574>