

Noticias de Investigación y Desarrollo Tecnológico

Mayo 2020

Año 19 N° 34

ISSN 2422-7447

Equipo Editorial

Editores

Dr. Juan Galantini

Dr. Roberto Distel

Dra. Cecilia Popovich

Dr. Diego Zappacosta

Dr. Alejandro Presotto

Secretario Editorial

Abg. Mariano Anderete Schwal

Contribuciones:

El boletín electrónico del CERZOS - CCT CONICET Bahía Blanca da la bienvenida a contribuciones de su personal para sus diferentes secciones.

Los artículos y notas que aparecen en el boletín representan la opinión de los autores y no necesariamente la política del CERZOS (CONICET-UNS). En cuanto al derecho de autor, los artículos en su totalidad o parcialmente no podrán ser reproducidos por terceros sin previa autorización del autor/ autores.

Versión on line

Alejandra Olazabal

Adrián Zunini

Área Cómputos y Comunicaciones de la UAT,
CONICET-BB

Su opinión, sugerencias o colaboraciones serán bien recibidas.

Para suscribirse, enviar información o comunicarse con la redacción, envíe su correo a:

boletín_cerzos@criba.edu.ar

Centro de Recursos Naturales Renovables de la Zona Semiárida (CERZOS)

Edificio E-1, Centro Científico Tecnológico (CCT CONICET Bahía Blanca)

Camino La Carrindanga, Km 7.

B8000FWB - Bahía Blanca. ARGENTINA

Tel: 54 (0291) 4861124, Fax: 54 (0291) 4862882

Contenido

EDITORIAL 3

Viviana Echenique*

Rizosfera: la dimensión subterránea de la biósfera 5

Marianela Morales*

¿Engorde a campo o a corral? Esa es la cuestión..... 7

Maria Sol Villaverde*

No somos lo que fuimos 10

Juan Manuel Rodrigo*

Algoritmos de aprendizaje automático para la predicción del modo reproductivo de
Eragrostis curvula en base a marcadores moleculares 12

Jimena Gallardo, Cristian Andrés Gallo, Diego Zappacosta y Viviana Echenique.....

Algoritmos de aprendizaje automático para la predicción del modo reproductivo de *Eragrostis curvula* en base a marcadores moleculares

Jimena Gallardo, Cristian Andrés Gallo, Diego Zappacosta y Viviana Echenique

A la hora de realizar una investigación, es fundamental contar con métodos simples, cuali o cuantitativos, que faciliten las tareas y permitan ahorrar tiempo, esfuerzos e insumos. Para la determinación precisa del modo reproductivo de *Eragrostis curvula* es menester utilizar técnicas citoembriológicas, que son laboriosas, requieren de la colección de sacos embrionarios en los estadíos adecuados y, por lo tanto, consumen tiempo. Es por ello que en nuestro estudio, se vislumbró la posibilidad de aplicar técnicas de minería de datos (*datamining*), utilizando algoritmos de aprendizaje automático, a fin de determinar la posibilidad de predecir el modo reproductivo de *E. curvula* en base a la presencia o ausencia de un marcador molecular.

Eragrostis curvula, vulgarmente llamada pasto llorón, es una planta originaria de Sudáfrica, naturalizada en zonas semiáridas y de suelos arenosos de Argentina, donde se la ha utilizado como forrajera y fijadora de médanos. Nuestro grupo de trabajo la utiliza como especie modelo para el estudio de su modo reproductivo, que puede ser sexual o por apomixis. En la reproducción sexual la formación de semillas se logra a partir de la unión (fecundación) de los gametos femeninos y masculinos, generando progenies de plantas genéticamente diversas. En la apomixis, en cambio, no hay intercambio de gametos, ya que los embriones de las semillas se forman a partir de la ovocélula misma, que no experimenta reducción meiótica, por lo tanto contienen embriones genéticamente idénticos a la planta madre. Esta forma de reproducción clonal por semillas puede aportar múltiples beneficios para el mejoramiento genético vegetal, a través de la obtención de híbridos permanentes, preservando caracteres multigénicos complejos y fijando características deseables a través de la semilla. Su introducción en maíz, trigo, arroz, sorgo y otros cultivos permitiría propagar híbridos en forma clonal, perpetuando el vigor en forma indefinida.

Para poder caracterizar a las plantas de *E. curvula* en base a su modo reproductivo se realiza un análisis citoembriológico. Para ello, durante el periodo de floración, las espiguillas (órganos donde se forman las semillas), se fijan en FAA (10 % formol, 5 % ácido acético, 50 % alcohol y 35 % agua destilada) y luego son incluidas, por separado, en parafina (Paraplast) y cortadas con un micrótomo. Se realizan cortes seriados de 10 μm de espesor que se tiñen con safranina-fast green y luego se observan en microscopio óptico, donde se estudian las estructuras típicas de procesos apomícticos y sexuales (presencia/ausencia de meiosis, y número y posición de los núcleos en el saco embrionario).

Otra forma de determinar el modo reproductivo de las plantas, es utilizando marcadores moleculares, que identifican las variaciones (polimorfismos) en la secuencia del ADN entre individuos. En este caso, ciertos marcadores se asocian a plantas apomícticas, permitiendo diferenciar a las plantas de *E. curvula* de manera

rápida y eficiente, evitando la complejidad y la demanda de tiempo impuesta por las técnicas citoembriológicas.

Como parte de nuestro estudio, muestras de ADN de 74 individuos de una población de de pasto llorón fueron enviadas al CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo, México), donde se analizaron con miles de marcadores moleculares, a fin de determinar que marcadores estaban presentes en plantas previamente clasificadas por métodos citoembriológicos en apomícticas o sexuales. Se generó una enorme base de datos que hizo indispensable el uso de herramientas informáticas que permitan la organización y análisis de los datos, la inferencia de relaciones entre éstos y el reconocimiento de patrones, entre otras cosas. Se utilizó el programa WEKA 3.8 (*Waikato Environment for Knowledge Analysis*), que se halla disponible de manera gratuita en el sitio oficial de la Universidad de Waikato (Nueva Zelanda) y contiene diversos algoritmos de aprendizaje automatizado.

Con *datamining* puede predecirse el modo reproductivo de las plantas dado que la computadora “aprende” reglas o generaliza comportamientos, usando como ejemplo datos proporcionados por el usuario. Al comenzar el análisis, WEKA ofrece distintas posibilidades con respecto al uso de los datos para el entrenamiento de los clasificadores (algoritmos). En este trabajo, se aplicó la opción de división del porcentaje, utilizando el 66% de los datos para entrenamiento y el 34% de los datos como conjunto de validación, es decir, 49 individuos para entrenamiento y 25 para validación. En la Tabla 1 se presentan los resultados obtenidos en la prueba de diferentes algoritmos, buscando aquel que pueda clasificar correctamente el mayor número de individuos utilizando los parámetros definidos por defecto y así hacer uso de esta herramienta para la clasificación futura de nuevos individuos de pasto llorón.

Tabla 1. Algoritmos utilizados para la determinación del modo reproductivo del pasto llorón y el porcentaje de individuos clasificados correctamente.

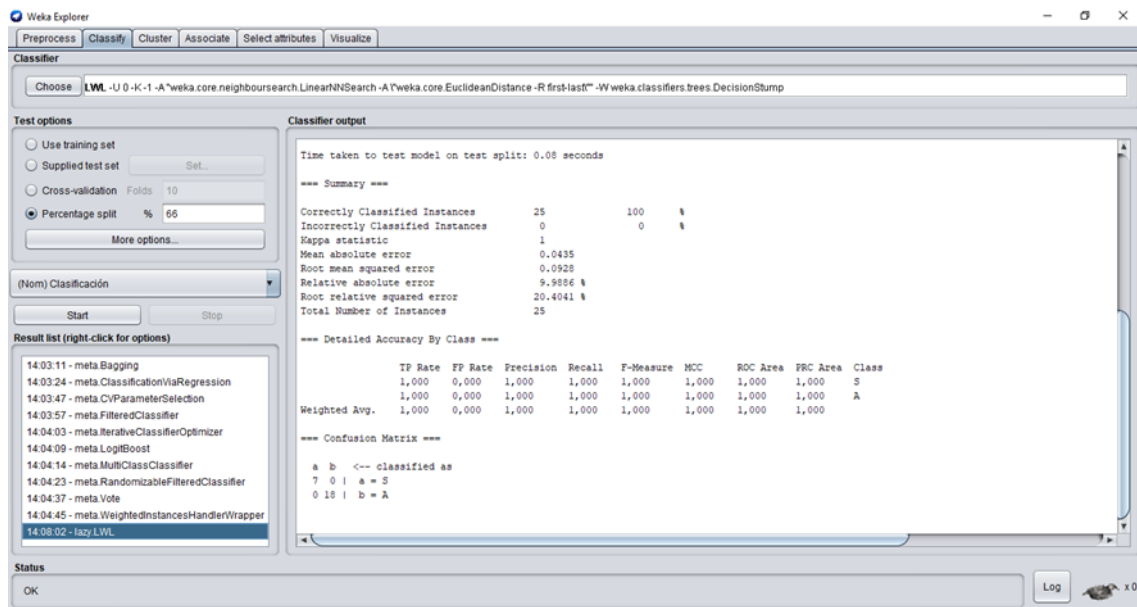
Algoritmo	%Clasificación correcta
BayesNet	68
NaiveBayes	64
Logistic	84
SimpleLogistic	100
IBK	72
LWL	100
AdaBoostM1	92
ClassificationViaRegression	96
InputmappedClassifier	72
DesicionTable	92
OneR	100
DisicionStump	100
HoeffdingTree	88

Utilizando los algoritmos LWL (Figura 1), OneR y DecisionStump se logró clasificar correctamente al 100% de los individuos, con lo cual se podría afirmar que estos son los que mejor predicen la forma reproductiva de nuestra población de *E. curvula*. Contrariamente, los que resultaron en una menor precisión en la predicción fueron los algoritmos NaiveBayes, BayesNet, IBK y InputmappedClassifier.

Este trabajo permitió determinar que es posible predecir el modo reproductivo de *Eragrostis curvula* utilizando algoritmos de aprendizaje automático, con datos de presencia/ausencia de marcadores moleculares, ya que se observa un alto grado de precisión en la validación de los métodos.

La predicción del modo reproductivo de los diferentes individuos con ayuda de técnicas de minería de datos brinda la posibilidad de reducir los tiempos, insumos y esfuerzos necesarios en la determinación citoembriológica, permitiendo además, simplificar el procesamiento de los datos, facilitando el análisis, comprensión e interpretación de los resultados.

Figura 1: Visualización en WEKA de la pestaña Classify, utilizando el algoritmo LWL.



*Ing. Agr. Jimena Gallardo, becaria doctoral CERZOS.