DETECCIÓN DE ROYA EN EL CAFÉ A TRAVÉS DE ARBOLES DE DECISIÓN

Felipe Álvarez Benítez
EAFIT
Medellín, Colombia
falvarezb@eafit.edu.co

Andrés Ospina Patiño
EAFIT
Medellín, Colombia
aospinap1@eafit.edu.co

Mauricio Toro
Universidad Eafit
Colombia
mtorobe@eafit.edu.co

RESUMEN

Palabras clave elegidas por el autor:
Café – roya – arboles de decisión – algoritmos de distribución – estructuras de datos
Palabras clave de la clasificación de la ACM:
Informática aplicada → Agricultura Computación aplicada → Biología computacional Sistemas de información → Estructuras de datos

1. INTRODUCCIÓN

Los tres principales productores de café en el mundo son Brasil (con 43.2 M de sacos al año), Vietnam (con 27.5M de sacos al año) y Colombia (con 13.5M de sacos al año). En Colombia, el café es nuestra principal exportación agrícola y, aproximadamente, 563.000 familias dependen de él. La plaga de la roya es el principal problema fitosanitario que afecta al café. El problema es agravado porque se hace un diagnóstico muy tarde. Esto hace que su control sea difícil y hay, inevitablemente, altas pérdidas. Aunque existen diversas variedades de café que son más resistentes a la roya, la variedad caturra (*coffea arabica*), que es la de exportación, es de las más susceptibles a esta plaga.

2. PROBLEMA

La caficultura es un motor para el desarrollo en la economía del país, en especial en el campo, donde en el año 2018 se contaba con cerca de 900 000 hectáreas, sin embargo según canicafé tres cuartas partes del área sembrada en café tienen variedades susceptibles, que están expuestas a ataques de roya y que con el presente cambio climático y el cambio drástico de las condiciones climáticas (precipitación, temperatura y humedad relativa, entre otros factores) han generado estrés en las plantaciones de café y han favorecido circunstancias propicias para esta plaga.

El problema de la roya radica principalmente en su detección tardía, imposibilitando el tratamiento oportuno y la erradicación de la plaga, es por esto por lo que este proyecto tiene como fin alertar oportunamente de la existencia de roya en los cafetales por medio de redes de sensores inalámbricos y arboles de decisión que digan si el cultivo esta afectado o no por el hongo.

3. TRABAJOS RELACIONADOS

3.1 Métodos heurísticos en problemas geométricos visibilidad, iluminación y vigilancia:

En este trabajo relacionado se habla afondo acerca de un problema que ha estado presente en varias ocasiones y que nadie podría haber pensado que podía ser solucionado mediante a algoritmos. El problema radica en la visibilidad de los espacios y cual es la manera mas optima de vigilar e iluminar un espacio que esta limitado por una figura geométrica.

3.2 Algoritmos de estimación de distribuciones en problemas de optimización combinatoria:

En este trabajo relacionado, se fundamenta principalmente en el ámbito de la ciencia donde los métodos algorítmicos radican principalmente en la sustitución de el cruce y la mutación por estimación y muestreo de una distribución de probabilidad. Es decir, lo que se busca con los algoritmos computacionales es optimizar la combinatoria para proceso de probabilidad u otros.

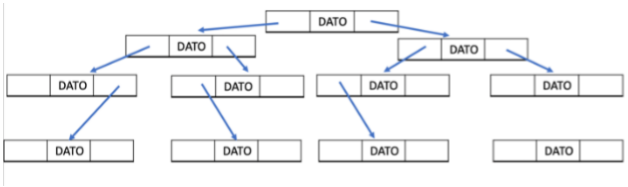
3.3 Los buses de campo aplicados al control de procesos industriales

En este articulo, se resalta la importancia y objetivo de los buses de campo (sistema de transmisión de información (datos) que simplifica enormemente la instalación y operación de máquinas y equipamientos industriales utilizados en procesos de producción) para mejorar la calidad del producto, reducir los costos y mejorar la eficiencia.

3.4 Influencia de algunas variables instruccionales

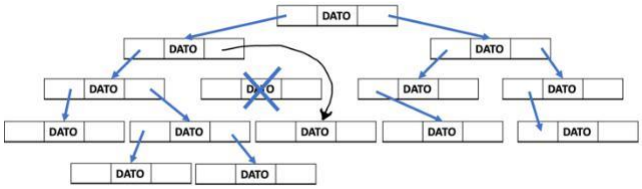
En este articulo se basa principalmente, en el uso de los algoritmos y como influyen estos en las aulas de algunos colegios en los grados de primaria y secundaria. El autor afirma que por medio de los algoritmos los estudiantes podrían dejar de memorizar algunas cosas y comenzarlas a implementar por medio de algoritmos. “El objetivo principal del presente estudio es analizar la influencia que tienen distintas variables instruccionales sobre el éxito en la resolución de problemas *algorítmicos* y *conceptuales* en educación secundaria, aunque también compararemos las capacidades de resolución de los sujetos en ambos tipos de problemas”.

4. Arboles Binarios

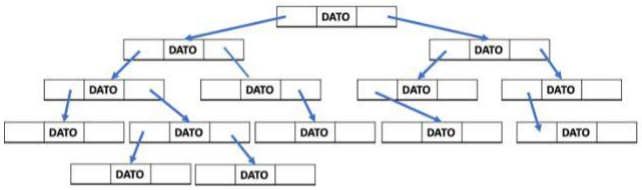


4.1 Operaciones de la estructura de datos

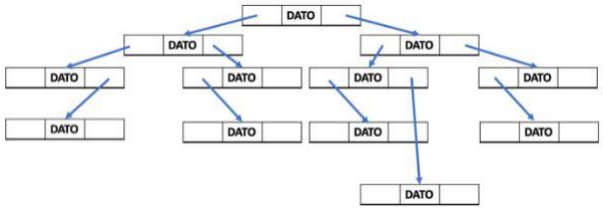
Eliminar



Mostrar el árbol



Insertar Nodo



4.2 Criterios de diseño de la estructura de datos

Debido a que el problema de la roya a veces tiene diversos cambios debido a diferentes factores, vimos necesario implementar un árbol binario en donde se puedan tomar decisiones mediante a los cambios que se ven y que el problema se pueda ir desarrollando de la mejor manera tomando en cuenta todos los posibles cambios ya sea por clima, temperatura, humedad entre otros…

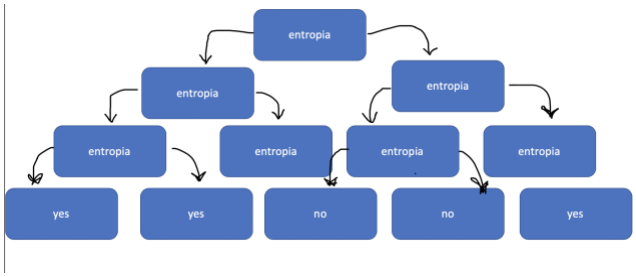
Por otra parte, este es una estructura de datos en donde los datos están muy organizados y es fácil de recorrer el árbol y ver sus componentes, hijos y hojas del árbol. En otras palabras, esta estructura de datos permite organizar toda la información de una manera mas jerarquizada que las listas donde se van a poder encontrar los datos de una mejor manera.

4.3 Análisis de Complejidad

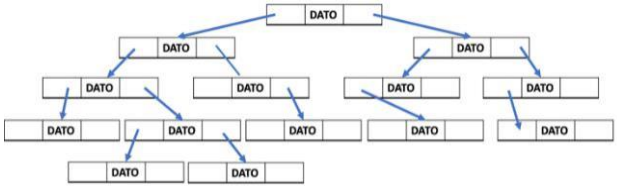
Metodos	Complejidad
Crear Nodo	O (1).
Insertar Nodo	O (n)

Estos son los métodos que de momento hemos realizado y que ya los tenemos implementados en una clase c++, a lo largo de las clases vamos a ir añadiendo métodos y sus respectivas complejidades.

5. Árbol de decisión con C4.5



5.1 Partición de los datos



Nuestro árbol a partir de la entropía y de su raíz, toma la decisión de partir el nodo para llegar a la respuesta de si o no.

5.2 Criterios de diseño de la estructura de datos

A la hora de escoger el algoritmo correcto para poder realizar nuestro proyecto tomamos en cuenta varios algoritmos y nos decidimos por el algoritmo C4.5, debido a que este es una mejora avanzada del famoso algoritmo ID3, porque mejora atributos, tiene un mejor manejo de los datos de información. Este algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la prueba que resulta con mayor ganancia de información. Este algoritmo a su vez también permite trabajar con valores continuos para los atributos separando los posibles resultados en dos ramas, también los arboles son menos frondosos ya que cada hoja cubre una distribución de clase no una clase en particular. A partir de esto podemos concluir que el C4.5 era óptimo para este problema debido a que cumple con todos los requisitos de crear un árbol binario de solo dos ramas, y hace que nuestra información sea mas compacta y podamos trabajar mejor con ella.

5.3 Análisis de la Complejidad

Aunque en el peor de los casos la búsqueda en un árbol puede ser O(n) , la mayoría de las veces este tiene una complejidad de O(log n), siendo n la profundidad del árbol, sin embrago hay que considerar también la lectura y orden de los datos los cuales tienen una complejidad de O(n * m) siendo m el número de atributos y n el número de datos que hay por atributo.

5.4 Tiempos de Ejecución

Métodos	Tiempos de ejecución
Cargar Datos	Tiempo: 0.0560691357 seconds
Crear el arbol	Tiempo: 0.0051579475 seconds
Validar clasificador	Tiempo: 0.0051579475 seconds

5.5 Memoria

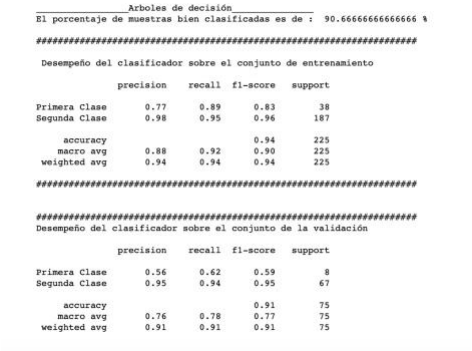
Mencionar la memoria que consume el programa para los conjuntos de datos

	Conjunto de Datos 1	*Conjunto de Datos 2*	*...Conjunto de Datos n*
Consumo de memoria	10 MB	20 MB	5 MB

Tabla 7: Consumo de memoria de la estructura de datos con diferentes conjuntos de datos

5.6 Análisis de los resultados

Como se puede evidenciar en la tabla, aunque el árbol toma buenas decisiones y se aproxima mucho a tener bastante precisión en los datos analizados, sin embargo hay un pequeño porcentaje de error debido a que hay datos que presentan diversa dispersión, además se puede observar que la primera clase tiene una precisión menor debido a que presenta menos datos(Primera clase: label = 10 y Segunda clase: label = no)



```
_____Arboles de decisión_____
El porcentaje de muestras bien clasificadas es de :  90.66666666666666 %

######################################################################

 Desempeño del clasificador sobre el conjunto de entrenamiento

              precision    recall  f1-score   support

Primera Clase       0.77      0.89      0.83        38
Segunda Clase       0.98      0.95      0.96       187

     accuracy                           0.94       225
    macro avg       0.88      0.92      0.90       225
 weighted avg       0.94      0.94      0.94       225

######################################################################

######################################################################
Desempeño del clasificador sobre el conjunto de la validación

              precision    recall  f1-score   support

Primera Clase       0.56      0.62      0.59         8
Segunda Clase       0.95      0.94      0.95        67

     accuracy                           0.91        75
    macro avg       0.76      0.78      0.77        75
 weighted avg       0.91      0.91      0.91        75
```

6. CONCLUSIONES

Los arboles de decisión son una gran herramienta para filtrar un gran número de datos, y lograra obtener conclusiones a partir de esto, en este caso especifico podemos ver la influencia de diferentes factores como el ph, la humedad, la iluminación o la temperatura en la planta del café mas específicamente en saber si son mas propensos en si pueden padecer de este fenómeno o no.

Por otra parte, es necesario resaltar la relevancia de la complejidad en los arboles, puesto que facilitan mucho algunas operaciones, como la búsqueda de información, comparándolos con otras estructuras de datos como Arrays o ArrayList. Sin embargo, hay que apoyarse de diferentes herramientas matemáticas para el uso adecuado de los datos.

REFERENCIAS

Aprende Machine Learning. (2018). Arbol de Decisión en Python: Clasificación y predicción.. [online] Available at: https://www.aprendemachinelearning.com/arbol-de-decision-en-python-clasificacion-y-prediccion/

Aprende Machine Learning. (2018). Arbol de Decisión en Python: Clasificación y predicción.. [online] Available at: https://www.aprendemachinelearning.com/arbol-de-decision-en-python-clasificacion-y-prediccion/

Cano, S. (2004). Métodos heurísticos en problemas geométricos. visibilidad, iluminación y vigilancia. [online] Dialnet. Available at: https://dialnet.unirioja.es/servlet/dctes?codigo=2840

Larrañaga, P. and Lozano, J. (2003). [online] Redalyc.org. Available at: https://www.redalyc.org/pdf/925/92571910.pdf

LÓPEZ TAKEYAS, B. (2005). [online] Itnuevolaredo.edu.mx. Available at: http://www.itnuevolaredo.edu.mx/takeyas/Apuntes/Inteligencia%20Artificial/Apuntes/tareas_alumnos/C4.5/C4.5(2005-II-B).pdf