



Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones

Programas de Maestría y Doctorado en Ingeniería Telemática
Seminario de Investigación

Detección de condiciones favorables para la ocurrencia de enfermedades y plagas en el café basada en similitud de grafos

Emmanuel Gerardo Lasso Sambony

Estudiante de Doctorado

13 de mayo de 2016

1- Introducción

En esta intervención en el seminario, será presentada la motivación de la propuesta de trabajo de grado de doctorado, un avance en el estado del arte que resume diferentes investigaciones alrededor de las áreas de interés, un avance en la evaluación y selección de una técnica de cálculo de similitud entre grafos y un conjunto de conclusiones para guiar el trabajo futuro.

2- Motivación

La Agroindustria en Colombia está posicionada como un sector de producción tradicional en el país, donde existe la necesidad de una integración de procesos para la aplicación de medidas de contingencia de manera oportuna ante eventos que generen un riesgo para los cultivos. En particular, para el sector cafetero, existen varias enfermedades tal como la roya, gotera, muerte descendente; y además plagas como la broca, que afectan la calidad y costos de su producción para el agricultor en gran medida. Debido a esto, organizaciones gubernamentales, académicos y centros de investigación especializados se han enfocado en proponer técnicas de control químico y cambios en las propiedades agronómicas de los cultivos, los cuales deben ser aplicados en un momento oportuno, con el propósito de que la enfermedad o plaga pueda ser atacada antes que genere efectos irreversibles en los cafetos (Rivillas, Serna, Cristancho, & Gaitán, 2011). En éste sentido, algunos investigadores del sector (Avelino, Willocquet, & Savary, 2004; Kushalappa & Eskes, 1989; Muller, Berry, Avelino, & Bieysse, 2004; Rivillas et al., 2011; Waller, Bigger, & Hillocks, 2007) han centrado sus esfuerzos en determinar a través del tiempo las relaciones existentes entre las condiciones climáticas y las propiedades agronómicas de los cultivos, con los episodios de epidemias de enfermedades e infestaciones de plagas. Para este propósito las autoridades cafeteras Colombianas han desplegado estaciones de monitoreo climático a lo largo del país y desarrollado una *Plataforma Agroclimática Cafetera* para el seguimiento

de la variabilidad climática en zonas cafeteras¹. Sin embargo, solo existen unas cuantas iniciativas que se enfoquen en analizar intensivamente los datos monitoreados con el fin de generar una respuesta oportuna y medidas de contingencia ante enfermedades y plagas que afectan los cafetos, a su vez generadoras de grandes pérdidas y disminución de la calidad en los cultivos.

En este sentido, son identificadas algunas tareas necesarias:

- Identificar los patrones de clima y propiedades de cultivo que generan enfermedades y plagas en el café.
- Encontrar la similitud entre las condiciones de las diferentes enfermedades y plagas, con el fin de determinar la condición de un cultivo en un determinado instante hacia estos problemas

3- Estado de la Investigación

Retomando de la intervención anterior en el seminario, para la revisión del estado del arte, fueron identificadas tres áreas importantes para la investigación:

- Detección de enfermedades y plagas en cultivos: Identificar oportunamente condiciones favorables para la ocurrencia de enfermedades y plagas.
- Aprendizaje Automático: Generar conocimiento a partir de los datos.
- Representación basada en grafos: identificación de estructuras de información similares y explotación de su naturaleza dinámica.

A partir de estas áreas, ha ido llevándose a cabo un mapeo sistemático que comprende: formulación de la pregunta de investigación, selección de la fuente de datos y estrategia de búsqueda, selección de estudios, clasificación de estudios y extracción de datos y síntesis.

A partir de este mapeo se han identificado los estudios más relevantes con las intersecciones de las áreas del estado del arte:

- Representación basada en grafos: 12 estudios.
- Detección de enfermedades y plagas en cultivos / Representación basada en grafos: 2 estudios.
- Detección de enfermedades y plagas en cultivos / aprendizaje automático: 16 estudios.
- Aprendizaje automático / Representación basada en grafos: 15 estudios.

¹ <http://agroclima.cenicafe.org/>

Desde las ciencias de la computación se han formulado diferentes propuestas (Cintra, Meira, Monard, Camargo, & Rodrigues, 2011; Corrales Muñoz, Corrales Muñoz, & Figueroa, 2015; Meira, Rodrigues, & Moraes, 2008) que pretenden generar soluciones para el sector agropecuario a partir de la aplicación de técnicas de minería y análisis de datos. En algunos de estos trabajos, los autores relacionan las distintas variables que intervienen en el desarrollo de un cultivo, con el fin de generar modelos de predicción, a partir de una base de conocimiento, que facilite determinar las condiciones de riesgos en los cultivos que generan las enfermedades o plagas. Dichos modelos de predicción tienen la opción de utilizar representaciones tal como los conjuntos de reglas y árboles de decisión, los cuales proporcionan una estructura de fácil entendimiento para el usuario y generan una predicción a partir de la definición de rangos en los valores de cada variable utilizada para tal fin. Sin embargo, dentro del café, los trabajos encontrados se enfocan en una sola enfermedad (roya), dejando de lado otras enfermedades y plagas de este cultivo donde las condiciones que las generan pueden comportarse de manera similar.

Con esto, son identificadas algunas necesidades:

- Extraer los patrones que describen las condiciones favorables para que cada enfermedad o plaga se presente en cultivos de café. Para esto, se debe mejorar el desempeño de los clasificadores usados (comúnmente de aprendizaje supervisado), a través de el aprovechamiento de la información que pueda ser extraída de datos no etiquetados presentes en el dominio de aplicación.
- Identificar la cercanía del estado de un cultivo en un momento determinado con los patrones de enfermedades y plagas. Para esto, se debe considerar una representación de la información del dominio basada en grafos, obtener una medida de similitud entre patrones y grafo de datos y parametrizar las funciones de similitud.

El aprovechamiento de la información del dominio no etiquetada puede hacerse desde una perspectiva de grafos, a partir del aprendizaje semi-supervisado basado en grafos. Esta es una formulación del aprendizaje semi-supervisado (SSL – Semi-Supervised Learning) que combina datos etiquetados (relacionados con una variable objetivo) y no etiquetados a través de una representación basada en grafos y mecanismos de propagación de etiquetas entre nodos similares, con el fin de generar clasificadores. El etiquetado de una instancia no etiquetada, a partir del peso de las relaciones entre nodos de un grafo se conoce como **Propagación de Etiqueta**. Los grafos son usados debido a que son una representación uniforme para datos homogéneos, fácilmente escalable y efectivo en la práctica frente a otras técnicas de aprendizaje automático (Subramanya & Talukdar, 2014), como se ve en la figura 1.

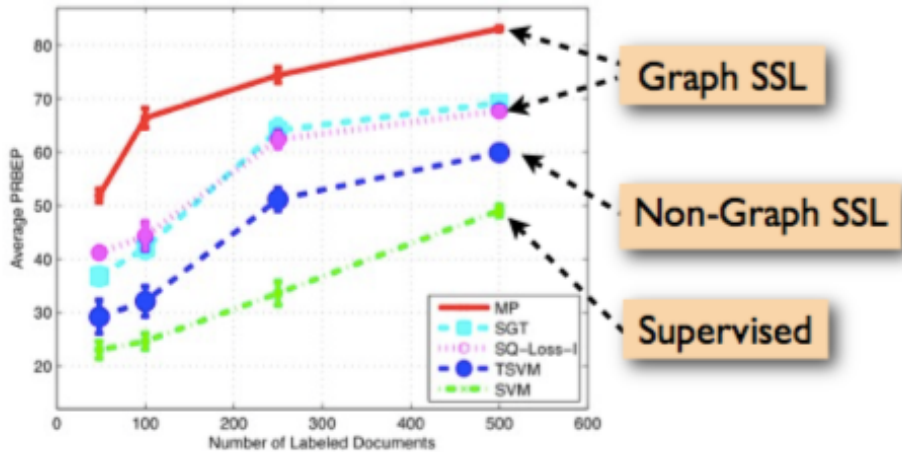


Figura 1. Comparación de técnicas de aprendizaje automático (Subramanya & Talukdar, 2014)

De manera especial, en la investigación de Tang (Tang et al., 2011) es propuesta una propagación de etiquetas de grafo basada en los k vecinos más cercanos de una instancia (representada por un nodo). Para esto, se construyen nuevas relaciones en un grafo de datos, teniendo en cuenta la similitud entre nodos vecinos y esta relación expresa la fuerza de dicha similitud. Este enfoque tiene mejor precisión que los enfoques de aprendizaje supervisado (SVM, KNN) y semi-supervisado (LNP, S-Recon), como se ve en la figura 2.

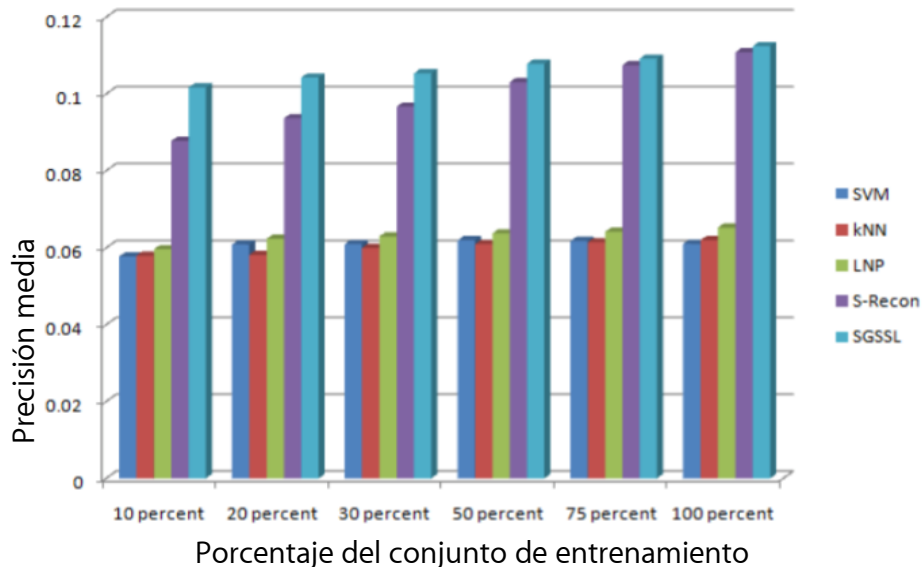


Figura 2. Comparación de ASSG basado en k vecinos cercanos con otras técnicas

Para la presente propuesta de doctorado, el comportamiento de una enfermedad o plaga en el cultivo puede ser representadas a partir de patrones de grafo. Las

estructuras de información basadas en grafos consisten en un conjunto de nodos (para el caso específico de la roya, un nodo puede representar las instancias de monitoreo de datos) que están relacionados a través de aristas (nuevamente, para el caso de la roya, una arista puede representar el contexto semántico entre variables medidas en los cultivos), los cuales tienen una naturaleza generalmente dinámica, con capacidades para la manipulación de grandes cantidades de datos. Como se puede concluir, los grafos pueden ser aprovechados para el almacenamiento y análisis de los distintos tipos de variables que están presentes en un ambiente de cultivos.

Ahora bien, la extracción del conocimiento puede darse a partir del enfoque de ASSG, que permite reducir los errores en la construcción de clasificadores y aumentar su precisión. Con esto, se podría obtener un conjunto de patrones para cada enfermedad o plaga, con lo cual puede pasarse a una fase de uso del conocimiento obtenido (patrones), buscando la similitud de los patrones representados como grafos con subgrafos presentes en un repositorio de información de cultivos.

Específicamente, en el emparejamiento de patrones en grafos de tipo exacto, los autores de (Lasso & Corrales, 2016; Lasso, Thamada, Meira, & Corrales, 2015) han desarrollado un prototipo que permite la detección de condiciones favorables para tres tasas de infección de roya en el café de Colombia, a partir de la aplicación de emparejamiento de patrones de grafo de tipo exacto, dejando de lado la posibilidad de encontrar condiciones propicias para otras enfermedades o plagas que pueden tener características similares en los patrones que las caracterizan. Por otro lado, en el caso del emparejamiento de tipo inexacto, es posible discretizar la cercanía de un patrón a un subgrafo a través de medidas como: distancia de edición de grafo (Gao, Xiao, Tao, & Li, 2010), máximo común subgrafo (Bunke, 1997) y puntaje de similitud (Zager & Verghese, 2008). Las medidas de **similitud** entre grafos deben tener en cuenta las diferencias entre los **tipos de nodo**, valores de **etiquetas** y **estructura** de los grafos comparados (Gallagher, 2006).

Un avance del trabajo de doctorado ha sido la evaluación y selección de un algoritmo para emparejamiento de patrones en grafo inexacto, con el fin de obtener medidas de similitud entre patrones y un grafo de datos de monitoreo de cultivos. Para esto, se realizó una revisión sistemática de investigaciones alrededor del emparejamiento de tipo inexacto en grafos, encontrando 18 artículos relevantes. A estos 18 artículos fue aplicado un proceso de evaluación del algoritmo o técnica usada, con criterios como:

- Número de veces usado
- Base de funcionamiento
- ¿Considera costos?
- ¿Existe implementación disponible?
- Precisión
- Tiempo de ejecución
- Complejidad

- ¿Soporta grafos etiquetados?
- ¿Soporta grafos dirigidos?

Como resultado, se eligieron dos algoritmos: Una mejora del algoritmo A* (Messmer, 1995) y Algoritmo Greedy (Dijkman, Dumas, & García-Bañuelos, 2009).

Actualmente, fue creado un generador de archivos para representación de grafos en formato GXL, a partir del grafo de datos usado y patrones extraídos en la tesis de maestría del doctorando (Lasso et al., 2015). Además, los algoritmos están siendo modificados para que estos soporten la lectura y validación de rangos dentro de las etiquetas de nodos de los grafos.

En este sentido, a partir de los datos ingresados por un usuario, el emparejamiento inexacto permitiría establecer con un alto margen de probabilidad la ocurrencia de una u otra enfermedad (roya, gota, muerte descendente) o en caso tal la presencia de una plaga (broca), basado en el análisis de las condiciones (climáticas y agronómicas) presentes en la zona donde se encuentra el cultivo (la similitud entre los patrones de grafo extraídos para cada enfermedad o plaga, pueden determinar qué tan semejantes son las condiciones que llevan a la ocurrencia de cada uno de ellas).

4- Conclusiones

A partir de lo anteriormente presentado, son identificados tres elementos o tareas relevantes para la propuesta de doctorado:

- Extracción de patrones para enfermedades y plagas del café
- Reducir el error en la construcción de clasificadores (patrones), enfocado en el uso de ASSG.
- Verificar la “cercanía” entre el estado de un cultivo y los patrones extraídos. Teniendo en cuenta los avances en la adaptación de los algoritmos para el cálculo de la similitud entre grafos.

Además, para esto fueron identificados algunos trabajos futuros:

- Caracterizar las enfermedades y plagas del café.
- Seleccionar la técnica dentro de ASSG (Aprendizaje semi-supervisado basado en grafos) con mejor desempeño y que mejor se adapte al dominio de aplicación tratado.
- Seleccionar la técnica dentro del cálculo de Similitud de Grafos con mejor desempeño y que mejor se adapte al dominio de aplicación tratado.

Bibliografía

- Avelino, J., Willocquet, L., & Savary, S. (2004). Effects of crop management patterns on coffee rust epidemics. *Plant Pathology*, 53(5), 541–547.
- Bunke, H. (1997). On a relation between graph edit distance and maximum common subgraph. *Pattern Recognition Letters*, 18(8), 689–694.
- Cintra, M. E., Meira, C. A. A., Monard, M. C., Camargo, H. A., & Rodrigues, L. H. A. (2011). The use of fuzzy decision trees for coffee rust warning in Brazilian crops. In *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on* (pp. 1347–1352). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6121847
- Corrales Muñoz, D. C., Corrales Muñoz, J. C., & Figueroa, A. (2015). Toward Detecting Crop Diseases and Pest by Supervised Learning. *Revista Ingeniería Y Universidad*, 19(1).
- Dijkman, R., Dumas, M., & García-Bañuelos, L. (2009). Graph matching algorithms for business process model similarity search. In *Business Process Management* (pp. 48–63). Springer. Retrieved from http://link.springer.com/10.1007/978-3-642-03848-8_5
- Gallagher, B. (2006). Matching structure and semantics: A survey on graph-based pattern matching. *AAAI FS*, 6, 45–53.
- Gao, X., Xiao, B., Tao, D., & Li, X. (2010). A survey of graph edit distance. *Pattern Analysis and Applications*, 13(1), 113–129.
- Kushalappa, A. C., & Eskes, A. B. (1989). Advances in coffee rust research. *Annual Review of Phytopathology*, 27(1), 503–531.

- Lasso, E., & Corrales, J. C. (2016). Expert System for Crop Disease based on Graph Pattern Matching: A proposal. *Revista Ingenierías Universidad de Medellín*, 29.
- Lasso, E., Thamada, T. T., Meira, C. A. A., & Corrales, J. C. (2015). Graph Patterns as Representation of Rules Extracted from Decision Trees for Coffee Rust Detection. In E. Garoufallou, R. J. Hartley, & P. Gaitanou (Eds.), *Metadata and Semantics Research* (pp. 405–414). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-24129-6_35
- Meira, C. A. A., Rodrigues, L. H. A., & Moraes, S. A. (2008). Analysis of coffee leaf rust epidemics with decision tree. *Tropical Plant Pathology*, 33(2), 114–124. <http://doi.org/10.1590/S1982-56762008000200005>
- Messmer, B. T. (1995). Efficient graph matching algorithms. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.33.4206>
- Muller, R. A., Berry, D., Avelino, J., & Bieysse, D. (2004). Coffee diseases. *Coffee: Growing, Processing, Sustainable Production: A Guidebook for Growers, Processors, Traders, and Researchers*, 491–545.
- Rivillas, C., Serna, C., Cristancho, M., & Gaitán, A. (2011). Roya del Cafeto en Colombia: Impacto, Manejo y Costos del Control. *Chinchiná: Boletín Técnico*, (36).
- Subramanya, A., & Talukdar, P. P. (2014). Graph-based semi-supervised learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 8(4), 1–125.

- Tang, J., Hong, R., Yan, S., Chua, T.-S., Qi, G.-J., & Jain, R. (2011). Image annotation by k nn-sparse graph-based label propagation over noisily tagged web images. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(2), 14.
- Waller, J. M., Bigger, M., & Hillocks, R. J. (2007). *Coffee pests, diseases and their management*. CABI. Retrieved from https://books.google.com/books?hl=es&lr=&id=qm54fhoV1U4C&oi=fnd&pg=PR5&dq=Coffee+pests,+diseases+and+their+management.&ots=weOQn4miqr&sig=6sldmdp3am_x7X7Heis9f26lxGc
- Zager, L. A., & Verghese, G. C. (2008). Graph similarity scoring and matching. *Applied Mathematics Letters*, 21(1), 86–94.