

CLASIFICACION AUTOMATICA PARA LA PREVENCION DEL ESTRÉS DE LOS SUELOS Y LA FATIGA DE SOJA EN EL NOROESTE ARGENTINO

Sanson, E., Britos, P., Rodriguez, D., García-Martínez, R.

Centro de Ingeniería de Software Ingeniería del Conocimiento. Escuela de Postgrado. ITBA
Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA
Área Ingeniería del Software. Licenciatura Sistemas. UNLa
Área Ingeniería del Software. Unidad Académica Río Gallegos. UNPA

{pbritos,drodrigu}@itba.edu.ar, rgarciamar@fi.uba.ar

CONTEXTO

El Proyecto articula líneas de trabajo del Proyectos “Aplicaciones de Explotación de Información basada en Sistemas Inteligentes”, con financiamiento de la Secretaria de Ciencia y Técnica de la Universidad de Buenos Aires (UBACYT 2008-2010 código I012) y acreditado por Resolución Rector-UBA N° 576/08 con radicación en el Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires

RESUMEN

El crecimiento de la soja y sus derivados en la Argentina en los últimos 30 años se ha constituido en un fenómeno trascendente y fundamental para el país, tanto desde el punto de vista económico como social. Para la campaña 2002/03, la soja representó más del 50% del total de los granos producidos, considerando los cinco cultivos mas difundidos en la pampa húmeda. Lo anterior implica que el sector agropecuario, especialmente en la pampa húmeda, ha tornado hacia un modelo de "especialización sojera", que se expande rápidamente hacia regiones extrapampeanas, al que muchos definen como el proceso de "sojificación", y el nuevo paradigma de la agricultura argentina basado en los materiales transgénicos y la siembra directa. [Begenisic, F; *et al*; 2004; Formento, N; de Souza, J.; 2007; Graciela, E; 2006; Sinavimo, 2008; Rearte, D; 2007].

Ante este presente las tecnologías informáticas han dado soporte desde lo las funciones esenciales del negocio y sin llegar a profundizar en la forma analítica-predictiva masivamente.

El presente trabajo intenta acortar dicha brecha y de esa forma por ayudar a la industria agrícola-ganadera del país, por el medio de ayudar en la prevención de la Roya de Soja y la Fatiga de los suelos. Para lograr este objetivo se utilizaran las mediciones meteorológicas en la zona noreste de la provincia de Córdoba.

Palabras clave: Minería de datos, agronomía, sistemas inteligentes.

1. INTRODUCCION

La soja ha significado en el mundo una gran revolución productiva-comercial en la cadena agroalimentaria mundial a partir de los años setenta por ser a la vez principal fuente de proteína vegetal para la alimentación animal en sistemas de cría intensiva así como proveedora de aceites para alimentación humana. Adicionalmente, en estos últimos años es también uno los cultivos claves para la obtención de biocombustibles.

Argentina es el tercer exportador de grano (dependiendo del semestre del año), luego de Estados Unidos y Brasil pero es el primer exportador mundial de productos procesados: Aceites –tanto crudo como envasados - y harinas proteicas en un nivel de oferta muy superior al de Brasil que es el segundo exportador mundial.

Si bien Argentina exporta productos procesados oleaginosos a más de 100 países en los cinco continentes, los mercados principales de aceite de soja son China e India –países de altísima demanda debido al intenso crecimiento de sus ingresos a partir de este milenio.

A continuación se describen dos situaciones que afectan a la industria agrícola-ganadera del país, la primera es la Roya de Soja y la Fatiga de los suelos. La roya de soja es un hongo que afecta a varios procesos de crecimiento de la planta tales como la capacidad fotosintética y del crecimiento del tallo, aborto de las vainas e interrupción del llenado de granos. Esta considerada como la enfermedad más importante para este tipo de sembrado que reducen los rendimientos y que pueden incluso la pérdida total de la producción de un lote.

Por otra parte, el término fatiga del suelo o estrés de los suelos engloba un amplio concepto que podría definirse como la pérdida de vigor y rendimiento productivo de las plantas cuando se efectúa un cultivo reiterado sobre un mismo suelo. Evidentemente, existe un gran número de factores que pueden, de forma más o menos conjunta, desencadenarla [[Begenisic, F; *et al*; 2004; Formento, N; de Souza, J.; 2007; Graciela, E; 2006; Sinavimo, 2008; Rearte, D; 2007].

Para este fin se tomo, en particular, el estudio de las medidas meteorológicas y como estas inciden en los suelos, para de esa forma obtener indicadores sobre

alguno de los problemas agrícolas en la zona. Puntualmente se buscara generar patrones a partir de los datos meteorológicos de esa región del país, para inferir en enfermedades referentes a los cultivos de soja, como es la roya de soja y la fatiga de los suelos. Al establecer el uso de patrones en los datos para inferir en dichas problemáticas, es donde se hará uso de las herramientas de Explotación de la Información.

2. LINEAS DE INVESTIGACION y DESARROLLO

Se denomina Explotación de la Información (data mining) al análisis de archivos y bitácoras de transacciones, trabaja a nivel del conocimiento con el fin de descubrir patrones, relaciones, reglas, asociaciones o incluso excepciones útiles para la toma de decisiones.

Las técnicas asociadas a la Explotación de la Información son diversas y se pueden nombrar las pertenecientes al aprendizaje automático, el análisis estadístico de los datos, la visualización de datos y las redes neuronales artificiales (RNA). La Explotación de la Información se refiere a la aplicación de técnica de aprendizaje automático, entre otros métodos, para encontrar importantes patrones en los datos. [García Martínez, R., *et al*, 2003].

Por lo tanto podemos establecer que un proceso de descubrimiento de conocimientos consiste de una repetición iterativa de los siguientes pasos:

- Limpieza de datos (Data Cleaning) procesamiento de los datos ruidosos, erróneos, faltantes o irrelevantes
- Integración de datos (Data integration) integración de múltiples y heterogéneas fuentes de datos en una única fuente.
- Selección de datos (Data selection) extracción de los datos relevantes al área de análisis del almacenamiento de datos.
- Transformación de datos (Data transformation) transformación o consolidación de los datos en formas apropiadas para la minería mediante procedimientos de agregación.
- Explotación de la Información: proceso esencial donde se aplican diversos métodos para extraer patrones de los datos.
- Evaluación de patrones (Pattern evaluation) identificación de patrones interesantes basándose en algún parámetro de comparación impuesto por el usuario.
- Presentación de los conocimientos (Knowledge presentation) técnicas de visualización y representación de los conocimiento

A si mismo, para este proyecto se empleo la metodología CRISP-DM [Chapman, P; *et al*, 1999], para el análisis y diseño ya que este tipo de herramienta son especificas para los distintos problemas de la Explotación de la Información con

una estructura o fases, bien definidas. En el presente trabajo se emplearon dichas fases.

La metodología CRISP-DM [Chapman, 1999], la cual es un estándar para la minería de datos y el análisis predictivo, presenta un ciclo de vida en orden jerárquico, que consiste en un conjunto de tareas que se describen en cuatro niveles de abstracción:

- Fases: por cada una de estas fases, existen un número determinado de tareas. Forman parte del Modelo Genérico de CRISP-DM.
- Tareas Genéricas: estas tareas son llamadas genéricas por que intentan cubrir todas las posibles situaciones de la minería de datos, como por ejemplo extracción de los datos.
- Tareas Específicas: estas tareas describen como algunas de las tareas genéricas van a ser efectuadas en específicas situaciones, por ejemplo extracción de los datos específicamente de humedad de los suelos en la provincia de Córdoba.
- Instancia de Procesos: es decir la instancia en que las tareas son efectuadas, las decisiones que son necesarias para dicha tarea y los resultados para esa tarea especifica de la minería de datos.

Para la metodología CRISP DM se puede distinguir entre la referencia del modelo y la guía del usuario. La primera presenta una rápida vista general de las fases del modelo, tareas y las salidas esperadas en el modelo de minería de datos.

En el presente trabajo y ante la problemática antes descrita, en una primera instancia se puede observar como las mediciones meteorológicas que son generadas por la consola de medición climática Groweather, estas se exportan en un archivo de texto (archivo plano) hacia el servidor. Por medio de un básico proceso de extracción, transformación y carga de los datos, estos son llevados a la base de datos para de esa forma ser depurados y relacionados con otros datos.

En este punto, vale destacar que entre los datos con los que se cuentan provenientes de la consola de Groweather son fecha y hora de la muestra, temperatura máxima, temperatura media, humedad relativa del ambiente, radiación solar, velocidad del viento, sensación térmica entre otros.

Con los datos ya procesados y relacionados entre si, se puede emplear cualquiera de las herramientas de Explotación de la Información para de esta forma poder inferir en la problemática vigente. Dicho de otra forma, se podría generar por ejemplo, un árbol de decisiones a partir del modelo ID3 por el cual obtendríamos la precisión para saber si ante determinado rango de temperatura media, humedad relativa del ambiente y en determinada temporada del año hay un alto riesgo de que los suelos de esa zona geográfica puedan tener un índice de fatiga alto o no.

3. RESULTADOS OBTENIDOS/ESPERADOS

Se utiliza una muestra de 731 casos, a los cuales se le aplicó en primera instancia dos métodos inductivos C4.5 y MC4 para determinar el stress del suelo. Pudiéndose obtener los siguientes resultados:
Reglas que se desprenden del árbol aplicando C4.5:

Estado del suelo: Estrés

1. Si la Temperatura Media es mayor o igual a 26,05 entonces los suelos están en el estado de Estrés
2. Si la Temperatura Media es mayor o igual a 25,55 y menor a 26,05 y la Humedad Relativa es menor o igual a 69,9 entonces los suelos están en el estado de Estrés.

Estado del suelo: Normal

3. Si la Temperatura Media es menor a 22,25 entonces los suelos están en el estado Normal.
4. Si la Temperatura Media es mayor o igual a 22,25 y menor a 22,85 y la Humedad Relativa es menor a 66,45 entonces los suelos están en el estado Normal.

Estado del suelo: Fatiga

5. Si la Temperatura Media es mayor o igual a 22,85 y menor a 25,55 entonces los suelos están en el estado de Fatiga.
6. Si la Temperatura Media es mayor o igual a 22,25 y menor a 22,85 y la Humedad Relativa es mayor o igual a 66,45 entonces los suelos están en el estado de Fatiga.
7. Si la Temperatura Media es mayor o igual a 25,55 y menor a 26,05 y la Humedad Relativa es mayor a 69,9 entonces los suelos están en el estado de Fatiga.

Los resultados del modelo son representativos de la realidad de la muestra. Se puede ver como en el estudio de los datos, para los del tipo Normal, el modelo solo acertó en el 99,57% de los casos. En el caso de los del tipo Fatiga, acertó en el 98,9% de los casos y en el tipo Estrés acertó el 92,7% de los casos.

A continuación se aplica el método MC4 (este método es similar al C4.5 pero se distingue por ser más sensitivo a los costos que el anterior. Emplea una generalización de la estimación de Laplace, con lo que minimiza las pérdidas en la clasificación).

Estado del suelo: Estrés

1. Si la Temperatura Media es mayor o igual a 26,05 entonces los suelos están en el estado de Estrés
2. Si la Temperatura Media es mayor o igual a 25,55 y menor a 26,05 y la Humedad Relativa es menor o igual a 69,9 entonces los suelos están en el estado de Estrés.

Estado del suelo: Normal

3. Si la Temperatura Media es menor a 22,25 entonces los suelos están en el estado Normal.
4. Si la Temperatura Media es mayor o igual a 22,25 y menor a 22,85 y la Humedad Relativa es menor a 70,9 entonces los suelos están en el estado Normal.

Estado del suelo: Fatiga

5. Si la Temperatura Media es mayor o igual a 22,85 y menor a 25,55 entonces los suelos están en el estado de Fatiga.
6. Si la Temperatura Media es mayor o igual a 22,25 y menor a 22,85 y la Humedad Relativa es mayor o igual a 70,9 entonces los suelos están en el estado de Fatiga.
7. Si la Temperatura Media es mayor o igual a 25,55 y menor a 26,05 y la Humedad Relativa es mayor a 69,9 entonces los suelos están en el estado de Fatiga.

Se puede ver como al comparar las reglas generadas por el método C4.5 y el MC4 las mismas son iguales en general. La diferencia se puede ver en el ajuste de valores de las reglas 4 y 6 en donde el método MC4 discrimina por valores de la Humedad Relativa en 70,9 en lugar de 66,45 del C4.5.

Seguidamente a este estudio se utilizarán redes neuronales SOM para agrupamiento y posteriormente inducción para determinar reglas de comportamiento en grupo de acuerdo a los procesos planteados en [Britos, P; 2008]

4. FORMACION DE RECURSOS HUMANOS

En la línea de investigación cuyos resultados parciales se reportan en esta comunicación, se encuentran trabajando: dos tesis de maestría, un tesis de especialización y tres investigadores formados.

5. BIBLIOGRAFIA

- Begenisic, F., Ploper, D., Ivancovich, A. 2004. Plan Nacional de Soja de la Soja. www.sinavimo.gov.ar/?q=node/3123. Vigente a 03/09.
- Britos, P. 2008. Procesos de Explotación de Información Basados en Sistemas Inteligentes. Tesis Doctoral. Facultad de Informática UNLP
- Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 1999. CRISP-DM 1.0 Step by step BGuide. www.crispdm.org/CRISPWP-0800.pdf. Vigente a 03/09
- [Formento, N., de Souza, J. 2007. Folleto Soja De La Soja – INTA EEA Paraná – SAGPyA. 2007. www.sagpya.mec.gov.ar/new/0-0/nuevositio/agricultura/sanidad/index.php?fondo_agri_01=Sanidad&fondo_agri_sanidad=soja. Vigente a 03/09.
- García Martínez, R., Servente, M. y Pasquini, D. 2003. Sistemas Inteligentes. Editorial Nueva Librería. Buenos Aires.
- Graciela, E; 2006. Grupo Economía Agrícola. 2006. Márgenes Brutos de los Principales Cultivos. www.inta.gov.ar/saenzpe/info/documentos/economia/2006/2008-12s-1.pdf. Vigente a 03/09
- Rearte, D. 2007. Situación de la Ganadería Argentina en el contexto Mundial. www.inta.gov.ar/balcarce/carnes/sitgan.htm. Vigente a Marzo 2009.
- Sinavimo. P. 2008. Soja de la soja en Argentina. www.planetasoja.com/trabajos/trabajos800.php?id1=27914&publi=&idSec=107&id2=27915. Vigente a 03/09.