การวิจัยเชิงสำรวจ - การพยากรณ์ศัตรูพืชโดยใช้วิธีการเรียนรู้ของเครื่องกล

Survey of Crop Pests Prediction by using Machine Learning Techniques

Maniphone Khenphimsoulivong[1], นทพล แสงกล้า[1], รัฐปัตย์ ประทุมศิริ[1], ธิดารัตน์ วันแสวง[1]

คณะวิทยาศาสตร์ ภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

Abstract – This paper has aims to present the prediction of crop pests by using and comparing many algorithms of Machine Learning techniques. This paper also classifies and introduces Multiple Linear Regression (MLR), Generalized Regression Neural Network (GRNN), Support Vector Machine (SVM), Bayesian Network, Neural Network, and describes some cases of their utilization. The predicted results can help the farms to reduce damages and increase their income, and also helpful in term of pest control and pest eradication.

บทคัดย่อ - บทความนี้ได้นำเสนอถึงการพยากรณ์โดยใช้วิธี Machine Learning เข้าในงานทางด้านวิชาการเกษตร, เพื่อทำนายการระบาดของ ศัตรูพืช และ โรคระบาดพืช. โดยได้กล่าวถึง อัลกอริธึม เช่น: Multiple Linear Regression (MLR), Generalized Regression Neural Network (GRNN), Support Vector Machine(SVM), Bayesian Network, และ ตัวแบบจำลอง Neural Network พร้อมยกตัวอย่าง และ อภิปรายผล โดยมีเป้าหมายหลักคือ ความแม่นยำ ของการพยากรณ์. การพยากรณ์ศัตรูพืช ยังเป็นการสนองข้อมูล เพื่อใช้ในการตัดสินใจ และ นำใช้เข้าในการวางแผน เพื่อควบคุมผลผลิต ไปจนถึงการปราบปรามศัตรูพืช.

คำสำคัญ: Machine Learning, SVM, Regression, Neural Network , Bayesian

1. บทนำ

การเกษตรนั้นมีความสำคัญมากในการขับเคลื่อนทุกกลไกภายในโลกนี้มาตั้งแต่ช้านาน, ไม่ว่าจะเป็นการสนองวัตถุดิบ อาหาร ข้าวของ อุปโภค บริโภค แล้วยังเป็นหัวใจหลักของระบบเศรษฐกิจแทบทุกประเทศภายในโลก หรือจะพูดว่า การเกษตรก็คือส่วนหนึ่งของสิ่งแวดล้อมก็ว่าได้.

ความท้าทายหลักๆในงานเกษตรก็คือการรักษาไว้ซึ่งทรัพยากร หรือ อาหาร ไม่ให้ขาดแคลน และสามารถนำใช้ไปได้เรื่อยๆจนถึงอานาคต, การพยากรณ์จึงเป็นสิ่งสำคัญอย่างมาก.

การพยากรณ์ศัตรู และ โรคระบาดพืชไม่ใช่แค่การทำนายทั่วไป, แต่ยังเป็นการสนองข้อมูล เพื่อใช้ประกอบการตัดสินใจ และ นำใช้เข้าในแผนงานต่างๆ เพื่อควบคุมผลผลิต ไปจนถึงการปราบปรามศัตรูพืช.

2. ทฤษฎีที่เกี่ยวข้อง

2.1. Multiple Linear Regressions (MLR)

เป็นเทคนิคการวิเคราะห์การถดถอยที่เกี่ยวข้องกับตัวแปรอิสระที่มากกว่าหนึ่งตัวแปร การเพิ่มตัวแปรอิสระที่เกี่ยวข้องเข้าในการวิเคราะห์จะทาให้ ความถูกต้องของการวิเคราะห์เพิ่มมากขึ้นและค่าคลาดเคลื่อนมาตรฐานของตัวประมาณค่า (standard error of estimates) จะลดลง.

2.2. Generalized Regression Neural Network (GRNN)

เป็นเครื่องมือที่มีประโยชน์มากในการพยากรณ์และการเปรียบเทียบประสิทธิภาพของระบบ นอกจากนี้ GRNN ได้ถูกจัดว่าเป็นโครงข่าย แบบใช้ความจำประเภทหนึ่งที่มีตัวประมาณค่าตัวแปรแบบต่อเนื่อง และลู่เข้าสู่ผิวหน้าถดถอยทั้งแบบเชิงเส้น (Linear) และที่ไม่ใช่เชิงเส้น (Nonlinear)

2.3. Support Vector Machine (SVM)

เป็นอัลกอริทึมในการคัดแยกที่มีการนำมาใช้กันอย่างกว้างขวางในด้านการประมวลผลเป็นภาพดิจิตอล หลักการของ SVM คือการให้อินพุทที่ใช้ฝึกเป็นเวคเตอร์ในสเปซ N มิติ เช่นถ้าในกรณีของ 2 มิติ และ 3 มิติ จะเป็นจุดที่อยู่ในระนาบ xy และสเปซ xyz ตามลำดับจากนั้นทำการสร้างไฮเปอร์เพลน (Hyperplane) ที่จะแยกกลุ่มของเวคเตอร์อินพุทออกเป็นประเภทต่างๆ ในกรณีที่เป็น 2 มิติ และ 3 มิติ ไฮเปอร์เพลน คือเส้นตรงและระนาบตามลำดับ ข้อเด่นของ SVM จะทำการเก็บแมพ (Map) เวคเตอร์ในสเปซอินพุทให้เข้าสู่ Feature Space โดยใช้ฟังก์ชั่นหรือเรียกว่าเคอร์นัล (kernel) ชนิดต่างๆ เช่น โพลีโนเมียล (Polynomial) เรเดียล (Radial) เป็นต้น.

2.4. Bayesian Network

เป็นแบบจำลองกราฟของความน่าจะเป็น ซึ่งเป็นวิธีการเรียนรู้ที่ลดข้อจำกัดของการเรียนรู้แบบเบย์อย่างง่ายในสมมุติฐานความไม่ขึ้นต่อกันระหว่างคุณสมบัติ โดย Bayesian network ใช้อธิบายความไม่ขึ้นต่อกันอย่างมีเงื่อนไข (Condition Independent) ระหว่างตัวแปร.

2.5. Neural Network

คือตัวแบบจำลองทางคณิตศาสตร์หรือโมเดลทางคอมพิวเตอร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (connectionist) แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (bioelectric network) ในสมอง ซึ่งประกอบด้วย เซลล์ประสาท (neurons) และ จุดประสานประสาท (synapses) ตามโมเดลนี้ ข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาท จนเป็นเครือข่ายที่ทำงานร่วมกัน.

3. งานวิจัยที่เกี่ยวข้อง

3.1. การพยากรณ์สารพิษ Deoxynivalenol ในข้าวสาลี โดยใช้ Multiple Linear Regression (MRL)

จากงานวิจัยของ D. C. Hooker และ คณะ, ได้พูดถึง ระยะเวลา 3 ขั้น และ เงื่อนไขของการเกิดสารพิษ Deoxynivalenol (DON) จากเชื้อรา (Mycotoxin) โดยใช้ Multiple Linear Regression (MRL) เพื่อพยากรณ์อัตราการเกิดสารพิษ DON. โดยกลุ่มตัวอย่างคือข้อมูลปี 1996-2000 จากฟาร์ม 399 แห่ง ในรัฐ Ontario ทางตอนใต้ของ Canada; กลุ่มข้อมูลจะเน้นไปที่ปัจจัยทางสภาพอากาศ ซึ่งประกอบด้วย ปริมาณน้ำฝนรายวัน, อุณหภูมิรายวัน ต่ำสุด/สูงสุด,และ ค่าความชื้นสัมพันธุ์ต่อชั่วโมง.

จากผลการทดลองสรุปได้ว่า, ช่วงเวลาเติบโตของข้าวสาลี, ปริมาณน้ำฝน และ อุณหภูมิ มีส่วนในการสร้างเชื้อรา (Mycotoxin) และทำให้เกิดสารพิษ Deoxynivalenol (DON). ในขณะเดียวกันกลับพบว่าความชื้น ไม่ได้มีส่วนเกี่ยวข้องกับอัตราการเกิดสารพิษเลย, โดยอ้างอิงจากการพยากรณ์ ได้ให้ความแม่นยำถึง 89%.

3.2. การพยากรณ์ ระดับความชื้นของใบไม้ โดยใช้ Generalized Regression Neural Network (GRNN)

Younes Chtioui และคณะ ได้กล่าวไว้ในบทความของพวกเขาว่า ระดับความชื้นของใบไม้ นั้นเป็นปัจจัยสำคัญของการระบาดของโรคพืช, ในงานวิจัยนี้ ได้วิเคราะห์ปัจจัยทางอุตุนิยมวิทยาโดยได้พัฒนา Generalized Regression Neural Network (GRNN) เพื่อพยากรณ์ระดับความชื้นของใบ.

หลักๆในงานวิจัยนี้ คือการเปรียบเทียบข้อผิดพลาดของการทำนายค่าสัมบูรณ์เฉลี่ย (Average absolute value prediction errors) ระหว่างวิธี Generalized Regression Neural Network (GRNN) และ Multiple Linear Regression(MLR). โดยใช้ปัจจัยหลักคือ เวลา (ภายใน 24 ชั่วโมง), อุณหภูมิ, ค่าสัมพันธุ์ความชื้น (Relative moisture), ความเร็วลม, รังสีดวงอาทิตย์ (Solar radiation), ฝน, ดัชนีความชุ่มชื้นในดิน (Soil Moisture Indexes) ซึ่งข้อมูลภายในปัจจัย เหล่านี้ได้ถูกนำไปใช้เป็น test set และ training set.

จากการทดลองสรุปผลออกมาว่า, MLR ได้แสดงค่าข้อผิดพลาดของการทำนายค่าสัมบูรณ์เฉลี่ย (Average absolute value prediction errors) คือ 0.1414 สำหรับ test set และ 0.1300 สำหรับ training set. ในขณะเดียวกัน, GRNN ได้แสดงค่าข้อผิดพลาดของการทำนายค่าสัมบูรณ์เฉลี่ย (Average absolute value prediction errors) คือ 0.0491 สำหรับ test set และ 0.0894 สำหรับ training set. จึงสรุปออกมาได้ว่า GRNN มีความแม่นยำสูงกว่า MLR.

3.3. การพยากรณ์ศัตรูพืช โดยใช้วิธี Naïve Bayes และ Wireless Sensor Network

จากตัวอย่างในงานวิจัยของ A. K. Tripathy และคณะ, พวกเขาได้พัฒนาระบบตัวช่วยการตัดสินใจแบบ real-time ซึ่งสามารถให้ผลการพยากรณ์ศัตรูพืช โดยนำขั้นตอนการทำเหมืองข้อมูล ประยุคใช้กับ wireless sensor network ออกมาเป็นวิธี Gaussian Naive Bayes พร้อมกฎ Rapid Association Rule Mining. ผลลัพธ์ที่ได้คือ ระบบสามารถพยากรณ์ว่าเมื่อไหร่ศัตพืชรูจะระบาด โดยใช้ข้อมูลจาก wireless sensor network ของแต่ละจุด.

3.4. การพยากรณ์โรคราสนิมของต้นกาแฟ โดยใช้ Bayesian Network

Cora B. Perez-Ariza และ คณะได้ทำการวิจัยเกี่ยวกับวิธีการพยากรณ์อัตราการเกิดของโรคราสนิมของต้นกาแฟ โดยใช้วิธี Bayesian Network มาสร้างตัวแบบจำลองในการพยากรณ์ ซึ่งใช้ข้อมูลกลุ่มตัวอย่างปี 1998 - 2006 จากไร่กาแฟในประเทศบราซิล, ตัวแปรหลักในงานวิจัยนี้คือสภาพอากาศ.

3.5. การทำนายโรคราสนิมที่ใบของข้าวสาลีโดยการเปรียบเทียบวิธี Regression กับ SVM

Haiguang Wang และ Zhanhong Ma ได้กล่าวไว้ในบทความว่า สำหรับโรคนี้เป็นหนึ่งในโรคที่สำคัญที่สุดในประเทศจีน การป้องกันด้วยวิธีพยากรณ์ล่วงหน้าจึงเป็นสิ่งสำคัญมาก โดยโรคนี้เกิดจากเชื้อราชื่อ Puccinia recondita มักพบในฤดูปลูกข้าวสาลีที่มีอากาศร้อนชื้น มีแผลเกิดขึ้นบนใบ ซึ่งเชื้อราสร้างขึ้นมา อาการของโรคจะเริ่มปรากฏจากใบล่าง แล้วลามขึ้นไปส่วนยอด และผลการศึกษาพบว่าวิธี SVM มีประสิทธิภาพมากกว่า

3.6. พยากรณ์การติดเชื้อจากหนอนซอนใบโดยใช้วิธี SVM

Wu Dake and Ma Chengwei ได้ดำเนินงานเกี่ยวกับการพยากรณ์ของการติดเชื้อจากหนอนชอนใบด้วยวิธีการ SVM ในการศึกษา สกัดข้อมูลโดยใช้การประมวลผลภาพ image processing และเทคโนโลยีการวิเคราะห์สเปกตรัม (spectrum analysis) และจากนั้นคาดการณ์การติดเชื้อโดยการจำแนกองศาความเสียหายของใบเนื่องจากการติดเชื้อกับสเปกตรัมและความยาวคลื่นที่สะท้อนจากใบ และ ผลที่มาจากการทดสอบ ได้อัตราความแม่นยำมากกว่า 90% โดยวิธีนี้อาจช่วยในการตรวจสอบแมลงศัตรูพืชได้อย่างรวดเร็ว.

งานนี้ได้นำเสนอเทคนิค Near Infrared Spectroscopy (NIRS) เป็นการใช้พลังงานคลื่นแม่เหล็กไฟฟ้าย่านความถี่ใกล้อินฟราเรด ที่อยู่อยู่บนพื้นฐานของ SVM โดยส่องไปยังใบมะเขือเทศเพื่อให้เกิดการดูดกลืนแสง แล้วมีการเปลี่ยนแปลงของพันธะระหว่างอะตอมที่เกิดการยืด-หด หรือบิด-งอ ในรูปแบบต่างๆ แล้วทำการตรวจวัดคลื่นแสงที่ไม่ถูกดูดกลืน ซึ่งสะท้อนหรือส่องผ่านออกมา แล้วประมวลผลเป็นค่าการดูดกลืนแสงที่ความยาวคลื่นต่างๆ เกิดเป็นสเปคตรัมที่มีลักษณะเฉพาะตัว เพื่อนำมาเทียบกับค่าองศาความเสียหายของใบ damaged degrees (DD)

3.7. ตัวแบบจำลองโครงข่ายประสาทเทียม เพื่อพยากรณ์การเกิดสารพิษ Deoxynivalenol (DON) ของข้าวสาลี โดยใช้ข้อมูลสภาพอากาศ และ ข้อมูลก่อนการปลูก

ในงานวิจัยนี้ K. Klem และ คณะ ได้ทำการพัฒนาตัวแบบจำลอง เพื่อใช้ในการพยากรณ์ การเกิดสารพิษ Deoxynivalenol (DON) ของข้าวสาลี, โดยใช้ข้อมูลของสภาพอากาศ และ สภาพพื้นที่ก่อนทำการเพาะปลูก เพื่อเป็นตัวแปรในการพยากรณ์ ผ่านโครงข่ายประสาทเทียม (Neural Network)

3.8. การพัฒนาตัวแบบจำลองพยากรณ์การเกิดสารพิษ Deoxynivalenol (DON) ของข้าวสาลี โดยใช้ข้อมูลสภาพอากาศ

Marie Vánová ได้นำงานของเดิมของ K. Klem มาต่อยอดแล้วพัฒนาตัวแบบการพยากรณ์ผ่านโครงข่ายประสาทเทียม, โดยเปลี่ยนมาใช้ข้อมูลสภาพอากาศรายเดือน ใน ออสเตรเลีย ตั้งแต่ปี 2002 - 2005. ภายหลังพัฒนาสำเร็จ พวกเขาได้เริ่มนำใช้ และ เรียกระบบนั้นว่า AgriClim, ซึ่งสามารถพยากรณ์ระยะเวลา และ อัตราการเกิดของสารพิษ Deoxynivalenol (DON) โดยใช้ตัวแปรหลักแค่สภาพอากาศ.

4. ข้อสรุป

ในบทความนี้ได้พิจารณาหลายๆบทความที่เกี่ยวข้องกับการพยากรณ์โดยใช้วิธี Machine Learning เข้าในงานทางด้านวิชาการเกษตร, ซึ่งสามารถสรุปออกมาได้ใน ตารางที่ 4. จากการสังเกตโดยรวมทำให้เรารู้ว่า วิธีที่ใช้ในการพยากรณ์นั้นมีหลากหลาย แต่ผลลัพธ์ที่ได้นั้นคล้ายกัน โดยมีเป้าหมายหลักคือความแม่นยำ ของการพยากรณ์.. ผลการพยากรณ์นี้ จะเป็นตัวช่วยในการตัดสินใจ การจัดทำระบบ ข้อมูลเพื่อการเกษตร ซึ่งอาจมีทั้งระดับท้องถิ่น ระดับชาติ และ ระดับนานาชาติ, อีกทั้งยังสามารถนำไปประยุคใช้เข้าในแผนปฏิบัติงานภายในองค์กร หรือ เพื่อควบคุมผลผลิต ไปจนถึงการปราบปรามศัตรูพืช.

No.	เป้าหมายของการพยากรณ์	กลุ่มตัวอย่าง	วิธีที่ใช้	ผลลัพธ์
1	การพยากรณ์สารพิษ Deoxynivalenol	ข้าวสาลี	Multiple Linear Regression (MLR)	ผลการพยากรณ์ ให้ความแม่นยำ 89%
2	ระดับความชื้นของใบไม้	ใบไม้	GRNN Vs. MLR	GRNN มีความแม่นยำสูงกว่า MLR
3	การระบาดของศัตรูพืช	ทั่วไป	Gaussian Naïve Bayes, Rapid Association Rule Mining	แสดงผลพยากรณ์แบบ Real-time ว่าเมื่อไหร่ศัตพืชรูจะระบาด โดยใช้ข้อมูลจาก wireless sensor network ของแต่ละจุด.
4	โรคราสนิม	ต้นกาแฟ	Bayesian Network	สามารถพยากรณ์อัตราการเกิดของโรคได้, แต่ไม่มีการประเมินผลความแม่นยำ
5	โรคราสนิม	ข้าวสาลี	SVM	สามารถพยากรณ์อัตราการเกิดของโรคได้, แต่ไม่มีการประเมินผลความแม่นยำ
6	อัตรการติดเชื้อจากหนอนชอนใบ		SVM	สามารถพยากรณ์อัตราการติดเชื้อได้, แต่ไม่มีการประเมินผลความแม่นยำ
7	การเกิดสารพิษ Deoxynivalenol (DON)	ข้าวสาลี	Neural Network	สามารถสร้างตัวแบบการพยากรณ์, แต่ไม่มีการประเมินผล
8	การเกิดสารพิษ Deoxynivalenol (DON)	ข้าวสาลี	Neural Network	สามารถพัฒนา AgriClim ซึ่งเป็นตัวแบบการพยากรณ์, แต่ไม่มีการประเมินผล

ตาราง 1. การเปรียบเทียบผลที่ได้จากงานที่เกี่ยวข้อง

5. อ้างอิง

1. D. C. Hooker. (2002). Using Weather Variables Pre- and Post- heading to Predict Deoxynivalenol Content in Winter Wheat, Plant Disease, vol. 86, pp. 611–619.
2. Younes Chtioui. (1999). A generalized regression neural network and its application for leaf wetness prediction to forecast plant disease, Chemometrics and Intelligent Laboratory Systems, vol. 48, pp. 47–58.
3. A. K. Tripathy. (2011). Data Mining and Wireless Sensor Network for Agriculture Pest/Disease Predictions. In: 2011 World Congress on Information and Communication Technologies, pp. 1229-1234.
4. Cora B. Perez-Ariza. (2012). Prediction of Coffee Rust Disease Using Bayesian Networks. In: 6th European Workshop on Probabilistic Graphical Models, pp. 259-266.
5. Haiguang Wang. (2011). Prediction of Wheat Stripe Rust Based on Support Vector Machine. In: 2011 Seventh International Conference on Natural Computation, pp. 378–382.
6. Wu Dake. (2006). The Support Vector Machine (SVM) Based Near-Infrared Spectrum Recognition of Leaves Infected by the Leafminers. In: First International Conference on Innovative Computing, Information and Control, vol. 3, pp. 448-451.
7. K. Klem. (2007). A neural network model for prediction of deoxynivalenol content in wheat grain based on weather data and preceding crop, Plant Soil and Environment, vol. 53, pp. 421-429.
8. Marie Vánová. (2009). Prediction Model for Deoxynivalenol in Wheat Grain Based on Weather Conditions, Plant Protection Science, vol. 45, pp. S33–S37.