

การทำนายปริมาณสารส้มที่ใช้ในกระบวนการผลิตน้ำประปาด้วยวิธีป่าของต้นไม้
ตัดสินใจและโปรแกรมเชิงพันธุกรรม

(Alum Dosage using in Water Supply Process Prediction by Decision Tree Forest
Method and Genetic Programming)

ปกรณ์ จารุตระกูลชัย, สุกรี สิ้นธุภิญโญ

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาวิธีการที่เหมาะสมในการทำนายปริมาณสารส้มที่ใช้ในกระบวนการผลิตน้ำประปา โดยใช้ซัพพอร์ตเวกเตอร์แมชชีน ต้นไม้ตัดสินใจ และป่าของต้นไม้ตัดสินใจ เปรียบเทียบกับวิธีการแบบนิวรอลเน็ตเวิร์กที่ถูกนำมาพัฒนาเพื่อใช้ในงานวิจัยส่วนใหญ่เพื่อทำนายปริมาณการเติมสารโคแอกกูแลนต์ ตัวแปรอินพุตเริ่มต้นที่นำมาใช้ในการทำนายปริมาณการเติมสารส้มมีทั้งหมด 7 ตัว ได้แก่ ความขุ่น ความเป็นด่าง พีเอช การนำไฟฟ้า สี ปริมาณสารแขวนลอย และแอมโมเนียไนโตรเจน จากนั้นทดลองปรับเปลี่ยนและลดจำนวนตัวแปรอินพุต โดยงานวิจัยนี้ได้เสนอเทคนิคการสร้างตัวแปรอินพุตเพิ่มเติมด้วยการประยุกต์ใช้โปรแกรมเชิงพันธุกรรม และทำการทดลองลดตัวแปรอินพุตโดยใช้เฉพาะตัวแปรอินพุตหลักๆ ได้แก่ ความขุ่น ความเป็นด่าง และพีเอช สำหรับข้อมูลที่นำมาใช้ในฝึกสอนให้เกิดการเรียนรู้เป็นข้อมูลที่ได้จากสำนักงานประปาสาขาบางเขน การประปานครหลวง ที่เก็บในช่วงระยะเวลา 1 ปี ตั้งแต่วันที่ 1 มกราคม 2549 ถึง 31 ธันวาคม 2549 ผลการศึกษาพบว่าในกรณีตัวแปรอินพุตแบบ 7 ตัว นิวรอลเน็ตเวิร์กเป็นวิธีที่สามารถทำนายได้แม่นยำที่สุด กรณีอินพุตแบบ 8 ตัว ซึ่งเป็นการเพิ่มค่าคุณลักษณะใหม่ 1 ค่า ด้วยวิธีโปรแกรมเชิงพันธุกรรม วิธีป่าของต้นไม้ตัดสินใจเป็นวิธีที่ทำนายได้แม่นยำที่สุด กรณีอินพุต 3 ตัว ซึ่งมีเฉพาะตัวแปรอินพุตหลัก นิวรอลเน็ตเวิร์กเป็นวิธีที่ทำนายได้แม่นยำที่สุด และเมื่อเปรียบเทียบวิธีในการทำนายทุกวิธีพบว่าวิธีที่สามารถทำนายผลได้แม่นยำที่สุดได้แก่ วิธีป่าของต้นไม้ตัดสินใจโดยประยุกต์ใช้โปรแกรมเชิงพันธุกรรมในการสร้างตัวแปรอินพุตเพิ่มเติม

Abstract

This paper presents a comparison of prediction methods for alum dosage used in water supply treatment process. Neural networks which is a common method has been used in many works. In this research, we compared results from support vector machine, decision tree and decision tree forest to those from neural networks. Seven input variables, i.e. turbidity, alkalinity, pH, conductivity, color, suspended solids, and $\text{NH}_3\text{-N}$, relating to reaction of coagulation were used. Then we inserted a new input variable by applying genetic programming and decrease input variables by using only main input variables, i.e. turbidity, alkalinity and pH. The data in this research had been collected from Bangkok Branch Office of Metropolitan Waterworks Authority from 1 January 2006 to 31 December 2006. Our experimental results showed that neural networks yielded the most accuracy in case of seven input variables. Decision tree forest yielded the most accuracy in case of eight input variables (seven input variables with a new input variable generated from genetic programming). Neural networks yielded the most accuracy in case of three input variables (only three main variables). And decision tree forest with eight input variables yielded the highest accuracy compared to all cases.