

บทที่ 3

ระเบียบวิธีการศึกษา

การศึกษาการพยากรณ์ราคาหลักทรัพย์และดัชนีตลาดหลักทรัพย์โดยใช้แบบจำลองโครงข่ายประสาทเทียมเปรียบเทียบกับวิธีอาร์มา

ขอบเขตและวิธีการศึกษา

1. ขอบเขตการศึกษา

เนื้อหาในการศึกษาค้นคว้าครั้งนี้ประกอบด้วยการหารูปแบบของการพยากรณ์ดัชนีราคาหลักทรัพย์และดัชนีตลาดหลักทรัพย์แห่งประเทศไทย โดยใช้ข้อมูลราคาปิดรายวันของหลักทรัพย์รายตัวในกลุ่ม SET50 จำนวน 50 ตัวและค่าดัชนี SET และดัชนี SET50 ย้อนหลังตั้งแต่ พ.ศ. 2548 ถึง 2550 จากทางระบบอินเทอร์เน็ต www.set.or.th, www.setsmart.com และ www.bot.or.th มาผ่านกระบวนการเรียนรู้โครงข่ายประสาทเทียม

2. วิธีการศึกษา

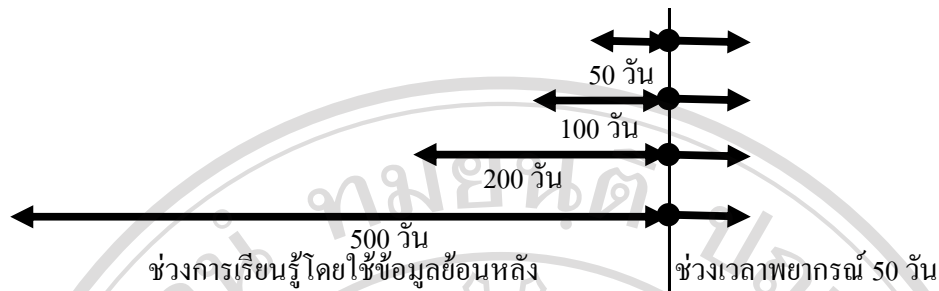
2.1 การเก็บรวบรวมข้อมูล

ใช้ข้อมูลทุติยภูมิ (Secondary Data) ซึ่งข้อมูลที่ได้ศึกษาในครั้งนี้ได้รวบรวมข้อมูลราคาปิดรายวันของดัชนี SET และ ดัชนี SET50 และหลักทรัพย์รายตัว 50 ตัว ซึ่งอยู่ในกลุ่ม SET50 ระหว่างช่วงเวลา 3 ปี ตั้งแต่ วันที่ 1 มกราคม 2548 – 31 ธันวาคม 2550 โดยเก็บข้อมูลจากทางระบบอินเทอร์เน็ต www.set.or.th, www.setsmart.com และ www.bot.com และเก็บข้อมูลจากศูนย์การเงินและการลงทุนมหาวิทยาลัยเชียงใหม่ (FIC) และ ศูนย์การเรียนรู้ทางการเงิน (Financial Street@CMU)

2.2 การวิเคราะห์ข้อมูล

1. แบ่งข้อมูลตั้งแต่วันที่ 1 มกราคม 2548 – 31 ธันวาคม 2550 รวมทั้งสิ้น 733 วันออกเป็น 2 ส่วนคือ ส่วนแรกคือส่วนที่นำไปวิเคราะห์ตามแบบจำลองโครงข่ายประสาทเทียมและวิธีอาร์มาโดยสุ่มเลือกเป็นข้อมูลช่วง 50 วัน, 100 วัน, 200 วัน และ 500 วันของช่วงก่อน 50 วันสุดท้าย และส่วนที่ 2 เป็นส่วนที่ใช้เปรียบเทียบกับค่าพยากรณ์ที่ได้จากแบบจำลองโครงข่ายประสาทเทียมในแบบต่างๆ โดยใช้ข้อมูลในช่วง 50 วันสุดท้าย ดังแสดงในรูปที่ 3.1

รูปที่ 3.1 แสดงช่วงเวลาที่ใช้เรียนรู้และช่วงเวลาในการพยากรณ์



2. นำข้อมูลที่จะใช้เข้าไปวิเคราะห์ในแต่ละชุด คือ ข้อมูลชุด 50 วัน 100 วัน 200 วัน และ 500 วัน เข้าไปวิเคราะห์ในแบบจำลองโครงข่ายประสาทเทียมซึ่งสร้างโดยใช้โปรแกรม NeuroSolutions 5.0 เป็นจำนวนทั้งหมด 11 แบบจำลองซึ่งเป็น โครงสร้างแบบดั้งเดิมและแบบประยุกต์ของแบบโครงข่ายแบบป้อนไปข้างหน้าและโครงข่ายแบบป้อนกลับ โดยใช้ชั้นซ่อนเท่ากันที่ 1 ชั้นและจำนวนการเรียนรู้ที่ 1,000 รอบ(เป็นข้อจำกัดของโปรแกรม NeuroSolution ที่เป็นรุ่นทดลองใช้)ส่วนวิธีการมาใช้โปรแกรม NCSS2007 ในการสร้างแบบจำลอง โดยแบบจำลองประกอบด้วย

โครงข่ายประสาทเทียมแบบป้อนไปข้างหน้า

- แบบ Multilayer Perceptions Networks (MLPs)
- แบบ Generalized Feed Forward Networks
- แบบ Modular Feed Forward Networks
- แบบ Jordan and Elman Networks
- แบบ Principal Component Analysis Networks (PCAs)
- แบบ Radial Basis Function (RBF) Networks
- แบบ Self-Organizing Feature Maps (SOFMs)

โครงข่ายประสาทเทียมแบบป้อนกลับ

- แบบ Time Lagged Recurrent Networks (TLRNs)
- แบบ Fully Recurrent Networks
- แบบ Co-Active Neuro-Fuzzy Inference System (CANFIS)
- แบบ Support Vector Machine (SVM)

การวิเคราะห์ด้วยวิธีอาร์มา

3. เปรียบเทียบข้อมูลที่ได้ที่จากการวิเคราะห์จากแบบจำลองโครงข่ายประสาทเทียมในแบบต่างๆ และวิเคราะห์ด้วยวิธีอาร์มา มาเปรียบเทียบกับข้อมูลในส่วนที่ 2 ที่เราได้ทำการแบ่งไว้แล้วจำนวน 50 วัน โดยใช้ค่าสถิติดังต่อไปนี้ในการเปรียบเทียบคือ

- ค่าความคลาดเคลื่อนที่แท้จริงเฉลี่ย (Mean Absolute Deviation หรือ MAD) ใช้วัดความแม่นยำของค่าพยากรณ์ โดยการเฉลี่ยค่าความผิดพลาดของค่าพยากรณ์ ซึ่งจะมีประโยชน์มากสำหรับการวิเคราะห์ที่ต้องการวัดความผิดพลาดในการพยากรณ์ที่มีหน่วยเดียวกันกับข้อมูลอนุกรมเวลาเดิม มีรูปแบบสมการในการพิจารณาดังนี้

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|; \text{เมื่อ } n \text{ เป็นจำนวนคาบเวลา } t \text{ ทั้งหมด}$$

- ค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean Square Error หรือ MSE) ใช้สำหรับหาค่าความแตกต่างระหว่างค่าพยากรณ์และค่าจริงเป็นวิธีการวัดความถูกต้องแม่นยำในการพยากรณ์และกำจัดทิศทางของการพยากรณ์ (ไม่คิดทิศทางบวกหรือลบ ซึ่งค่าบวกกับค่าลบจะหักล้างกันทำให้ค่า MAD น้อยแต่ไม่ได้หมายความว่ามีความแม่นยำมากเสมอไป) ซึ่งสามารถคำนวณได้จากการนำ MSE มาหารากที่สอง ซึ่งมีรูปแบบสมการในการพิจารณาดังนี้

$$MSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}; \text{เมื่อ } n \text{ เป็นจำนวนคาบเวลา } t \text{ ทั้งหมด}$$

- ร้อยละของค่าเฉลี่ยความผิดพลาด (Mean Absolute Percentage Error หรือ MAPE) เพื่อใช้สำหรับการเปรียบเทียบความแม่นยำสำหรับชุดข้อมูลที่อยู่ในช่วงแตกต่างกัน โดยปรับให้ค่าเฉลี่ยความผิดพลาดอยู่ในรูปของร้อยละเหมือนกัน ซึ่งมีรูปแบบสมการในการพิจารณาดังนี้

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|; \text{เมื่อ } n \text{ เป็นจำนวนคาบเวลา } t \text{ ทั้งหมด}$$

ระยะเวลาในการดำเนินงาน

การศึกษาครั้งนี้ใช้ระยะเวลา 5 เดือน ตั้งแต่เดือน ตุลาคม 2551 – กุมภาพันธ์ 2552 โดยการเก็บรวบรวมข้อมูลดัชนี SET และ SET50 และราคาหุ้นในกลุ่ม SET50 จากนั้นเข้าไปวิเคราะห์ในแบบจำลองโครงข่ายประสาทเทียมซึ่งสร้างโดยใช้โปรแกรม NeuroSolutions 5.0 และวิธีอาร์มา