

การฝึกโมเดล ARIMA: คำอธิบายฉบับเข้าใจง่ายแต่ถูกต้องตามหลักวิชาการ

ARIMA (Autoregressive Integrated Moving Average) เป็นโมเดลยอดนิยมสำหรับการวิเคราะห์และพยากรณ์ข้อมูลอนุกรมเวลา ซึ่งประยุกต์ใช้ได้หลายวงการ ทั้งการเงิน เศรษฐศาสตร์ การพยากรณ์ยอดขาย และอื่นๆ

องค์ประกอบของ ARIMA

ARIMA ประกอบด้วยองค์ประกอบ 3 ส่วนสำคัญ ได้แก่:

1. **AR (Autoregressive):** การใช้ค่าข้อมูลในอดีตมาพยากรณ์ค่าในอนาคต^{[1] [2]}
2. **I (Integrated):** การแปลงข้อมูลให้อยู่ในรูปแบบที่มีคุณสมบัติหยุดนิ่ง (stationary) โดยการหาผลต่าง (differencing)^{[3] [1]}
3. **MA (Moving Average):** การใช้ค่าความคลาดเคลื่อนในการพยากรณ์ในอดีตเพื่อปรับปรุงการพยากรณ์ในอนาคต^{[1] [2]}

โมเดล ARIMA จะแสดงในรูปแบบ ARIMA(p, d, q) โดย:

- **p:** จำนวนค่าในอดีตที่นำมาใช้ (AR order)
- **d:** จำนวนครั้งที่ต้องทำ differencing เพื่อให้ข้อมูลเป็น stationary
- **q:** จำนวนค่าความคลาดเคลื่อนในอดีตที่นำมาใช้ (MA order)^{[3] [2]}

ขั้นตอนการฝึกโมเดล ARIMA

1. การตรวจสอบและเตรียมข้อมูล

เริ่มต้นด้วยการพล็อตข้อมูลอนุกรมเวลาเพื่อตรวจสอบว่ามีแนวโน้ม (trend) ฤดูกาล (seasonality) หรือวัฏจักร (cycle) หรือไม่^[1] การทำความเข้าใจลักษณะข้อมูลเบื้องต้นจะช่วยให้การเลือกพารามิเตอร์โมเดลได้อย่างเหมาะสม

2. การแบ่งข้อมูล

แบ่งข้อมูลออกเป็นชุดฝึกสอน (training set) และชุดทดสอบ (test set) โดยทั่วไปจะใช้สัดส่วน 70-80% สำหรับชุดฝึกสอน และ 20-30% สำหรับชุดทดสอบ^{[4] [5]} การแบ่งข้อมูลนี้จะช่วยให้เราสามารถประเมินประสิทธิภาพของโมเดลได้อย่างเหมาะสม

3. การตรวจสอบคุณสมบัติ Stationarity

ข้อมูลที่จะใช้กับโมเดล ARIMA ต้องมีคุณสมบัติ stationary คือมีค่าเฉลี่ยและความแปรปรวนคงที่ตลอดช่วงเวลา ตรวจสอบโดยใช้ Augmented Dickey-Fuller (ADF) test^{[1] [2]} หากข้อมูลไม่เป็น stationary ต้องทำ differencing เพื่อปรับให้ข้อมูลมีคุณสมบัติ stationary ก่อน จำนวนครั้งที่ต้องทำ differencing คือค่า d ใน ARIMA(p,d,q)^{[3] [6]}

4. การกำหนดค่า p และ q

มีวิธีกำหนดค่า p และ q ได้หลายวิธี:

1. **การวิเคราะห์ ACF และ PACF:** ดูกราฟ Autocorrelation Function (ACF) และ Partial Autocorrelation Function (PACF) เพื่อกำหนดค่า p และ q^[6]
 - ถ้า ACF ลดลงอย่างช้าๆ และ PACF ตัดหลังจาก lag p → AR(p)
 - ถ้า ACF ตัดหลังจาก lag q และ PACF ลดลงอย่างช้าๆ → MA(q)
 - ถ้าทั้ง ACF และ PACF ลดลงอย่างช้าๆ → ARMA(p,q)
2. **การเปรียบเทียบค่า AIC:** สร้างโมเดลหลายๆ แบบโดยใช้ค่า p และ q ที่แตกต่างกัน แล้วเลือกโมเดลที่มีค่า AIC (Akaike Information Criterion) ต่ำที่สุด^{[3] [11]} AIC เป็นค่าที่ช่วยสร้างสมดุลระหว่างความพอดีและความเรียบง่ายของโมเดล^[11]

5. การประมาณค่าพารามิเตอร์

เมื่อได้ค่า p, d, q แล้ว นำไปสร้างแบบจำลอง ARIMA(p,d,q) ซึ่งค่าพารามิเตอร์จะถูกประมาณโดยวิธีทางสถิติ เช่น Maximum Likelihood Estimation^{[3] [14]}

6. การตรวจสอบความเหมาะสมของโมเดล

ตรวจสอบว่าค่าความคลาดเคลื่อน (residuals) ของโมเดลเป็น white noise หรือไม่ โดย residuals ควรมีการกระจายแบบปกติ ไม่มีสหสัมพันธ์ในตัวเอง และมีค่าเฉลี่ยเท่ากับศูนย์^[6] ถ้า residuals ไม่เป็น white noise แสดงว่าโมเดลยังไม่เหมาะสม ต้องกลับไปปรับค่า p, d, q ใหม่

7. การประเมินประสิทธิภาพโมเดล

ประเมินประสิทธิภาพของโมเดลโดยใช้ชุดข้อมูลทดสอบ ตัววัดที่นิยมใช้ได้แก่:

- **RMSE (Root Mean Square Error):** รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง^{[4] [5]}
- **MAPE (Mean Absolute Percentage Error):** ค่าเฉลี่ยของเปอร์เซ็นต์ความผิดพลาดสัมบูรณ์^[3]
- **MAE (Mean Absolute Error):** ค่าเฉลี่ยของความผิดพลาดสัมบูรณ์

ค่ายิ่งต่ำยิ่งแสดงถึงความแม่นยำที่ดีของโมเดล

8. การพยากรณ์

เมื่อได้โมเดลที่เหมาะสมแล้ว สามารถนำไปใช้พยากรณ์ข้อมูลในอนาคตได้^{[3] [6]}

เทคนิคเพิ่มเติม

1. **Auto ARIMA**: เป็นอัลกอริทึมที่ช่วยค้นหาค่า p , d , q ที่เหมาะสมโดยอัตโนมัติ โดยทดลองหลายๆ ค่า และเลือกค่าที่ให้ AIC หรือ BIC ต่ำที่สุด^[7]
2. **SARIMA (Seasonal ARIMA)**: เป็นการขยาย ARIMA เพื่อรองรับข้อมูลที่มีลักษณะเป็นฤดูกาล โดยเพิ่มพารามิเตอร์อีก 3 ตัว (P,D,Q) เพื่อจัดการกับความเป็นฤดูกาล^[3]
3. **ARIMA-ANN แบบผสม**: เป็นการผสมผสานระหว่าง ARIMA และ Artificial Neural Network เพื่อเพิ่มความแม่นยำในการพยากรณ์ โดย ARIMA จะจัดการกับส่วนที่เป็นเชิงเส้น ขณะที่ ANN จะจัดการกับส่วนที่ไม่เป็นเชิงเส้น^{[4] [5]}

สรุป

การฝึกโมเดล ARIMA เป็นกระบวนการที่มีขั้นตอนชัดเจน เริ่มจากการเตรียมข้อมูล การตรวจสอบคุณสมบัติ stationarity การกำหนดค่า p , d , q การประมาณค่าพารามิเตอร์ การตรวจสอบความเหมาะสมของโมเดล และการประเมินประสิทธิภาพ โดยแต่ละขั้นตอนมีความสำคัญและส่งผลต่อความแม่นยำของการพยากรณ์ทั้งสิ้น เมื่อเข้าใจหลักการพื้นฐานแล้ว การฝึกโมเดล ARIMA จึงไม่ใช่เรื่องยากอีกต่อไป

✱✱

1. <https://www.quantseras.com/post/arima-model-ที่ออกแบบจำลองในการวิเคราะห์หอนุกรมเวลา>
2. <https://algoaddict.wordpress.com/2019/06/22/หลักการทํางานของ-arima-ในการท/>
3. <https://grad.dpu.ac.th/upload/content/files/year10-2/10-10.pdf>
4. <https://li01.tci-thaijo.org/index.php/tstj/article/download/210424/170099>
5. <https://sure.su.ac.th/xmlui/bitstream/handle/123456789/11717/Fulltext.pdf?sequence=1&isAllowed=y>
6. <https://research.kpru.ac.th/research2/pages/filere/1556607621.pdf>
7. <https://www.datacamp.com/tutorial/arima>