บทที่ 5

การตรวจจับข้อมูลที่ผิดพลาด

วิธีการประมาณค่าตัวแปรสถานะรวมถึงการเลือกตำแหน่งติดตั้งเครื่องมือวัดที่ได้นำเสนอนั้น เป็นความพยายามที่จะประมาณค่าตัวแปรสถานะโดยอาศัยข้อมูลที่มีอยู่อย่างจำกัดให้เกิดประโยชน์สูงที่สุด อย่างไรก็ดีการประมาณค่าตัวแปรสถานะนั้น ยังมีอีกปัญหาหนึ่งซึ่งมีความสำคัญมากนั่นคือ การตรวจจับข้อมูลที่ผิดพลาดจากเครื่องมือวัด

การตรวจจับข้อมูลที่ผิดพลาด (Bad Data Detection) มักใช้ในระบบที่มีข้อมูลอยู่เป็นจำนวนมาก โดยปกติจะมีจำนวนข้อมูลมากกว่าจำนวนตัวแปรสถานะที่เราจะทำการประมาณ ($N_m > N_s$) แม้ว่าการมีจำนวนข้อมูลเป็นจำนวนมากนั้นจะทำให้การประมาณค่านั้นมีค่าใกล้เคียงกับค่าจริงมาก แต่หากข้อมูลที่ได้รับมานั้นมีความผิดพลาด หรือ เป็นข้อมูลที่คลาดเคลื่อนไปจากความเป็นจริงอันเนื่องมาจากสาเหตุต่างๆ เช่น เครื่องมือวัดกระแสไม่ทำงาน หรือการติดตั้งเครื่องมือวัดที่ผิดพลาด ข้อมูลที่มาจากความผิดพลาดเหล่านี้ หากนำมาคำนวณเพื่อประมาณตัวแปรสถานะย่อมทำให้เกิดความผิดพลาดจากการประมาณค่าได้ หนทางในการแก้ไขดังกล่าวเราจำเป็นต้องใช้วิธีการตรวจจับข้อมูลที่ผิดพลาดเข้ามาแก้ไขปัญหาดังกล่าว โดยในหัวข้อนี้จะได้อธิบายถึงการตรวจจับข้อมูลที่ผิดพลาดเบื้องต้น และจะได้ทำการทดสอบการประยุกต์ใช้วิธีการประมาณค่าที่นำเสนอกับการตรวจจับข้อมูลที่ผิดพลาดในการทดสอบขั้นต่อไป

ในการตรวจจับข้อมูลที่ผิดพลาดนั้นโดยปกติแล้วจะใช้ค่า $J(\mathbf{x})$ ที่ทำการคำนวณจากตัวแปรสถานะภายหลังจากการประมาณค่ามาเป็นตัวชี้วัด ค่าของ $J(\mathbf{x})$ นั้นควรจะมีค่าต่ำๆภายหลังจากการประมาณค่า ซึ่งนั่นหมายถึงค่าฟังก์ชั่นต่างๆที่เราคำนวณได้จากตัวแปรสถานะที่ทำการประมาณนั้นมีค่าใกล้เคียงกับข้อมูลที่เราได้รับจากทั้งเครื่องมือวัด และโครงสร้างของระบบ ในทางตรงกันข้ามหากค่า $J(\mathbf{x})$ นั้นมีค่าสูงจะบ่งชี้ให้เห็นว่าข้อมูลดังกล่าวมีแนวโน้มที่จะมีความผิดพลาดเนื่องจากการวัด

ดังนั้นเราจำเป็นต้องมีการพิจารณาค่าสูงสุดของ $J(\mathbf{x})$ ที่เรายังยอมรับได้และจะถือว่าไม่มีการนำข้อมูลที่ผิดพลาดมาใช้ในการประมาณค่าตัวแปรสถานะ โดยแนวทางหนึ่งในการทดสอบค่า $J(\mathbf{x})$ นั่นคือ การทดสอบสมมุติฐาน (Hypotheis testing) [1] โดยเราจะสมมติให้ค่าความคลาดเคลื่อนนั้นมีการกระจายตัวแบบปกติ (Normal distribution) ผลคือจะทำให้การกระจายตัวของ $J(\mathbf{x})$ นั้นเป็นแบบไคแสควร์ (Chi-Squared distribution; $\chi^2(K)$) โดยค่า K คือองศาเสรี (degree of freedom;DOF) ซึ่งสามารถคำนวณได้จาก

$$K = N_m - N_s \tag{5.1}$$

โดยที่ N_m คือ จำนวนข้อมูลจากเครื่องมือวัดและโครงสร้างของระบบ

N_s คือ จำนวนตัวแปรสถานะที่เราจะทำการประมาณค่า

โดยในวิทยานิพนธ์นี้เรากำหนดให้ทราบข้อมูลขนาดของแรงดันที่บัสที่มีเครื่องกำเนิดไฟฟ้าต่ออยู่ และกำหนดให้มุมของแรงดันที่บัสอ้างอิงมีค่าเป็นศูนย์ ดังนั้นจำนวนตัวแปรสถานะที่เราจะทำการประมาณค่าสามารถคำนวณได้ดังนี้

$$N_s = (2 \times N_{bus}) - N_g - 1 \tag{5.2}$$

โดยที่ N_{bus} คือ จำนวนบัสทั้งหมดในระบบ

N_g คือ จำนวนบัสที่มีเครื่องกำเนิดไฟฟ้าต่ออยู่ในระบบ

และหากค่าเฉลี่ยของ $J(\mathbf{x})$ มีค่าเท่ากับ K แล้วเราจะสามารถพิสูจน์ได้ว่าค่าเบี่ยงเบนมาตรฐาน (Standard deviation; $\sigma_{J(\mathbf{x})}$) จะมีค่าดัง (5.3)

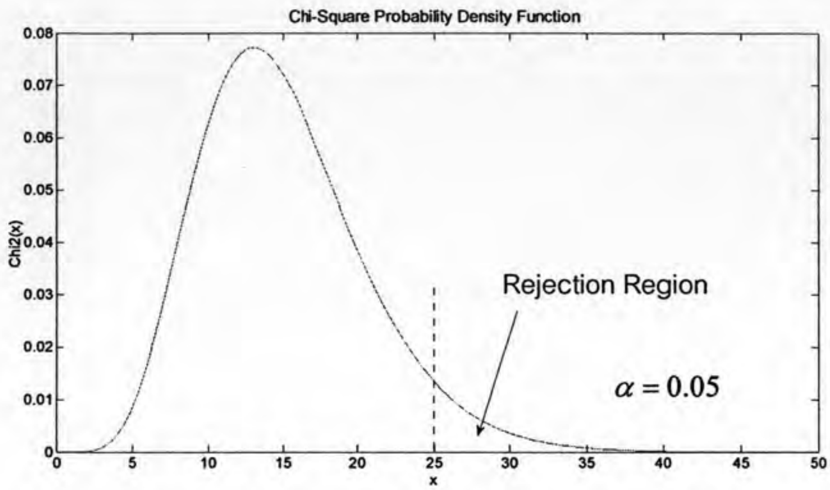
$$\sigma_{J(\mathbf{x})} = \sqrt{2K} \tag{5.3}$$

โดยในการทดสอบไคสแควร์นั้นเราจะทำการกำหนดค่าสูงสุดของ $J(\mathbf{x})$ โดยจะให้สัญลักษณ์เป็น $\chi_{K,\alpha}$ โดยเราจะถือว่าค่า $J(\mathbf{x})$ นั้นมีข้อมูลที่ผิดพลาดร่วมอยู่ด้วยเมื่อค่า $J(\mathbf{x})$ มีค่ามากกว่า $\chi_{K,\alpha}$ ($J(\mathbf{x}) > \chi_{K,\alpha}$) นั่นคือ

$$prob(J(x) > \chi_{K,\alpha} \mid J(x) \text{ is chi-square with K degree of freedom}) = \alpha \tag{5.4}$$

จาก (5.4) หมายถึงโอกาสที่ $J(\mathbf{x})$ จะมีค่ามากกว่า $\chi_{K,\alpha}$ นั้นมีค่าเท่ากับ α โดยที่กระจายตัวของความน่าจะเป็นของ $J(\mathbf{x})$ นั้นสามารถอธิบายได้ด้วยการกระจายแบบไคสแควร์ด้วยองศาเสรีเท่ากับ K

ค่า α เป็นค่าระดับนัยสำคัญของการทดสอบ (Significance level) โดยการเลือกระดับนัยสำคัญที่ค่าเท่ากับ α นั้น จะทำให้เราได้ค่าสูงสุดของการทดสอบค่า $J(\mathbf{x})$ ได้โดยการหาค่าฟังก์ชั่นอินเวอร์สของไคสแควร์ที่ระดับขั้นความเสรีเป็น K และมีค่าระดับนัยสำคัญเท่ากับ α



รูปที่ 5.1 การแจกแจงความน่าจะเป็นแบบไคสแควร์เมื่อมีขั้นความเสรีเท่ากับ 15

รูปที่ 5.1 เป็นตัวอย่างการแจกแจงความน่าจะเป็นแบบไคสแควร์เมื่อมีองศาเสรีเท่ากับ 15 โดยจะเห็นได้ว่าเมื่อเรากำหนดค่าระดับนัยสำคัญของการทดสอบเท่ากับ 0.05 จะพบว่าเราจะได้ค่ามากที่สุดของ $J(\mathbf{x})$ ที่จะยังยอมรับว่าไม่มีข้อมูลผิดพลาดอยู่ได้มีค่าประมาณ 25 หาก $J(\mathbf{x})$ ที่คำนวณจากตัวแปรสถานะที่ได้จากการประมาณค่านั้นมีค่ามากกว่านี้เราจะถือว่ามีข้อมูลผิดพลาดอยู่ และจะทำการตรวจสอบหาข้อมูลที่ผิดพลาดต่อไป

ในกรณีที่ทำการทดสอบไคสแควร์แล้วพบว่ามีข้อมูลที่ผิดพลาดรวมอยู่ด้วย จะต้องมีการตรวจสอบข้อมูลแต่ละข้อมูลเพื่อที่จะทราบว่าข้อมูลใดมีแนวโน้มที่จะเป็นข้อมูลที่ผิดพลาดมากที่สุด โดยปกติแล้วการตรวจสอบข้อมูลแต่ละข้อมูลนั้นนิยมใช้ค่าความคลาดเคลื่อนปกติ (Normalize residual; r_i^N) เพื่อมาตรวจสอบความผิดพลาด และเนื่องจากเราได้สมมติให้การกระจายตัวของข้อมูลเป็นแบบปกติ ดังนั้นเราจึงสามารถนำค่าความคลาดเคลื่อนปกติมาเปรียบเทียบกับการกระจายแบบ Z ซึ่งเป็นการกระจายแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และมีค่าเบี่ยงเบนมาตรฐานเท่ากับ 1 ได้ โดยในการตรวจสอบจะต้องกำหนดนัยสำคัญของการตรวจสอบเพื่อหาค่าความคลาดเคลื่อนปกติสูงสุดที่เราจะยอมรับได้ หากมีค่าเกินค่าสูงสุดนี้จะหมายความว่าข้อมูลดังกล่าวมีแนวโน้มที่จะเป็นข้อมูลที่ผิดพลาดได้ ค่าความคลาดเคลื่อนปกตินั้นสามารถคำนวณได้ดัง (5.5)

$$r_i^N = \frac{(z_i - f_i(\mathbf{x}))}{\sigma_i} \tag{5.5}$$

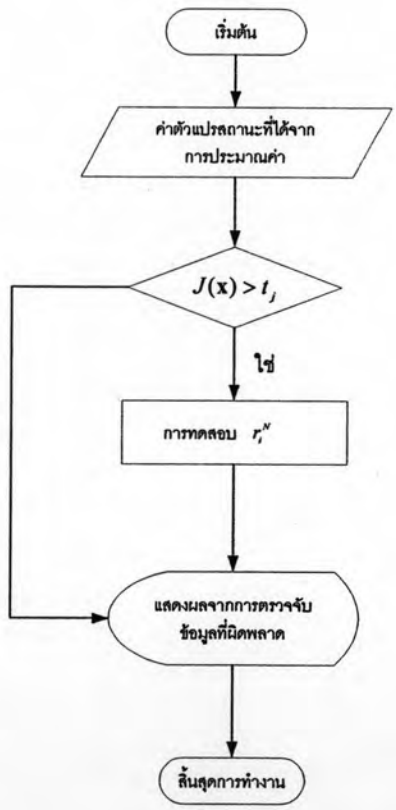
โดยที่ r_i^N คือ ค่าความคลาดเคลื่อนแบบปกติมาตรฐานของข้อมูลที่ i

σ_i คือ ส่วนเบี่ยงเบนมาตรฐานของข้อมูลที่ i ซึ่งประมาณให้เท่ากับค่าความแม่นยำของเครื่องมือวัด

z_i คือ ข้อมูลที่ i ที่ได้จากเครื่องมือวัดที่ติดตั้ง

$f_i(\mathbf{x})$ คือ ค่าฟังก์ชั่นที่คำนวณได้จากตัวแปรสถานะที่มาจากการประมาณค่า

หากข้อมูลใดที่ได้ทำการตรวจสอบแล้วพบว่าเป็นข้อมูลที่มีแนวโน้มมากที่สุดที่จะเป็นข้อมูลที่ผิดพลาด ข้อมูลดังกล่าวจะถูกกำจัดออกไปจากชุดข้อมูลที่ใช้ในการประมาณค่าตัวแปรสถานะ ขั้นตอนการตรวจจับข้อมูลที่ผิดพลาดเป็นดังรูปที่ 5.2



รูปที่ 5.2 ขั้นตอนการตรวจจับข้อมูลที่ผิดพลาด

อย่างไรก็ดีในการทดสอบ r_i^N ในกรณีที่มีข้อมูลผิดพลาด ข้อมูลที่มีค่าความคลาดเคลื่อนแบบปกติสูงสุดเป็นเพียงข้อมูลที่มีแนวโน้มที่จะเป็นข้อมูลที่ผิดพลาด (Suspected data) เท่านั้น เราจำเป็นต้องทำการทดสอบโดยการนำออกจากการคำนวณค่า $J(\mathbf{x})$ และทำการประมาณค่าตัวแปรสถานะใหม่ หากค่า $J(\mathbf{x})$ ที่ได้จากค่าตัวแปรสถานะจากการประมาณครั้งใหม่มีค่าน้อยกว่าค่าเดิมก็จะถือว่าข้อมูลนั้นเป็นข้อมูลที่ผิดพลาด หากไม่ใช่เราจะทำการพิจารณาข้อมูลที่มีค่า r_i^N สูงรองลงมาและทำการทดสอบอีกครั้งจนได้ค่า $J(\mathbf{x})$ ที่ผ่านการทดสอบไคสแควร์