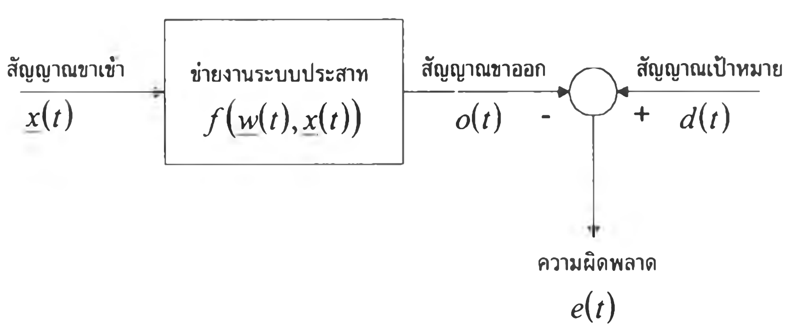
บทที่ 1
บทนำ

ความเบื้องต้น

ข่ายงานระบบประสาท ได้ถูกนำมาประยุกต์ใช้งานทางวิศวกรรมหลายๆ ด้าน เช่น ในด้านวิศวกรรมสื่อสาร ได้นำมาใช้กับการรู้จำตัวอักษร (Character Recognition) หรือคำพูดของมนุษย์ ในด้านวิศวกรรมระบบควบคุม ก็นำมาใช้ออกแบบเพื่อให้แสดงสัญญาณอย่างที่ต้องการ [11] เป็นต้น

เมื่อพิจารณาถึงการใช้ข่ายงานระบบประสาท แสดงดังรูปที่ 1.1 สัญญาณขาเข้าของระบบสามารถแสดงด้วยลำดับของเวกเตอร์ $\underline{x}(0), \underline{x}(1), \underline{x}(2), \ldots$ และเวกเตอร์ $\underline{w}(t)$ ซึ่งประกอบด้วยพารามิเตอร์ของระบบที่เวลา t โดยที่เวกเตอร์ทั้งสองมีขนาดเท่ากับ N สัญญาณขาออกของระบบ $o(t)$ ที่เวลา t จะขึ้นกับสัญญาณขาเข้าและพารามิเตอร์ที่เวลา t

$$o(t) = f\left(\underline{w}(t), \underline{x}(t)\right), \quad t = 1, 2, 3, \ldots \tag{1.1}$$



รูปที่ 1.1 บล็อกไดอะแกรมแสดงการใช้ข่ายงานระบบประสาท

สัญญาณขาออก $o(t)$ จะลู่เข้าหาสัญญาณเป้าหมาย $d(t)$ ซึ่งเป็นค่าอ้างอิง การวัดความผิดพลาดระหว่างสัญญาณเป้าหมาย $d(t)$ และสัญญาณขาออก $o(t)$ แสดงดังสมการ (1.2)

$$\begin{aligned} e(t) &= d(t) - o(t) \\ &= d(t) - f\left(\underline{w}(t), \underline{x}(t)\right) \end{aligned} \tag{1.2}$$

จากสมการ (1.2) เห็นได้ว่า เมื่อกำหนดสัญญาณขาเข้า $\underline{x}(t)$ และสัญญาณเป้าหมาย $d(t)$ ความผิดพลาด $e(t)$ และสัญญาณขาออก $o(t)$ จะเป็นฟังก์ชันของพารามิเตอร์ $\underline{w}(t)$

การปรับพารามิเตอร์ของข่ายงานระบบประสาทมีจุดมุ่งหมายเพื่อหาจุดเหมาะสมที่สุด ที่ทำให้ฟังก์ชันเป้าหมาย

$$\varepsilon(t) = E\left[\left(d(t) - f\left(\underline{w}(t), \underline{x}(t)\right)\right)^2\right] \tag{1.4}$$

มีค่าต่ำสุด จากสมการ (1.4) ฟังก์ชันเป้าหมาย $\varepsilon(t)$ เป็นความผิดพลาดกำลังสองเฉลี่ย โดยใช้สัญญลักษณ์ E เป็นเครื่องหมายการหาค่าคาดหวัง (expectation value) การปรับพารามิเตอร์ $\underline{w}(t)$ เพื่อให้ลู่เข้าสู่จุดเหมาะสมที่สุด ทำได้ดังสมการ

$$\underline{w}(t+1) = \underline{w}(t) - \frac{\beta}{2}\nabla_{\underline{w}}\varepsilon(t) \tag{1.5}$$

ซึ่งเป็นไปตามขั้นตอนการลดระดับตามแนวเกรเดียนต์ (Gradient Descent) โดยที่ $\nabla_{\underline{w}}\varepsilon(t)$ เป็นเกรเดียนต์ของฟังก์ชัน $\varepsilon(t)$ เทียบกับเวกเตอร์พารามิเตอร์ $\underline{w}(t)$ และใส่ $\frac{1}{2}$ คูณไว้เพื่อไม่ให้เทอมนี้มีเลข 2 คูณอยู่หลังจากหาเกรเดียนต์แล้ว ค่า β เป็นอัตราการเรียนรู้สำหรับควบคุมลักษณะการลู่เข้า ซึ่งปกติจะกำหนดเป็นค่าคงตัวที่มีค่าเป็นบวกน้อย ๆ

กรณีที่ใช้งานแบบเชื่อมตรง (on-line) กับการประยุกต์ใช้งานหลายอย่าง เราสามารถวัดประสิทธิภาพของการลู่เข้าได้ 2 อย่าง คือ ความเร็วในการลู่เข้าและความถูกต้องของการประมาณค่าจุดเหมาะสมที่สุด การใช้อัตราการเรียนรู้ β เป็นค่าคงตัว จะไม่สามารถให้ประสิทธิภาพทั้ง 2 อย่างนี้พร้อมกันได้ การประยุกต์ใช้งานส่วนใหญ่ ถ้ากำหนด β ให้มีค่ามากจะสามารถลู่เข้าได้เร็ว แต่ประมาณค่าจุดเหมาะสมที่สุดได้ไม่ดี ในทางกลับกัน ถ้ากำหนดให้มีค่าน้อยจะสามารถประมาณค่าได้ใกล้เคียงกว่าเมื่อเข้าสู่สถานะอยู่ตัว แต่ทำให้ลู่เข้าได้ช้า

การปรับพารามิเตอร์ตามสมการ (1.5) โดยใช้ β เป็นค่าคงตัว ทำให้เกิดปัญหาเกี่ยวกับประสิทธิภาพของการลู่เข้า โดยไม่สามารถให้ความเร็วและการประมาณค่าจุดเหมาะสมที่สุดพร้อมกันได้ ดังนั้น เพื่อหลีกเลี่ยงปัญหาที่เกิดขึ้น จึงมีผู้นำเสนอให้ปรับอัตราการเรียนรู้ β ให้มีค่าเปลี่ยนตามเวลา เช่น การใช้เทอมอันดับหนึ่งของอนุกรมเทเลอร์ (Taylor's Series) เพื่อประมาณค่า $\Delta o(t)$ ซึ่งเป็นการเปลี่ยนแปลงของสัญญาณขาออก $o(t)$ เพื่อทำให้ฟังก์ชัน $\varepsilon(t)$ ที่อยู่ในรูปของ β มีลักษณะที่ง่าย และสามารถหาค่าที่ทำให้ฟังก์ชัน $\varepsilon(t)$ ต่ำสุดในแต่ละรอบการคำนวณได้ [5] วิธีการนี้ต้องใช้การคำนวณมากเป็นระดับกำลังสองของจำนวนพารามิเตอร์ และการประมาณมีค่าใกล้เคียงเมื่อเทอมอันดับสูงในอนุกรมเทเลอร์มีค่าน้อยเท่านั้น

Franzini [14] ได้เสนอให้ปรับอัตราการเรียนรู้โดยอาศัยโคซายน์ของทิศทาง (direction cosine) ซึ่งเป็นผลคูณภายใน (inner product) ของ $\Delta\underline{w}_n(t)$ และ $\Delta\underline{w}_n(t-1)$ ซึ่งเป็นเวกเตอร์ปกติของการเปลี่ยนแปลงพารามิเตอร์ของรอบการคำนวณที่อยู่ติดกัน และถูกปรับปรุงโดย Hsin, Li, Sun และ Sclabassi [7] โดยให้อัตราการเรียนรู้ β เป็นผลรวมถ่วงน้ำหนักของโคซายน์ของทิศทางในรอบการคำนวณติดกันในอดีตจำนวนหนึ่งและให้ผลรวมของค่าน้ำหนักเท่ากับ 1 แต่เนื่องจากโคซายน์แสดงทิศทางมีค่าอยู่ในช่วง -1 ถึง 1 ทำให้อัตราการเรียนรู้ที่ได้จากวิธีทั้งสองมีค่าเป็นลบได้ ซึ่งทำให้ฟังก์ชันเป้าหมาย $\varepsilon(t)$ มีค่าเพิ่มขึ้น และไม่ลู่เข้าได้ การใช้งานวิธีทั้งสองจึงจำเป็นต้องมีการปรับปรุงเพื่อให้อัตราการเรียนรู้เป็นบวกเสมอ

Nachtcheim [6] เสนอการปรับอัตราการเรียนรู้ โดยใช้การประมาณฟังก์ชันเป้าหมายด้วยอนุกรมเทเลอร์ถึงเทอมอันดับหนึ่ง ดังนี้

$$\varepsilon\left[\underline{w}(t)+\Delta\underline{w}(t)\right]=\varepsilon\left[\underline{w}(t)\right]+\Delta\underline{w}(t)^T\frac{\partial\varepsilon(t)}{\partial\underline{w}(t)}+\ldots \tag{1.6}$$

โดยที่ การเปลี่ยนแปลงของพารามิเตอร์ $\Delta\underline{w}(t)$ ตามวิธีการลดระดับตามแนวเกรเดียนต์ เป็น

$$\Delta\underline{w}(t)=-\beta(t)\frac{\partial\varepsilon(t)}{\partial\underline{w}(t)} \tag{1.7}$$

และกำหนดให้การเปลี่ยนแปลงของพารามิเตอร์ $\Delta\underline{w}(t)$ มีผลให้ฟังก์ชันเป้าหมายที่ปรับพารามิเตอร์แล้ว $\varepsilon\left[\underline{w}(t)+\Delta\underline{w}(t)\right]$ ในสมการ (1.6) มีค่าเป็น 0 จะได้ค่าอัตราการเรียนรู้ β ในแต่ละรอบการเรียนรู้ เป็นไปตามสมการ

$$\beta(t)=\frac{\varepsilon(t)}{\sum_{i=1}^{N}\left[\partial\varepsilon(t)/\partial w_i(t)\right]^2} \tag{1.8}$$

แต่เนื่องจาก วิธีนี้ไม่ได้จำกัดขนาดอัตราการเรียนรู้ไว้ ทำให้อัตราการเรียนรู้ที่ได้มีขนาดใหญ่เกินไป ทำให้ไม่ลู่เข้าหรือเกิดการแกว่งได้ง่าย จึงจำเป็นต้องมีการปรับปรุงโดยจำกัดขนาดสูงสุดไว้หรือลดขนาดอัตราการเรียนรู้ในแต่ละรอบลง Nachtsheim ได้เสนอให้ใช้ค่าคงตัวระหว่าง 0 ถึง 1 คูณจนกว่าค่าฟังก์ชันเป้าหมายลดลงในแต่ละรอบการเรียนรู้ ซึ่งทำให้การคำนวณและเวลาของแต่ละรอบไม่เท่ากัน จึงไม่สามารถกำหนดเวลาสำหรับการทำงานแต่ละรอบได้ วิธีนี้จึงไม่เหมาะสมกับการใช้งานในแบบเชื่อมตรง และไม่สามารถประยุกต์ใช้งานได้กว้างขวางเท่าที่ควร

Jacobs ได้พัฒนาวิธีปรับอัตราการเรียนรู้แบบ heuristic เรียกว่า Delta-bar-delta (Dbd) [2] ซึ่งแต่ละพารามิเตอร์จะถูกกำหนดให้มีอัตราการเรียนรู้ของตัวเอง แต่ละอัตราการเรียนรู้จะถูกปรับเพิ่มขึ้นหรือลดลงอย่างอิสระบนการเปลี่ยนแปลงของพารามิเตอร์ที่ใช้อัตราการเรียนรู้นั้น ถ้าเครื่องหมายของการเปลี่ยนแปลงพารามิเตอร์แสดงออกมาเหมือนกันในหลาย ๆ รอบการเรียนรู้ อัตราการเรียนรู้จะเพิ่มขึ้น แต่ถ้าต่างกัน อัตราการเรียนรู้จะลดลง

ต่อมา Sutton ได้นำเสนอการปรับอัตราการเรียนรู้ที่พัฒนาจากวิธี Dbd โดยปรับอัตราการเรียนรู้เพื่อให้ฟังก์ชันเป้าหมาย $\varepsilon(t)$ ลู่เข้าสู่จุดเหมาะสมที่สุด วิธีนี้เรียกว่า Incremental delta-bar-delta (Idbd) [3,4] Sutton ยังได้นำวิธีนี้ไปใช้งานร่วมกับตัวกรองคาลมาน (Kalman filter) เพื่อพัฒนาให้มีการคำนวณน้อยลง วิธีที่พัฒนาได้ใหม่นี้เรียกว่า K1 และ K2 แต่การปรับค่าที่ใช้ในตัวกรองคาลมานมีการคำนวณในระดับกำลังสองของจำนวนอัตราการเรียนรู้ ทำให้วิธีทั้งสองยังคงมีการคำนวณมากและไม่เหมาะกับการใช้งานแบบเชื่อมตรง

เนื่องจาก การปรับอัตราการเรียนรู้มีผู้เสนอไว้หลายวิธีและกระจายกันอยู่ ดังนั้น ในวิทยานิพนธ์ฉบับนี้ จึงได้รวบรวมเอาวิธีต่าง ๆ ที่มีการเสนอไว้ แล้วคัดเลือกเฉพาะวิธีที่เหมาะสมกับการใช้งานแบบเชื่อมตรงซึ่งประยุกต์ใช้งานได้กว้างขวาง มาทดสอบเปรียบเทียบกัน โดยใช้ฟังก์ชันเชิงเส้นที่มีและไม่มีการเปลี่ยนแปลงตามเวลา เพื่อทดสอบเปรียบเทียบประสิทธิภาพที่มีต่อการใช้งานเบื้องต้น และใช้การชดเชยช่องสัญญาณ (Channel Equalization) เป็นตัวอย่างสำหรับทดสอบกับการประยุกต์ใช้งานจริง โดยมีวัตถุประสงค์เพื่อเป็นแนวทางสำหรับปรับปรุงหรือเพิ่มประสิทธิภาพให้กับวิธีการปรับอัตราการเรียนรู้ให้เหมาะสมกับการประยุกต์ใช้งานเฉพาะด้าน หรือค้นหาหลักเกณฑ์ในการคัดเลือกวิธีการปรับสำหรับการประยุกต์ใช้งานที่กำหนดให้ ผลการทดสอบเปรียบเทียบในวิทยานิพนธ์นี้ คาดว่าจะเป็นประโยชน์ต่อผู้ที่สนใจในการประยุกต์ใช้ข่ายงานระบบประสาท หรือระบบที่มีการปรับพารามิเตอร์ด้วยวิธีการลดระดับตามแนวเกรเดียนต์ โดยต้องการเพิ่มประสิทธิภาพให้ระบบสามารถใช้งานได้กว้างขวาง

โครงสร้างของวิทยานิพนธ์

จากความเบื้องต้น ทำให้เข้าใจถึงปัญหาและแนวทางในการแก้ปัญหาโดยใช้วิธีการปรับอัตราการเรียนรู้ตามเวลา ในบทต่อไป จะกล่าวถึงทฤษฎีการหาค่าเหมาะสมที่สุดของพารามิเตอร์ในโครงสร้างแบบชั้นเดียวที่เป็นเชิงเส้น เพื่อให้ทราบถึงสาเหตุ ข้อจำกัด และแนวทางสำหรับการปรับอัตราการเรียนรู้เพื่อเพิ่มประสิทธิภาพของระบบ โดยกล่าวถึงกรณีที่สัญญาณขาเข้าเป็นกระบวนการสโตคาสติกส์ (stochastic process) และใช้งานระบบในแบบเชื่อมตรง ส่วนในบทที่ 3 กล่าวถึงขั้นตอนวิธีการปรับอัตราการเรียนรู้ที่นำมาทดสอบในวิทยานิพนธ์เพื่อแสดงแนวทาง

และเหตุผลในการสร้างขั้นตอน วิธีที่นำมาทดสอบ ได้แก่ Dbd, Idbd, direction cosine ซึ่งมีการปรับปรุงเพื่อให้อัตราการเรียนรู้เป็นบวกเสมอ และเอนโทรปี (Entropy) วิธีที่เลือกมาทดสอบเป็นวิธีที่เหมาะสมกับการใช้งานแบบเชื่อมตรง

ในบทที่ 4 และ 5 ได้แสดงผลการทดสอบวิธีการปรับอัตราการเรียนรู้กับข่ายงานระบบประสาท เปรียบเทียบกับการใช้อัตราการเรียนรู้เป็นค่าคงตัว ในบทที่ 4 เป็นการทดสอบโดยใช้การประมาณฟังก์ชันเชิงเส้น ซึ่งสามารถประมาณให้แม่นยำได้โดยข่ายงานระบบประสาทแบบชั้นเดียว การทดสอบเริ่มจากการประมาณฟังก์ชันที่ไม่มีการเปลี่ยนแปลงตามเวลา แล้วจึงให้ระบบประมาณฟังก์ชันที่มีเทอมค่าคงตัวและค่าสัมประสิทธิ์เปลี่ยนแปลงตามเวลา ตามลำดับ ส่วนในบทที่ 5 เป็นการทดสอบข่ายงานระบบประสาทกับตัวอย่างการประยุกต์ใช้งานจริง คือ การชดเชยช่องสัญญาณ (Channel Equalization) โดยกำหนดให้สัญญาณเป็นแบบเชิงเลข เพื่อแสดงผลการทดสอบและเปรียบเทียบการปรับอัตราการเรียนรู้แบบต่าง ๆ บนตัวอย่างการใช้งานจริง

บทที่ 6 เป็นบทสรุปของวิทยานิพนธ์ โดยสรุปผลการทดสอบที่ผ่านมาและเสนอแนะแนวทางสำหรับการทำงานวิจัยในชั้นต่อไป