การวิเคราะห์และออกแบบออโตมาตาความน่าจะเป็นแบบเชื่อมตรง

นาย จิตรกร พูลโพธิ์ทอง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2548

ISBN 974-17-2529-9

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

ON-LINE PROBABILISTIC AUTOMATA ANALYSIS AND DESIGN

Mr Jittakorn Pullpothong

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Computer Engineering

Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University

Academic Year 2005

ISBN 974-53-2517 -1

หัวข้อวิทยานิพนธ์	การวิเคราะห์และออกแบบออโตมาตาความน่าจะเป็นแบบเชื่อมตรง
โดย	นาย จิตรกร พูลโพธิ์ทอง
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร. อรรถสิทธิ์ สุรฤกษ์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์

(ศาสตราจารย์ ดร. ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(รองศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา)

..... อาจารย์ที่ปรึกษา

(อาจารย์ ดร. อรรถสิทธิ์ สุรฤกษ์)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร อานนท์ รุ่งสว่าง)

..... กรรมการ

(อาจารย์. ดร อาทิตย์ ทองทักษ์)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จิตรกร พูลโพธิ์ทอง : การวิเคราะห์และออกแบบออโตมาตาความน่าจะเป็นแบบเชื่อมตรง. (ON-LINE PROBABILISTIC AUTOMATA ANALYSIS AND DESIGN) อ. ที่ปรึกษา: ดร. อรรถสิทธิ์ สุรฤกษ์ 50 หน้า. ISBN 974-53-2517-1.

ปัจจุบันแบบจำลองภาษาสโตแคสติกหรือภาษาเชิงน่าจะเป็นมีอยู่หลายชนิด ได้แก่ แบบจำลองมาร์คอฟแบบซ่อน แบบจำลองลูกโซ่มาร์คอฟ ออโตมาตาอิงความน่าจะเป็น และอื่นๆ งานวิจัยเกี่ยวกับแบบจำลองเหล่านี้ส่วนมากเน้นที่การสร้างแบบจำลองเพื่อการเรียนรู้เชิงสถิต นั่นคือไม่เน้นการเปลี่ยนแปลงภายหลังการเรียนรู้ เนื่องจากเทคโนโลยีทางคอมพิวเตอร์ในปัจจุบันทำให้เกิดข้อมูลรูปแบบใหม่เกิดขึ้น จำเป็นต้องอาศัยการเรียนรู้ที่สามารถเรียนรู้ได้ตลอดเวลาและสามารถตัดสินใจได้ตามสถานการณ์ที่แปรเปลี่ยนไป เพื่อรองรับกรณีที่ข้อมูลมีการแปรผันอยู่ตลอดเวลา ทำให้เกิดอุปสงค์ต่อแบบจำลองการเรียนรู้เชิงพลวัต ที่สามารถเรียนรู้และพัฒนาตัวเองได้ตลอดเวลาที่มีข้อมูลใหม่เข้ามา ในงานวิจัยนี้สนใจแบบจำลองที่มีความสามารถดังกล่าวซึ่งสามารถเรียนรู้ได้จากแหล่งข้อมูลขนาดใหญ่ที่เป็นสายข้อมูลแบบต่อเนื่อง และปราศจากการกำหนดข้อกำหนดเบื้องต้นจากมนุษย์ งานวิจัยนี้ได้เสนอแบบจำลองชนิดใหม่และวิธีการเรียนรู้ของแบบจำลองดังกล่าวเพื่อให้มีความสามารถตามที่กล่าวไว้ข้างต้น โดยอิงกระบวนทัศน์ในการเรียนรู้แบบ การจำแนกภายในจำกัด เรียกแบบจำลองชนิดใหม่นี้ว่า สโทแคสติกออโตมาตาอิงอักขระนำหน้าแบบจำกัด โดยที่อัลกอริทึมในการเรียนรู้ของแบบจำลองชนิดนี้มีฟังก์ชันการเติบโตของเวลาในการเรียนรู้ไม่เกินฟังก์ชันพหุนามและขนาดของแบบจำลองขึ้นอยู่กับบริบทของข้อมูล ซึ่งมีอัตราการเติบโตของขนาดไม่เกินฟังก์ชันพหุนามเช่นกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา.....วิศวกรรมคอมพิวเตอร์.....ลายมือชื่อนิสิต.....
สาขาวิชา.....วิศวกรรมคอมพิวเตอร์.....ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา.....2548.....

4670258821 : MAJOR COMPUTER ENGINEERING

KEY WORD: STOCHASTIC AUTOMATA / ON-LINE LEARNING / GRAMMAR INFERENCE / PROBABILISTIC AUTOMATA / IDENTIFICATION IN THE LIMIT

JITTAKORN PULLPOTHONG : ON-LINE PROBABILISTIC AUTOMATA ANYLYSIS AND DESIGN. THESIS ADVISOR : ATHASIS SURARERKS Ph.D., 50 pp. ISBN 974-53-2517-1.

Nowadays, there are several probabilistic (stochastic) language models such as Hidden Markov Model, Markov Chain, Probabilistic Automaton, etc. Most of recent researches on probabilistic models are to improve how to construct a static learning model. However, some work use dynamic mechanism to learn and update. In practice, we need a dynamic model which should be easier and faster to induce, and it should be able to update in real time. We are interested in a model which is able to learn from continuous data stream without prior knowledge or finite length of input. In this research, Identification in the limit is a learning criterion. We propose a novel probabilistic automaton named stochastic finite precedent automaton and its induction algorithm. Particularly, its induction or update algorithm can perform in an on-line mode within polynomial time complexity while the size of the model is dependent on the context of data.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department.....Computer Engineering.....Student's signature.....

Field of study...Computer Engineering.....Advisor's signature.....

Academic year....2005.....

กิตติกรรมประกาศ

ขอกราบขอบพระคุณอาจารย์ ดร. อรรถสิทธิ์ สุรฤกษ์ ผู้เป็นอาจารย์ที่ปรึกษาซึ่งคอยอบรมตักเตือนและให้คำปรึกษาที่มีคุณค่ามาโดยตลอด 3 ปี และขอกราบขอบพระคุณ อาจารย์ธงชัย โรจนกังสดาล ที่คอยชี้แนะศิลปะแห่งการจัดการความคิด รวมถึงหนังสือธรรมะที่อาจารย์ให้มา ซึ่งมีส่วนสำคัญอย่างมากที่ทำให้วิทยานิพนธ์นี้สำเร็จลุล่วงไปด้วยดี

ขอบคุณเพื่อนๆร่วมห้องปฏิบัติการทุกท่าน ที่มีส่วนทำให้การทำงานในห้องปฏิบัติการมีความสุขและมีความอบอุ่นเวลาทำงาน

ขอขอบคุณคณะวิศวกรรมศาสตร์มหาวิทยาลัยรามคำแหงที่คอยสนับสนุนด้านปัจจัย

สุดท้ายขอขอบพระคุณบิดา มารดาและคนที่อยู่ใกล้ที่คอยเป็นกำลังใจให้เสมอ รวมถึงท่านอื่นๆ ที่มีส่วนเกี่ยวข้องที่ดีที่ไม่ได้เอ่ยนามในที่นี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....ง
บทคัดย่อภาษาอังกฤษ.....จ
กิตติกรรมประกาศ.....ฉ
สารบัญ.....ช
สารบัญภาพ.....ฌ

บทที่

1. บทนำ..... 1
   1.1 ความเป็นมาและความสำคัญของปัญหา.....1
   1.2 วัตถุประสงค์ของการวิจัย.....4
   1.3 ขอบเขตของการวิจัย..... ...4
   1.4 ข้อจำกัดของการวิจัย.....4
   1.5 ประโยชน์ที่คาดว่าจะได้รับ.....5
   1.6 วิธีดำเนินการวิจัยและลำดับขั้นตอนในการเสนอผลการวิจัย.....5
2. เอกสารและงานวิจัยที่เกี่ยวข้อง..... 7
   2.1 แนวคิดและทฤษฎี.....7
   2.2 ทฤษฎีพื้นฐาน.....9
   2.3 แนวความคิดในการเรียนรู้แบบเชื่อมตรง.....11
   2.4 การเรียนรู้ด้วยการอุปนัย.....12
   2.5 ทฤษฎีการเรียนรู้ภาษา.....13
   2.6 เอกสารและงานวิจัยที่เกี่ยวข้อง.....16
3. วิธีดำเนินการวิจัย.....19
   3.1 การวิเคราะห์และเลือกทฤษฎีการเรียนรู้.....19
   3.2 เกณฑ์เทียบประสิทธิภาพ.....19
   3.3 การออกแบบคลาสของภาษา.....21
   3.4 แบบจำลองที่ใช้เรียนรู้ภาษา.....29
   3.5 อัลกอริทึมในการเรียนรู้.....31
   3.6 การคำนวณความน่าจะเป็นของสายอักขระ.....40

สารบัญ (ต่อ)

บทที่ หน้า

3.7 ผลการวิเคราะห์ข้อมูล.....41

4. สรุปผลการวิจัยและข้อเสนอแนะ.....45

4.1 สรุปผลการวิจัย.....45

4.2 ข้อเสนอแนะ.....46

รายการอ้างอิง.....48

ประวัติผู้เขียนวิทยานิพนธ์.....50

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

รูปที่ หน้า

1 แผนภาพต้นไม้อิงอักขระนำหน้าเชิงน่าจะเป็นที่สร้างได้จากสายอักขระ ‘235123’.....11

2 แผนภาพเอสเอฟพีเอที่ได้จากการเรียนรู้จากสายอักขระ ‘23512321351’.....40

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ระบบการจดจำรูปแบบข้อมูล (pattern recognition system) ทางวิทยาการคอมพิวเตอร์ ได้เข้ามามีบทบาทอย่างยิ่งต่อวิทยาการสาขาต่างๆ เช่น ระบบการตรวจจับการบุกรุก ระบบจดจำเสียงมนุษย์ ระบบจดจำรูปแบบข้อมูลดีเอนเอทางชีววิทยา และการเรียนรู้ของเครื่องจักร เป็นต้น วิธีการในการจดจำรูปแบบข้อมูลเหล่านี้อาศัยวิธีการเรียนรู้เป็นหลัก โดยมีงานวิจัยที่เกี่ยวข้องหลายด้าน เช่น การเรียนรู้ด้วยแบบจำลองโครงข่ายประสาทเทียม (artificial neural network) การเรียนรู้ด้วยอัลกอริทึมเชิงพันธุกรรม (genetic algorithm) และการเรียนรู้ด้วยการสร้างแบบจำลองการส่งผ่าน (transition model) ซึ่งแต่ละวิธีมีข้อดีข้อเสียแตกต่างกัน การเรียนรู้ด้วยแบบจำลองโครงข่ายประสาทเทียมและอัลกอริทึมเชิงพันธุกรรมจะอาศัยการคำนวณเพื่อให้ได้มาซึ่ง ฟังก์ชันที่เป็นสมการทางคณิตศาสตร์ที่ดีที่สุดซึ่งเหมาะสมต่อสภาพข้อมูล การเรียนรู้ด้วยแบบจำลองการส่งผ่านจะเป็นการหากฎเกณฑ์ที่แสดงวากยสัมพันธ์ (syntax) ของสายอักขระ ซึ่งการเรียนรู้ด้วยแบบจำลองการส่งผ่านที่มีประสิทธิภาพ ได้แก่ การเรียนรู้ด้วยแบบจำลองการส่งผ่านสถานะเชิงน่าจะเป็น (probabilistic state transition model) ซึ่งได้ถูกพัฒนาจนแยกย่อยเป็นแบบจำลองชนิดต่างๆมากมาย เช่น แบบจำลองมาร์คอฟเชิงซ่อน (hidden Markov model) แบบจำลองออโตมาตาเชิงน่าจะเป็น (probabilistic finite automata) แบบจำลองลูกโซ่มาร์คอฟ (Markov chain) เรียกอีกอย่างหนึ่งว่าแบบจำลอง เอ็นแกรม (N-gram) และแบบจำลองมาร์คอฟเชิงความยาวแปรผัน (variable memory length Markov model) เป็นต้น จากงานวิจัยที่ผ่านมาซึ่งได้กล่าวตามข้างต้น การเรียนรู้ด้วยแบบจำลองส่วนมากจะเป็นการเรียนรู้เชิงสถิต (static learning) [1, 2] ซึ่งจะเป็นการเรียนรู้เพียงครั้งเดียวเพื่อให้ได้แบบจำลองที่ครอบคลุมต่อเหตุการณ์ทั้งหมดในอนาคต การเรียนรู้เช่นนี้ค่อนข้างเหมาะสมต่อเหตุการณ์ที่ไม่ค่อยเปลี่ยนแปลงมาก เช่น การจดจำรูปแบบภาพ [3] เสียงพูด ลายนิ้วมือ ฯลฯ กระบวนทัศน์การเรียนรู้ชนิดนี้จึงยังค่อนข้างมองในมุมแคบ ซึ่งเหมาะสมเฉพาะแต่ละสถานการณ์ พบว่าในกรณีที่มีชุดข้อมูลใหม่จากสถานการณ์ใหม่เพิ่มเข้ามาซึ่งอาจจะทำให้สมมติฐานเดิมเปลี่ยนไป ทำให้การตัดสินใจด้วยความรู้เดิมของแบบจำลองผิดพลาด ในกรณีนี้จำเป็นต้องนำชุดข้อมูลเดิมและชุดข้อมูลใหม่รวมเข้าด้วยกันแล้วดำเนินวิธีการเรียนรู้ใหม่ ซึ่งอาจจะต้องกำหนดโครงสร้างของแบบจำลองใหม่ ทำให้เสียเวลาและเกิดความยุ่งยากตามมา

การพัฒนาทางเทคโนโลยีคอมพิวเตอร์ในปัจจุบันทั้งทางด้าน เทคโนโลยีอินเตอร์เน็ต หุ่นยนต์ และการพัฒนาของซอฟแวร์ ทำให้เกิดอุปสงค์ต่อการเรียนรู้ที่มีประสิทธิภาพมากกว่าเดิมที่มีโครงสร้างเชิงสถิต ซึ่งไม่อาจรองรับการเปลี่ยนแปลงที่รวดเร็วของข้อมูลที่มีขนาดมหาศาลในปัจจุบันได้ดีพอ ตัวอย่างของชุดข้อมูลขนาดใหญ่ที่มีการเปลี่ยนแปลงอยู่เสมอ เช่น การเรียนรู้ของหุ่นยนต์ด้วยข้อมูลหลากชนิด การสร้างแอพพลิเคชัน (application) ที่ตอบสนองต่อการใช้งานเว็บไซต์ (website) อย่างทันท่วงที เช่นการทำนายพฤติกรรมของผู้เรียกใช้งานเว็บไซต์ การโจรกรรมข้อมูลผ่านทาง ซิสเท็มคอล (system call) บนระบบปฏิบัติการ ยูนิกส์ (UNIX) และระบบวิเคราะห์ฐานข้อมูลการเงิน เป็นต้น จะเห็นได้ว่า รูปแบบของข้อมูลเหล่านี้มีขนาดใหญ่และซับซ้อนจนไม่สามารถคาดเดาได้ว่า อะไรจะเกิดขึ้นบ้างในอนาคต การตอบสนองข้อมูลเหล่านี้จึงจำเป็นต้องอาศัยวิธีการเรียนรู้ที่มีความสามารถในการเรียนรู้เพื่อเพิ่มความรู้แบบต่อเนื่องระยะยาวได้โดย ไม่จำกัดโครงสร้างของแบบจำลอง และในขณะเดียวกันจะต้องมีความสามารถตัดสินใจได้พร้อมกับการเรียนรู้ไปด้วย เรียกการเรียนรู้เช่นนี้ว่า การเรียนรู้เชิงพลวัต (dynamic learning) งานวิจัยนี้จึงสนใจวิธีการเรียนรู้ที่มีความสามารถในการเรียนรู้แบบเชื่อมตรง (on-line learning) จากสายอักขระ โดยไม่จำกัดทั้งความยาวของสายอักขระ จำนวนอักขระ จำนวนครั้งในการเรียนรู้ และโครงสร้างของแบบจำลอง เพื่อรองรับการเรียนรู้เชิงพลวัตดังกล่าว และเพื่อนำไปสู่การพัฒนาระบบการเรียนรู้อย่างอัตโนมัติมากขึ้น ในงานวิจัยนี้จะใช้คำจำกัดความการเรียนรู้ชนิดนี้ใหม่ว่า *การเรียนรู้เชิงตลอดชีพ* (lifelong learning) ซึ่งเป็นการเรียนรู้ที่สามารถเรียนรู้เพิ่มเติมเพื่อสร้างองค์ความรู้มากขึ้นได้ โดยที่สามารถเปรียบเทียบการเรียนรู้เชิงตลอดชีพนี้เช่นเดียวกับการเรียนรู้ของมนุษย์

ปัญหาการออกแบบการเรียนรู้ในแบบที่ต้องการจึงถูกแบ่งออกเป็น 2 กรณีหลัก คือ การเลือกแบบจำลองที่เหมาะสม และวิธีการเรียนรู้ที่เหมาะสมสำหรับการเรียนรู้เชิงตลอดชีพ ในส่วนของการเลือกแบบจำลองที่เหมาะสม งานวิจัยนี้จะให้ความสนใจแบบจำลองการส่งผ่านอิงความน่าจะเป็น เนื่องจากวิธีการเรียนรู้ด้วยแบบจำลองชนิดนี้อาศัยการพิจารณาจากหลักเกณฑ์ทางวากยสัมพันธ์ของสายอักขระ พร้อมกับการคำนวณความน่าจะเป็นเพื่อที่จะรองรับภาษาที่หลากหลายตามกระบวนทัศน์การเรียนรู้ที่ต้องการ ด้วยความรู้ในขณะนี้ของผู้วิจัย การเรียนรู้ด้วยแบบจำลองชนิดนี้จึงน่าจะเหมาะสมต่อการเรียนรู้ข้อมูลที่เป็นสายอักขระที่หลากหลาย มากกว่าแบบจำลองที่เน้นการคำนวณเพื่อหาแต่เฉพาะสมการทางคณิตศาสตร์ที่เหมาะสม ดังเช่นแบบจำลองโครงข่ายประสาทเทียม หรือ อัลกอริทึมเชิงพันธุกรรม ซึ่งมีกระบวนทัศน์ในการเรียนรู้ (learning paradigm) อันจำกัดขอบเขต จากงานวิจัยเกี่ยวกับแบบจำลองการส่งผ่านสถานะอิงความน่าจะเป็น [1, 2] พบว่า แบบจำลองที่มีประสิทธิภาพที่ได้กล่าวนำไว้ข้างต้น ไม่ว่าจะเป็น

แบบจำลองมาร์คอฟเชิงซ่อน หรือแบบจำลองลูกโซ่มาร์คอฟยังคงมีข้อเสียที่ไม่เหมาะต่อการนำมาเรียนรู้เชิงตลอดชีพ กล่าวคือแบบจำลองมาร์คอฟเชิงซ่อนไม่สามารถสร้างขึ้นได้อัตโนมัติโดยปราศจากการกำหนดโครงสร้างเบื้องต้นจากมนุษย์ อีกทั้งกระบวนการเรียนรู้เพื่อให้ได้มาซึ่งโครงสร้างและพารามิเตอร์ของแบบจำลองชนิดนี้ ต้องใช้ข้อมูลสายอักขระที่มีคุณลักษณะของความยาวคงตัวซึ่งมีลักษณะการเรียนรู้ที่เป็นแบบเชิงสถิต รวมถึงการเรียนรู้ด้วยแบบจำลองมาร์คอฟเชิงซ่อนต้องใช้เวลาสูงมาก จึงไม่เหมาะต่อการเรียนรู้ที่ต้องการทั้งความเร็วในการเรียนรู้และการตอบสนองต่อการจดจำข้อมูล ในส่วนของการเรียนรู้ด้วยแบบจำลอง เอ็น-แกรม จำเป็นต้องมีการกำหนดขนาดความยาวของสายอักขระหรือขนาดความยาวของลูกโซ่มาร์คอฟ ให้มีขนาดที่แน่นอนก่อนทำการเรียนรู้ ข้อสังเกตคือความยาวเท่าไรจึงจะเหมาะสมและมีความมาตรฐานมากที่สุดต่อข้อมูลทั่วไป ด้วยเหตุนี้แบบจำลองชนิดนี้จึงไม่เหมาะต่อการเรียนรู้เชิงตลอดชีพ และสำหรับแบบจำลองมาร์คอฟเชิงความยาวแปรผันนั้น พบว่าวิธีการนี้สามารถคำนวณหาความยาวลูกโซ่ที่เหมาะสมซึ่งแปรผันตามความน่าจะเป็นที่คำนวณได้ อย่างไรก็ตามด้วยความรู้ที่มีในขณะนี้ของผู้วิจัย วิธีการเรียนรู้ด้วยแบบจำลองชนิดนี้ยังไม่อาจจะคำนวณได้ด้วยกระบวนการแบบเชื่อมตรง จึงกล่าวได้ว่าแบบจำลองชนิดนี้ยังไม่เหมาะสมต่อการเรียนรู้ในแบบที่ต้องการ งานวิจัยนี้สนใจไปที่การเรียนรู้ด้วยแบบจำลองออโตมาตาเชิงน่าจะเป็น เนื่องจากงานวิจัยหลายงาน [1, 2] ได้กล่าวสนับสนุนว่า ข้อกำหนดและกฎเกณฑ์ในการสร้างแบบจำลองดังกล่าวสามารถยืดหยุ่นได้สูง โดยข้อกำหนดความถูกต้องในการเรียนรู้นั้นไม่เข้มงวด (weak criterion) [4] ซึ่งเหมาะต่อสภาวะข้อมูลที่แปรผันต่อเนื่องที่ต้องการแบบจำลองที่ไม่จำกัดโครงสร้าง อีกทั้งในส่วนของวิธีการเรียนรู้เพื่อให้ได้มาซึ่งโครงสร้างและพารามิเตอร์ของแบบจำลองชนิดนี้ ได้มีทฤษฎีของโกลด์ [4] ที่มีชื่อเรียกกระบวนทัศน์การเรียนรู้ว่า *การจำแนกภายในจำกัด* (identification in the limit) เป็นตัวกำหนดความสำเร็จของการเรียนรู้ ซึ่งมีแนวความคิดที่ตรงและเหมาะสมกับการเรียนรู้เชิงตลอดชีพตามแบบที่ต้องการ ตามความเห็นของผู้วิจัยในขณะนี้ แบบจำลองออโตมาตาเชิงน่าจะเป็นจึงเป็นแบบจำลองที่เหมาะสมต่อการเรียนรู้เชิงตลอดชีพตามวิธีการของงานวิจัยนี้

จากคุณสมบัติของแบบจำลองออโตมาตาเชิงน่าจะเป็นและทฤษฎีการเรียนรู้ของ โกลด์ งานวิจัยนี้ได้ทำการพัฒนาวิจัยผลงานและนำเสนอ วิธีการเรียนรู้แบบใหม่ซึ่งมีคุณสมบัติในการเรียนรู้เชิงตลอดชีพซึ่งไม่จำกัดเหตุการณ์(event) ความยาวของสายอักขระ และโครงสร้างของแบบจำลอง โดยเสนอคลาสของภาษาที่ถูกออกแบบเพื่อใช้สำหรับการเรียนรู้ และแบบจำลองชนิดใหม่ ซึ่งมีคุณสมบัติเป็นคลาสย่อยของออโตมาตาเชิงน่าจะเป็น งานวิจัยนี้เรียกแบบจำลองชนิดใหม่นี้ว่า *แบบจำลองสโทแคสติกออโตมาตาอิงคำนำหน้าแบบจำกัด* (stochastic finite precedent automata: SFPA) ซึ่งมีจุดประสงค์เพื่อช่วยในการเรียนรู้เชิงตลอดชีพ นอกจากนี้

งานวิจัยนี้ได้นำเสนออัลกอริทึมที่ใช้ในการเรียนรู้ด้วยการอุปนัยโครงสร้างแบบจำลองดังกล่าว เพื่อให้ถูกต้องตรงกับคลาสของภาษาที่ออกแบบไว้ โดยอาศัยกระบวนทัศน์การเรียนรู้ของ โกลด์ มากำหนดขอบเขตความสำเร็จ งานวิจัยนี้ได้ทำการวิเคราะห์ภาษาที่ใช้ในการเรียนรู้ การประมาณความน่าจะเป็นที่คำนวณจากสายอักขระ และการลู่เข้าสู่ค่าความน่าจะเป็นที่ถูกต้องของสายอักขระใดๆที่ได้จากการเรียนรู้ ในการตรวจสอบว่าแบบจำลองการเรียนรู้มีประสิทธิภาพนั้น งานวิจัยนี้ได้ใช้หลักการทางทฤษฎีมาวิเคราะห์ประสิทธิภาพของอัลกอริทึมและวิธีการเรียนรู้ว่ามีประสิทธิภาพตรงกับที่ได้นิยามไว้

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อคิดค้นและได้มาซึ่งวิธีการเรียนรู้จากสายอักขระด้วยแบบจำลองออโตมาตาเชิงน่าจะเป็น ด้วยความสามารถเรียนรู้เพิ่มเติมได้ไม่จำกัดจำนวนครั้ง

1.2.2 เพื่อคิดค้นและได้มาซึ่งวิธีการเรียนรู้จากสายอักขระได้โดยไม่จำกัดโครงสร้างของแบบจำลองหรือ ความยาวสายอักขระ

1.2.3 เพื่อคิดค้นและได้มาซึ่งวิธีการเรียนรู้จากสายอักขระโดยไม่จำกัดจำนวนอักขระที่เป็นไปได้

1.2.4 เพื่อคิดค้นและได้มาซึ่งวิธีการเรียนรู้จากสายอักขระที่สามารถดำเนินการได้ด้วยวิธีการเชื่อมตรง

1.2.5 เพื่อคิดค้นและได้มาซึ่งวิธีการเรียนรู้ชนิดใหม่ซึ่งเป็นแนวทางในงานวิจัยทางด้านการเรียนรู้ในอนาคต

1.3 ขอบเขตของการวิจัย

1.3.1 งานวิจัยเน้นที่การเรียนรู้จากสายอักขระเพื่อให้ได้มาซึ่งออโตมาตาเชิงน่าจะเป็น

1.3.2 ออโตมาตาเชิงน่าจะเป็นภายหลังที่ได้จากการเรียนรู้ยังไม่มีการกำหนดขอบเขตของขนาด

1.3.3 แบบจำลองมีความสามารถในการรู้จำพร้อมกับปรับปรุงค่าพารามิเตอร์ต่างๆพร้อมกันไป

1.3.4 ในการเรียนรู้สามารถมีอักขระใหม่เกิดขึ้นในระหว่างการเรียนรู้ได้

1.3.5 แบบจำลองมีหลักการคำนวณความน่าจะเป็นของสายอักขระต่างๆเช่นเดียวกับหลักการแบบจำลองมาร์คอฟแบบหน่วยความจำแปรผัน (variable memory length Markov model)

1.4 ข้อจำกัดของการวิจัย

1.4.1 การเรียนรู้มีประสิทธิภาพจำกัดเฉพาะการเรียนรู้ภาษาในคลาสของภาษาสม่ำเสมอ (regular language)

1.4.2 การเกิดความคลาดเคลื่อนของความน่าจะเป็นที่ได้จากการเรียนรู้จะขึ้นอยู่กับข้อมูลที่นำมาเรียนรู้

1.4.3 ยังไม่สามารถทำให้ขนาดของแบบจำลองที่ได้จากการเรียนรู้มีขนาดเล็กที่สุดได้

1.4.4 ยังไม่สามารถจำกัดความคลาดเคลื่อน ณ เวลาใดๆ ระหว่างการเรียนรู้ให้น้อยที่สุดได้

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 เพิ่มความสามารถในเรียนรู้เชิงตลอดชีพของเครื่องจักร

1.5.2 เพิ่มความสามารถในการเรียนรู้โดยอัตโนมัติของเครื่องจักร

1.5.3 เพิ่มความสามารถในการเรียนรู้แบบเชื่อมตรง

1.5.4 เป็นแนวทางในการปรับปรุงปัญญาประดิษฐ์ให้มีประสิทธิภาพมากขึ้น

1.5.5 สามารถนำหลักการไปประยุกต์ใช้กับการจดจำรูปแบบข้อมูลหลากหลายชนิด เช่น การจดจำเสียงพูด การจดจำภาพ การจดจำรูปแบบการบุกรุก การจดจำดีเอ็นเอ การทำนายการใช้งานเว็บบราวเซอร์

1.6 วิธีดำเนินการวิจัยและลำดับขั้นตอนในการเสนอผลการวิจัย

การที่จะออกแบบแบบจำลองใหม่ขึ้นมาเพื่อรองรับการเรียนรู้สายอักขระแบบเชื่อมตรงจำเป็นจะต้องมีข้อพิจารณาหลายอย่างเช่น ทฤษฎีที่รองรับการเรียนรู้ หลักเกณฑ์ที่งานวิจัยนี้ยอมรับได้ว่าการเรียนรู้ถูกต้อง [5] และมีประสิทธิภาพ การออกแบบวิธีการเรียนรู้ทั้งหมดนี้งานวิจัยนี้เรียกว่ากรอบงานในการออกแบบเรียนรู้ (framework) โดยกรอบงานในวิธีการดำเนินการวิจัยของงานวิจัยนี้มีหลักการดังนี้

1.6.1 วิเคราะห์และเลือกทฤษฎีการเรียนรู้ที่กำหนดหลักการความถูกต้องให้เหมาะสมต่อสภาพปัญหา

1.6.2 ออกแบบคลาสของภาษาให้สอดคล้องกับทฤษฎีดังกล่าวโดยจะต้องพิจารณาตามหลักเกณฑ์ต่างๆดังจะได้กล่าวต่อไปเพื่อรับประกันว่าคลาสของภาษาที่ออกแบบไว้สามารถครอบคลุมทั่วถึงภาษาที่ได้จากข้อมูลตัวอย่างรวมถึงความสามารถในการทำนายข้อมูลที่ไม่เคยเห็นมาก่อนได้อย่างมีประสิทธิภาพ

1.6.3 ออกแบบหรือเลือกแบบจำลองทางคณิตศาสตร์ที่สอดคล้องกับคลาสของภาษาที่ออกแบบไว้

1.6.4 หาอัลกอริทึมที่สังเคราะห์แบบจำลองให้ตรงตามคลาสของภาษาที่ออกแบบไว้โดยอาศัยหลักเกณฑ์ของความถูกต้องและประสิทธิภาพให้ตรงกับหลักทฤษฎีที่วางไว้

1.6.5 พิสูจน์ว่าภาษาที่ออกแบบและอัลกอริทึมในการเรียนรู้นั้นมีประสิทธิภาพจริงตามหลักการ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 แนวคิดและทฤษฎี

งานวิจัยเกี่ยวกับการรู้จำระบบนั้น สามารถจำแนกได้หลายชนิดขึ้นอยู่กับชนิดและรูปแบบข้อมูลที่สนใจ สำหรับงานวิจัยนี้สนใจการเรียนรู้จากข้อมูลที่เป็นสายอักขระ โดยไม่จำกัดความยาวของสายอักขระที่จะนำมาใช้ในการเรียนรู้ การสร้างแบบจำลองที่ใช้ในการเรียนรู้สายอักขระนั้น จำเป็นต้องอาศัยทฤษฎีการคำนวณเกี่ยวกับเรื่องของภาษาและออโตมาตามาวิเคราะห์เนื่องจากการศึกษาทางทฤษฎีภาษาและออโตมาตาได้พิสูจน์ให้เห็นแล้วว่า เครื่องจักรหรือแบบจำลองที่มีความสามารถในการยอมรับภาษาชนิดต่างๆนั้นมีจำกัด ตามความต้องการของงานวิจัยนี้คือสายข้อมูลที่จะนำมาเรียนรู้ไม่ถูกจำกัดความยาว และไม่จำกัดจำนวนอักขระ ซึ่งแน่นอนว่าหากว่าสายอักขระเหล่านี้อยู่ในคลาสของภาษาเวียนเกิดแบบแจงนับ (recursive enumerable) ซึ่งเป็นคลาสภาษาที่ใหญ่มากแล้ว การหาแบบจำลองมายอมรับภาษาดังกล่าวย่อมสามารถมีได้หลายชนิดเช่น ออโตมาตาแบบจำกัด (finite automata) ออโตมาตาแบบกดลง (push down automata) เครื่องจักรทูริ่ง (turing machine) เป็นต้น ซึ่งไม่อาจจะรู้ได้ว่าสายอักขระเหล่านั้นอยู่ในคลาสใดของภาษา ดังนั้นการเลือกแบบจำลองเพื่อการยอมรับภาษาเหล่านั้นย่อมมีความสำคัญ ซึ่งสิ่งที่จะต้องพิจารณาคือขอบเขตความสามารถของแบบจำลองที่จะนำมาเรียนรู้ภาษา

ในงานวิจัยของโกลด์ ได้ทำการวิเคราะห์ว่า หากชุดข้อมูลตัวอย่างที่จะนำมาเรียนรู้เป็นภาษาที่อยู่ในคลาสของภาษาเวียนเกิดแบบแจงนับแล้ว การเรียนรู้ด้วยการอุปนัยเพื่อให้ได้มาซึ่งแบบจำลอง อาศัยเฉพาะสายอักขระอย่างเดียวนั้นจะไม่เพียงพอต่อการเรียนรู้ให้เกิดความสำเร็จได้ [4] จำเป็นจะต้องใช้ตัวช่วยในการบอกว่า สายอักขระใดอยู่ในกลุ่มข้อมูลที่ถูกต้อง หรือเรียกอีกอย่างหนึ่งว่ากลุ่มข้อมูลเชิงบวก (positive example) และสายอักขระใดไม่อยู่ในกลุ่มที่ถูกต้องเรียกอีกอย่างหนึ่งว่ากลุ่มข้อมูลเชิงลบ (negative example) ในเชิงปฏิบัติ การที่จะหากลุ่มข้อมูลเชิงลบได้ครบทุกกรณีนั้นเป็นไปได้ยาก จากการพัฒนางานวิจัยต่อมา พบว่าตัวช่วยดังกล่าวที่มีประสิทธิภาพชนิดหนึ่งก็คือ ความน่าจะเป็น [5, 6, 7, 8] การเรียนรู้ภาษาที่สามารถกำหนดความสำเร็จได้จึงใช้ความน่าจะเป็นเข้ามาร่วมด้วย ซึ่งภาษาดังกล่าวก็คือ *ภาษาสโทแคสติก* (stochastic language) [9] ในส่วนของแบบจำลองคณิตศาสตร์ที่ทำหน้าที่เป็นฟังก์ชันคำนวณความน่าจะเป็นในการเกิดสายอักขระใดเทียบเท่ากับ ภาษาสโทแคสติก เรียกแบบจำลองดังกล่าว

ว่า *ออโตมาตาเชิงน่าจะเป็น* (probabilistic finite automata) ในวิทยานิพนธ์นี้จะใช้คำว่า *สโทแคสติกออโตมาตา* (stochastic finite automata) เพื่อให้สอดคล้องกับความหมายของภาษาสโทแคสติก

งานวิจัยที่ผ่านมาเกี่ยวกับการเรียนรู้และจดจำสายอักขระด้วยแบบจำลองออโตมาตาเชิงน่าจะเป็นด้วยวิธีการเรียนรู้เชิงสถิต จะเน้นไปที่การคำนวณหาความน่าจะเป็นที่มากที่สุดตามเงื่อนไขซึ่งสอดคล้องกับข้อมูลตัวอย่างที่ออกแบบไว้เบื้องต้น ดังนั้นการคำนวณความน่าจะเป็นที่มากที่สุดจะเหมาะสมแต่เฉพาะข้อมูลที่เอามาเรียนรู้ จึงเรียกประสิทธิภาพของการเรียนรู้เช่นนี้ว่าประสิทธิภาพแบบเฉพาะที่ (local effective) แต่เนื่องจากการเรียนรู้ที่ต้องการเป็นการเรียนรู้เชิงตลอดชีพ ดังนั้นการเรียนรู้แบบที่ต้องการนั้นจึงไม่ควรเหมาะสมต่อเฉพาะข้อมูลที่นำมาสอนในขณะนั้นเท่านั้น แต่ควรเหมาะสมต่อทุกข้อมูลที่นำมาสอนในระยะยาวซึ่งสามารถรองรับกับหลักการเชื่อมตรงได้ เรียกการเรียนรู้เช่นนี้ว่าเป็นการเรียนรู้ให้เกิดประสิทธิภาพแบบถ้วนทั่ว (global effective)

ด้วยเหตุที่ต้องการประสิทธิภาพแบบถ้วนทั่วเพื่อรองรับการเรียนรู้ในระยะยาวนี้เอง งานวิจัยนี้จึงจำเป็นจะต้องออกแบบการเรียนรู้ใหม่ โดยที่คำนึงถึงการอุปนัยโครงสร้าง (Inductive Inference) ของสโทแคสติกออโตมาตาในระยะยาว ซึ่งรองรับกับสายอักขระทุกกรณี และการประมาณความน่าจะเป็นที่สามารถคำนวณได้ด้วยสโทแคสติกออโตมาตา ซึ่งสามารถลู่เข้าสู่ค่าที่ถูกต้องตามหลักการเรียนรู้ด้วยกระบวนการแบบเชื่อมตรงได้ งานวิจัยนี้จึงอาศัยทฤษฎีในการเรียนรู้เพื่อกำหนดขอบเขตของความสำเร็จ 2 ทฤษฎีคือ ในส่วนของการอุปนัยโครงสร้างของสโทแคสติกออโตมาตาจะใช้กระบวนทัศน์การเรียนรู้ของโกลด์ เพื่อกำหนดว่าเมื่อใดจึงจะหยุดปรับเปลี่ยนโครงสร้างของออโตมาตาในกรณีที่การเรียนรู้รับประกันว่าสำเร็จ โดยที่มีประสิทธิภาพทางด้านเวลาที่เหมาะสม และในส่วนของการประมาณความน่าจะเป็นของสโทแคสติกออโตมาตาจะใช้กระบวนทัศน์การเรียนรู้แบบเชื่อมตรงเพื่อกำหนดว่าการเรียนรู้นี้มีประสิทธิภาพตามหลักการ

จากวิธีการอุปนัยโครงสร้างของแบบจำลองสโทแคสติกออโตมาตา ที่ผ่านมาจะใช้หลักการสร้างแผนภาพต้นไม้เพื่อยอมรับข้อมูลสายอักขระที่สามารถเป็นไปได้ทั้งหมดที่มาจากชุดข้อมูลตัวอย่าง เราเรียกแผนภาพต้นไม้ที่ยอมรับข้อมูลนี้นี้ว่า *ต้นไม้อิงคำนำหน้าเชิงน่าจะเป็น* (probabilistic prefix tree acceptor :PPTA) จากนั้นจะทำการผสานสถานะที่มีความน่าจะเป็นในการส่งผ่านในทุกเหตุการณ์ที่เหมือนกันเข้าด้วยกัน ในกรณีนี้จะทำให้โครงสร้างของสโทแคสติกออโตมาตาไม่ใหญ่เกินไป อย่างไรก็ดี ในกระบวนการแบบเชื่อมตรงเราไม่อาจจะใช้วิธีการนี้ได้ งานวิจัยนี้จึงได้เสนอการอุปนัยโครงสร้างออโตมาตาขึ้นใหม่ซึ่งจะกล่าวในบทที่ 3 ในบทนี้จะสรุปทฤษฎีพื้นฐานอันเริ่มแต่สัญลักษณ์ที่จำเป็นต่อการทำความเข้าใจในเรื่องต่อไป ทฤษฎีทางภาษาส

โทแคสติก สโทแคสติกออโตมาตา ทฤษฎีการเรียนรู้ ทฤษฎีกำหนดความสำเร็จในการเรียนรู้แบบของโกลด์ หลักการกำหนดความสำเร็จของการประมาณความน่าจะเป็นแบบเชื่อมตรง และงานวิจัยที่เกี่ยวข้อง

2.2 ทฤษฎีพื้นฐาน

2.2.1 สัญลักษณ์และคำจำกัดความพื้นฐาน

กำหนดให้ $\sum$ แทนเซตของอักขระใดๆ (Alphabet) และ $\sum^*$ แทนเซตของสายอักขระใดๆ ที่มาจาก $\sum$ และสายอักขระว่าง (empty string) ถูกแทนด้วย ε ให้ $u \in \sum^*$ แทนสายอักขระ (string) ใดๆ แล้ว $u\sum^*$ เป็นเซตของสายอักขระที่มีคำเติมหน้าเป็น u กำหนดให้ $s = s_1\ s_2\ s_3 \ldots\ s_l$ เป็นสายอักขระใดๆ ที่มีความยาว l โดยที่ s_i คือสมาชิกใน $\sum$ ขนาดของ s แทนด้วย $|s|$. และเซตของคำเติมท้ายทั้งหมดของ s แทนได้ด้วย $suffix^*(s) = \{s_i \ldots s_l \mid 1 \leq i \leq l\} \cup \{\varepsilon\}$ และคำเติมท้ายที่ยาวที่สุดของ s ซึ่งไม่เท่ากับ s เขียนแทนด้วย $suffix(s) = s_2 \ldots s_l$ ถ้ากำหนดให้ $s=uv$ จะได้ว่า $v = u^{-1}s$ และ $u = sv^{-1}$ กำหนดให้ $L \subseteq \sum^*$, L จะถูกเรียกว่าภาษา และสมาชิกในภาษาถูกเรียกว่าคำ (word)

2.2.2 ภาษาสโทแคสติก (Stochastic language)

ภาษาสโทแคสติก คือ ฟังก์ชันการกระจายข้อมูลใดๆ ที่มีโคโดเมน(co-domain)เป็นจำนวนจริง [0, 1] และมีโดเมนเป็น $\sum^*$ แทนได้ด้วย D: $\sum^* \rightarrow$ [0, 1] ซึ่งจะต้องมีคุณสมบัติ $\sum_{u \in \sum^*} D(u) \leq 1$ โดยมีภาษาที่รองรับฟังก์ชันการกระจายข้อมูล D นี้เป็นเซตย่อย (subset) ของ $\sum^*$ เขียนแทนได้ด้วย $L(D) \subseteq \sum^*$ โดยที่ $L(D)$ เป็นเซตของคำใดๆ $u \in \sum^*$ โดยที่ $D(u) > 0$ เราเรียก $L(D)$ ว่าเป็น *ภาษาสนับสนุน* (Support language) ของภาษาสโทแคสติก D และเรียก *ผลหาร* (Quotient) ของภาษาสโทแคสติกซึ่งมีภาษา $L(D)$ สนับสนุน เขียนแทนด้วย $x^{-1}L(D)$ ซึ่งเราสามารถนิยาม ผลหารของภาษาสโทแคสติก ด้วยความน่าจะเป็นของสายอักขระใดๆใน $L(D)$ ที่มีคำเติมหน้าด้วยสายอักขระ x เขียนแทนด้วยสมการ

$$p(u \mid x^{-1}L(D)) = \frac{p(xu \mid L(D))}{p(x\sum^* \mid L(D))}$$

ในส่วนของแบบจำลองทางคณิตศาสตร์ซึ่งมีความสามารถในการคำนวณความน่าจะเป็นของสายอักขระใดๆ เทียบเท่าภาษาสโทแคสติกได้แก่ แบบจำลองสโทแคสติกออโตมาตา ซึ่งนิยามดังต่อไปนี้

2.2.3 **สโทแคสติกออโตมาตาแบบจำกัด** (Stochastic finite automaton)

ให้ สโทแคสติกออโตมาตา ประกอบไปด้วย [Q, $\sum$, ε, τ, γ] โดยที่

Q คือ เซตจำกัดของสถานะ

$\sum$ คือเซตของอักขระที่เป็นไปได้ทั้งหมด

ε: คือสถานะเริ่มต้น

τ: Q×$\sum$→Q คือ ฟังก์ชันการส่งผ่านระหว่างสถานะ

γ: Q×$\sum$→[0,1] คือ ฟังก์ชันการกระจายของการส่งผ่าน

ซึ่งสโทแคสติกออโตมาตาจะมีคุณสมบัติที่สำคัญคือ *"ผลรวมของความน่าจะเป็นในทุกการส่งผ่านจากสถานะใดๆจะมีค่าไม่เกิน1"* และเซตของสายอักขระ x ใดๆที่สามารถสร้างขึ้นหรือยอมรับด้วยสโทแคสติกออโตมาตา จะเป็นเซตย่อยของ $\sum^*$ โดยที่เราสามารถคำนวณความน่าจะเป็นในการสร้างหรือยอมรับสายอักขระ x ได้จากสมการ

$$P(x) = \prod_{i=1}^{n} \gamma(q_{i-1}, x_i, q_i)$$

โดยที่ q_0 คือสถานะเริ่มต้นและมีเส้นทางในการส่งผ่านสถานะ q_{i-1} ไปยัง q_i ด้วยเหตุการณ์ $x_i \in \sum$ มีอยู่จริง

2.2.4 **ต้นไม้อิงคำนำหน้าเชิงน่าจะเป็น** (Probabilistic Prefix Tree)

ในการอุปนัยสโทแคสติกออโตมาตา M นั้นส่วนมากจะใช้วิธีการผสานสถานะ ซึ่งมีหลักการคือ สร้างแผนภาพต้นไม้เชิงน่าจะเป็นที่ยอมรับสายอักขระที่เป็นไปได้ทุกสาย จากนั้นจะทำการผสานสถานะที่มีความน่าจะเป็นในการส่งผ่านใกล้เคียงกันเข้าด้วยกัน ต้นไม้อิงคำนำหน้าเชิงน่าจะเป็นถูกนิยามได้ดังนี้ ให้ *ต้นไม้อิงคำนำหน้าเชิงน่าจะเป็น* T = [Q, $\sum$, q_i, τ, γ] ซึ่งแต่ละรายการถูกนิยามเช่นเดียวกับออโตมาตาเชิงน่าจะเป็น โดยที่มีโครงสร้างพารามิเตอร์ตามนิยามดังนี้

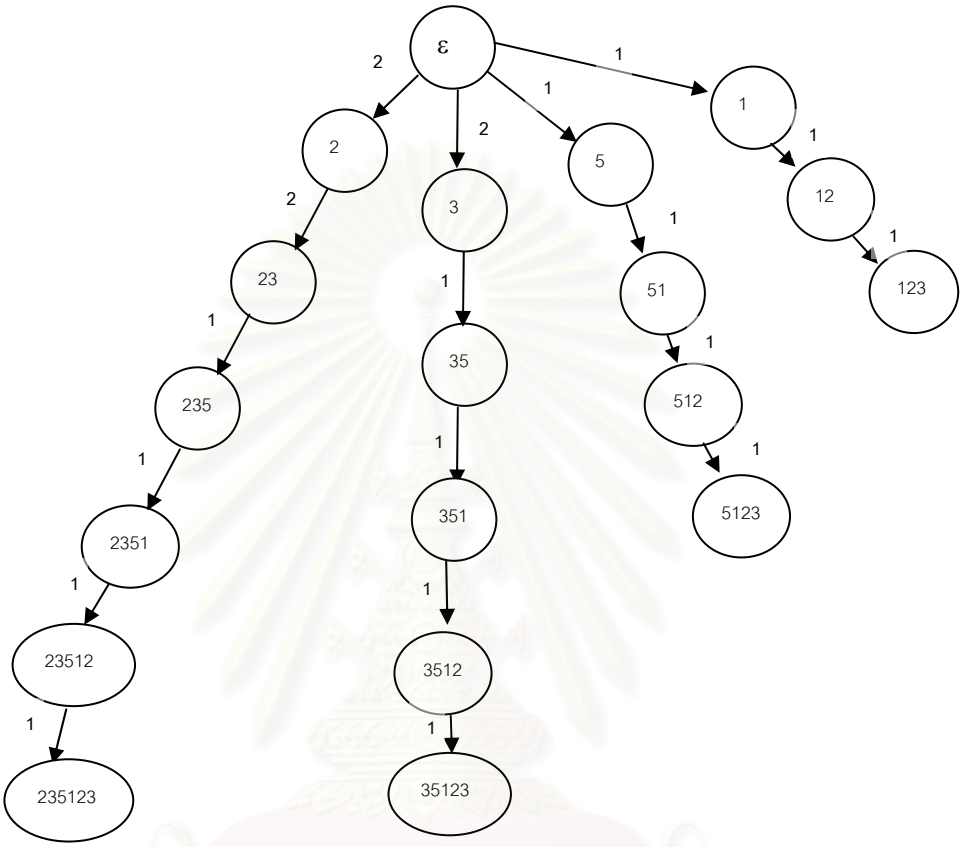
$$Q_t = \{ x \in \sum^* : x^{-1}L \neq NULL \}$$

$$\tau(x,a) = \begin{cases} xa & \text{if } xa \in Q_t \\ NULL & \text{otherwise} \end{cases}$$

$$q_i = \varepsilon$$

$$\gamma(x,a) = c(xa\sum^*)/c(x\sum^*)$$

โดยที่ c คือความถี่ในการส่งผ่านของแต่ละสถานะด้วยเหตุการณ์ต่างๆ



รูปที่ 1 แผนภาพต้นไม้อิงคำนำหน้าเชิงน่าจะเป็นที่สร้างได้จากสายอักขระ “235123”

2.3 แนวความคิดในการเรียนรู้แบบเชื่อมตรง (on-line learning concept)

เนื่องจากข้อมูลที่สนใจจะเรียนรู้เป็นสายอักขระไม่จำกัดความยาว ไม่จำกัดจำนวนอักขระที่เป็นไปได้ และไม่จำกัดโครงสร้าง เราจึงสนใจการเรียนรู้แบบเชื่อมตรงซึ่งจะใช้ข้อมูลตัวอย่างเป็นตัวกำหนดโครงสร้าง และอักขระที่เป็นไปได้ ความหมายของการเรียนรู้แบบเชื่อมตรงก็คือ การเรียนรู้ที่สามารถรับเอาข้อมูลปริมาณเล็กน้อยมาใช้เรียนรู้ได้ตลอดเวลา แทนที่จะใช้ข้อมูลทั้งหมดมาทำการเรียนรู้ในครั้งเดียว ด้วยลักษณะการเรียนรู้เช่นนี้เองที่เราไม่อาจจะคำนวณค่าความน่าจะเป็นที่มีค่าสูงสุดได้เหมือนกับการเรียนรู้แบบออฟไลน์ (off-line or batch) ซึ่งเป็นการเรียนรู้ที่มีประสิทธิภาพแบบเฉพาะที่ เนื่องจากปริมาณข้อมูลที่ใช้ในการเรียนรู้แต่ละครั้งมีจำกัด ไม่อาจจะคำนวณความน่าจะเป็นได้ทั้งหมด อีกทั้งการนำข้อมูลเดิมทั้งหมดมารวมเข้ากับข้อมูลใหม่แล้วทำการคำนวณก็ต้องใช้เวลาในการคำนวณมากเช่นกัน ซึ่งไม่ทันต่อการตอบสนองต่อข้อมูลที่ต้องการทั้งการเรียนรู้และการทำนายไปพร้อมกัน การเรียนรู้แบบเชื่อมตรงจึงเน้นไปที่แนวความคิดว่า ทำ

อย่างไรจะลดค่าความคลาดเคลื่อนให้น้อยที่สุด ณ ช่วงเวลาที่รับข้อมูลเข้ามาเรียนรู้ อย่างไรก็ตาม เมื่อเรียนรู้ไปแล้วระยะหนึ่ง วิธีการเรียนรู้จะต้องรับประกันได้ว่า ค่าความคลาดเคลื่อนจะลู่เข้าสู่ค่าศูนย์ หรือค่าที่มีปริมาณน้อยที่สุดเท่าที่จะทำได้ [10]

อย่างไรก็ตาม งานวิจัยเกี่ยวกับการเรียนรู้แบบเชื่อมตรงที่ผ่านมานั้น จะเน้นที่การเรียนรู้แบบเชื่อมตรงในกรอบการเรียนรู้แคบๆ ตัวอย่างเช่น การเรียนรู้บนการเดินหมากรุก การเรียนรู้บนเกมส์ การเรียนรู้บนหุ่นยนต์ที่โต้ตอบในแต่ละเฉพาะเหตุการณ์ วิธีการส่วนใหญ่ที่ใช้ในการเรียนรู้ชนิดนี้จึงเน้นไปที่การหาฟังก์ชันทางคณิตศาสตร์ที่ดีที่สุด เพื่อให้เกิดความคลาดเคลื่อนน้อยที่สุดระหว่างการเรียนรู้ ตัวอย่างวิธีการเหล่านี้ เช่น อัลกอริทึม ออนไลน์ เกรเดียน เดสเซนต์ (on-line gradient descent) อัลกอริทึม ควอไซ นิวตัน ออนไลน์ (Quasi-Newton on-line algorithm) [10] ฯลฯ ดังนั้นการเรียนรู้ในแบบที่เราต้องการจึงมีกรอบการเรียนรู้ที่แตกต่างจาก แนวความคิดเหล่านี้ ซึ่งไม่จำกัดเหตุการณ์ในการเรียนรู้ รวมถึงองค์ความรู้ที่ได้นั้นจะต้องสามารถเพิ่มขึ้นได้เรื่อยๆ อย่างไรก็ตาม เรายังคงอาศัยหลักการของการเรียนรู้แบบเชื่อมตรงที่มีลักษณะคล้ายกัน นั้นคือ การลดความคลาดเคลื่อนให้น้อยที่สุด ณ ช่วงเวลาใดๆ ระหว่างการเรียนรู้ โดยที่ผลโดยรวมของความคลาดเคลื่อนนั้นจะต้องลู่เข้าสู่ค่าที่น้อยที่สุด ตามสมการทางคณิตศาสตร์ดังนี้ ซึ่งเสนอขึ้นในปี 1971 -1973 [10]

$$J(w) = \mathsf{E}_z Q(z,w) = \int Q(z,w)dP(z)$$

โดยที่ *J(w)* คือฟังก์ชันที่มีชื่อว่า ฟังก์ชันคาดความเสี่ยง (expected risk function) ซึ่งเป้าหมายของการเรียนรู้แบบเชื่อมตรงคือ การทำให้ *J*(*w*) มีค่าน้อยที่สุด ในส่วนของ *Q*(*z*,*w*) จะหมายถึงฟังก์ชันค่าความสูญหาย (loss function) ซึ่งเป็นตัววัดความคลาดเคลื่อนที่ไม่ตรงกับความเป็นจริงของพารามิเตอร์ *w* ภายใต้สถานการณ์ *z*

2.4 การเรียนรู้ด้วยการอุปนัย

การเรียนรู้ด้วยการอุปนัย หมายถึง การเรียนรู้จากข้อมูลตัวอย่างเพื่อให้ได้มาซึ่งกฎเกณฑ์ทั่วไปของชุดข้อมูลตัวอย่างนั้น [5] ตัวอย่างเช่น การเรียนรู้จากชุดข้อมูลตัวอย่าง 011, 000011, 00111, 0001, 0011 อาจจะสรุปได้ว่ากฎเกณฑ์ที่อธิบายสภาพข้อมูลนี้คือ สายอักขระที่เริ่มต้นด้วย ‘0’ ไม่จำกัดจำนวน ตามด้วย ‘1’ ไม่จำกัดจำนวน ซึ่งเขียนแทนได้ด้วย 0*1* ซึ่งคำจำกัดความของการเรียนรู้ด้วยการอุปนัยนี้ อาจจะไม่เหมือนคำว่า การเรียนรู้ ในความหมายทั่วไปตรงที่ การเรียนรู้ด้วยการอุปนัยจะสนใจแต่ในกรณีของการได้มาซึ่งกฎเกณฑ์ทั่วไปของข้อมูล โดยที่ยังไม่ได้คำนึงถึงคำตอบว่าสิ่งใดอยู่หรือไม่อยู่ในกลุ่มข้อมูลตัวอย่าง แต่การเรียนรู้ในความหมายทั่วไป จะ

สนใจในแง่ของการตอบคำถามว่าใช่ หรือ ไม่ มากกว่ากฎเกณฑ์ที่อธิบายสภาพของข้อมูล ในการเรียนรู้ด้วยอุปนัยนั้นจะระบุปัญหาของการเรียนรู้ด้วยการอุปนัยดังนี้ [4, 5]

1) *คลาสของฟังก์ชันหรือคลาสของภาษา* ที่จะพิจารณา ตัวอย่างเช่น กฎเกณฑ์ที่ได้จากการเรียนรู้อยู่ในคลาสของภาษาสโทแคสติก

2) *สมมติฐานที่เป็นไปได้* (hypothesis space) หมายถึง ขอบเขตของสมมติฐานที่ได้จากการเรียนรู้ที่เป็นไปได้ทั้งหมด ตัวอย่างเช่น ในการออกแบบการเรียนรู้ต้องคำนึงว่า การเปลี่ยนสมมติฐานเป็นไปได้ทั้งหมดกี่กรณี

3) *กฎเกณฑ์ ที่ยอมรับได้ในการแสดงออก* (Admissible presentation) ยกตัวอย่างได้โดย สมมติฐานแต่ละสมมติฐานจะมีการยอมรับว่า สายอักขระแบบใดที่ยอมรับได้ว่ายังอยู่ในสมมติฐานที่ถูกต้อง และสายอักขระแบบใดที่ไม่อยู่ในสมมติฐานที่ถูกต้อง

4) *อัลกอริทึม* ที่ใช้ในการเรียนรู้

5) *ข้อกำหนดขอบเขตความสำเร็จของการเรียนรู้* (Criterion of Success) เช่น การยอมรับว่าการเรียนรู้สำเร็จเมื่อพบว่า ค่าความน่าจะเป็นที่คำนวณได้ คลาดเคลื่อนไม่เกินค่าที่ตั้งไว้

ด้วยหลักการการเรียนรู้ด้วยการอุปนัยนี้ ทำให้เกิดการวางกรอบงานในการออกแบบการเรียนรู้ตามบทที่ 1.6 ซึ่งการกำหนดขอบเขตความสำเร็จนี้ ได้มีทฤษฎีการเรียนรู้ต่างๆ หลายทฤษฎีที่เสนอขอบเขตของความสำเร็จ ซึ่งจะได้กล่าวถึงบางทฤษฎีที่สำคัญต่อไป

2.5 ทฤษฎีการเรียนรู้ภาษา

ตามกรอบงานที่วางไว้ขั้นตอนแรกคือ การเลือกทฤษฎีที่เหมาะสมต่อสภาพปัญหา ซึ่งปัญหาในที่นี้คือ การเรียนรู้ภาษาด้วยวิธีการเชื่อมตรง โดยเราได้ให้คำจำกัดความคำว่าเชื่อมตรง (on-line) เอาไว้ในบทก่อนหน้า ในที่นี้คือการหยิบเอาเหตุการณ์ที่เกิดขึ้นแต่ละครั้งมาทำการเรียนรู้ได้ทันที ถ้าเปรียบเทียบข้อมูลที่ใช้ในการเรียนรู้เป็นสายอักขระ การเรียนรู้แบบเชื่อมตรงคือ การหยิบอักขระแต่ละตัวในสายอักขระใดๆ มาทีละหนึ่งตัวตามลำดับ แล้วทำการเรียนรู้เพื่อให้ได้ผลคือองค์ความรู้ที่เป็นแบบจำลองในทันทีที่ประมวลผลเสร็จแต่ละตัว โดยที่โครงสร้างของแบบจำลองจะขึ้นอยู่กับสายอักขระที่นำมาเรียนรู้ และไม่ได้ถูกกำหนดตายตัวก่อนหน้าการเรียนรู้เช่นเดียวกันกับการเรียนรู้แบบออฟไลน์ สิ่งแรกที่จะต้องคำนึงถึงในการออกแบบการเรียนรู้คือ ปัญหาที่ว่าเรียนรู้แบบใดและเมื่อใดจึงเรียกว่าเรียนรู้สำเร็จ และที่สำคัญมากไปกว่านั้นคือ อะไรคือการเรียนรู้ที่มีประสิทธิภาพ

การเรียนรู้ของเครื่องจักรในที่นี้ คือ กระบวนการหาฟังก์ชันที่เหมาะสมที่สุดด้วยตัวเครื่องจักรที่อธิบายสภาพข้อมูลที่นำมาเรียนรู้ได้ ดังนั้นปัญหาที่ว่า การเรียนรู้แบบใดและเมื่อใดจึงจะเรียกว่าเรียนรู้สำเร็จ จึงกล่าวได้ว่าการเรียนรู้จะสำเร็จก็ต่อเมื่อเราพบฟังก์ชันที่ครอบคลุมข้อมูลทั้งหมดที่นำมาเรียนรู้ ยกตัวอย่างการเรียนรู้ที่สำเร็จอย่างสมบูรณ์และมีประสิทธิภาพเช่นการค้นพบ กฎแรงโน้มถ่วงของนิวตัน เป็นต้น ซึ่งสามารถครอบคลุมความจริงเกี่ยวกับแรงโน้มถ่วงโดยไม่ขึ้นอยู่กับกาลเวลาและกรอบของวัตถุที่เราต้องการพิจารณา แต่อย่างไรก็ตามนั่นคือการเรียนรู้จากมนุษย์ ในส่วนของการเรียนรู้ด้วยเครื่องจักรนั้น อาจจะไม่สามารถหาฟังก์ชันที่ครอบคลุมข้อมูลทั้งหมดที่มีอยู่จริง จึงมีทฤษฎีในการเรียนรู้ขึ้นมาเพื่อ กำหนดกรอบที่แคบลงในการยอมรับความถูกต้องของการหาฟังก์ชั่นเหล่านั้น ในที่นี้ทฤษฎีที่น่าสนใจมีอยู่ 2 ทฤษฎีคือ

1) การเรียนรู้ด้วยการประมาณความถูกต้องโดยความน่าจะเป็น (probabilistic approximately correcting: PAC)

2) การจำแนกภายในจำกัด (identification in the limit)

การเรียนรู้ด้วยการประมาณความถูกต้องด้วยความน่าจะเป็น หรือเรียกอีกอย่างหนึ่งว่าพีเอซี คือ หลักการเรียนรู้โดยการคำนวณเพื่อหาค่าความน่าจะเป็นให้ตรงตามข้อมูลตัวอย่าง โดยมีหลักเกณฑ์ว่า สามารถยอมรับค่าความน่าจะเป็นที่มีความคลาดเคลื่อนระดับหนึ่งได้ ถ้าความคลาดเคลื่อนนั้นไม่เกินค่าใดค่าหนึ่ง ซึ่งค่านี้อาจจะถูกกำหนดเอาไว้ล่วงหน้า และเมื่อใดก็ตามที่มีดำเนินการเรียนรู้จนได้ความคลาดเคลื่อนไม่มากไปกว่าที่กำหนดแล้ว เราสามารถยอมรับได้ว่าการเรียนรู้สำเร็จ นอกจากนี้ยังมีข้อจำกัดคือเวลาในการเรียนรู้และขนาดของแบบจำลองจะต้องไม่เติบโตเกินฟังก์ชั่นพหุนาม จากกระบวนการของการเรียนรู้เช่นนี้ จะเห็นได้ว่าด้วยการใช้การคำนวณความน่าจะเป็นเพื่อเป็นหลักเกณฑ์ในการสร้างโครงสร้างของแบบจำลองนั้น จะไม่เหมาะต่อการหาคำตอบในการเรียนรู้ได้ด้วยวิธีการเชื่อมตรง เนื่องจากการคำนวณความน่าจะเป็นนั้นต้องคำนวณจากข้อมูลทั้งหมดในครั้งเดียว ถึงกระนั้นก็ตามหากนับเอาความสำเร็จเฉพาะครั้งในการรับข้อมูลปริมาณน้อยแบบหลักการเชื่อมตรง ก็จะเกิดปัญหาที่ว่า ความคลาดเคลื่อนเล็กน้อยอาจจะเป็นความคลาดเคลื่อนมากในอนาคต ดังนั้นทฤษฎีนี้จึงไม่เหมาะสมต่อการเรียนรู้ภาษาด้วยวิธีการเชื่อมตรงตามที่ต้องการ

ทฤษฎีการเรียนรู้อีกอย่างหนึ่งที่น่าสนใจคือ หลักการจำแนกภายในจำกัด [4] ทฤษฎีนี้มีหลักการที่ว่า หากตั้งสมมติฐานขึ้น ซึ่งในที่นี้คือการสร้างฟังก์ชันขึ้นมาจากการเรียนรู้แล้ว พบว่าเมื่อใดก็ตามที่ฟังก์ชันนั้นยังไม่ครอบคลุมต่อข้อมูลทั้งหมดแล้ว เราจะสามารถเปลี่ยนฟังก์ชันได้เพื่อที่จะหาฟังก์ชันที่ครอบคลุมมากกว่า จนกว่าจะพบฟังก์ชันที่มีคำตอบต่อข้อมูลทั้งหมดโดยไม่ต้องเปลี่ยนแปลงแล้ว จึงจะสามารถยอมรับได้ว่าสมมติฐานหรือฟังก์ชันนั้นถูกต้อง โดยจำนวนครั้ง

ในการเปลี่ยนสมมติฐานนั้นจะต้องรับประกันได้ว่ามีจำนวนจำกัด ด้วยลักษณะของหลักเกณฑ์ของความถูกต้องนั้นไม่เข้มงวด ซึ่งเหมาะสมกับการเรียนรู้แบบเชื่อมตรงในกรณีที่มีการยอมรับการเปลี่ยนสมมติฐานได้ไม่จำกัด อีกทั้งในงานวิจัยของ โกลด์ ผู้ซึ่งเสนอหลักการเรียนรู้ชนิดนี้นั้นมุ่งเน้นที่การเรียนรู้ภาษามนุษย์ โดย โกลด์ เปรียบเทียบว่าการเรียนรู้ภาษามนุษย์ด้วยคอมพิวเตอร์ที่ดี น่าจะเหมือนกับเด็กที่พยายามเรียนรู้ภาษาพูดจากผู้ใหญ่ซึ่งไม่จำเป็นต้องมีถูกผิด [7] แต่ระยะเวลาและการรับข้อมูลเข้ามาเรื่อยๆนั่นเองจะเป็นตัวที่จะทำให้เด็กเข้าใจในภาษา และเลือกที่จะใช้ตามแบบฉบับที่เรียนมา จากลักษณะการเรียนรู้ที่เราต้องการนั้นคล้ายคลึงกับหลักการของ โกลด์ มากและด้วยเหตุผลนี้เองงานวิจัยนี้จึงเลือกใช้ กระบวนทัศน์ในการเรียนรู้แบบของ โกลด์ เป็นหลักการในการกำหนดขอบเขตความสำเร็จซึ่งจะได้กล่าวรายละเอียดต่อไป

2.5.1 **หลักการจำแนกภาษาภายในจำกัด** (language identification in the limit)

ตัวอย่างมุมมองของการเรียนรู้ด้วยหลักการจำแนกภายในจำกัด จากเกมการเดาว่าตัวเลขต่อจากลำดับ 1, 3, 5… ควรจะเป็นอะไร จากการเรียนรู้ด้วย 1, 3, 5 อาจจะเดาโดยตั้งสมมติฐานแรกว่าจะต้องเป็นชุดของเลขคี่หรือ 2n+1 ดังนั้นเมื่อตัวต่อไปเป็นเลข 7 เข้ามา สรุปว่าฟังก์ชันเลขคี่ยังคงถูกต้องและครอบคลุมอยู่ จากนั้นเมื่อตัวเลขต่อไปเข้ามาเป็น 11 ปรากฏว่าฟังก์ชันเลขคี่ไม่สามารถตอบได้ จึงจำเป็นต้องเปลี่ยนสมมติฐานด้วยการเดาเป็นเลขของจำนวนเฉพาะ ผลปรากฏออกมาว่าเลขที่เข้ามาใหม่ตัวต่อไปทั้งหมดเป็นลำดับดังนี้คือ 13, 17, 19, 23 ซึ่งถูกต้องตรงกับเลขของจำนวนเฉพาะทั้งหมดดังนั้นจึงกล่าวได้ว่า ฟังก์ชันเลขจำนวนเฉพาะคือฟังก์ชันหรือสมมติฐานการเรียนรู้ที่ถูกต้อง ตราบใดที่ยังไม่มีตัวเลขชุดใหม่ที่ทำให้สมมติฐานไม่เป็นจริง สรุปว่ากระบวนทัศน์การเรียนรู้ด้วยการจำแนกภายในจำกัด หมายถึง การกำหนดขอบเขตความสำเร็จของการเรียนรู้ด้วยหลักการว่า เมื่อมีการเรียนรู้ไปได้ระยะหนึ่งสมมติฐานที่ถูกต้องนั้นจะไม่เปลี่ยนแปลงและจำนวนครั้งในการเปลี่ยนสมมติฐานนั้นรับประกันได้ว่ามีจำกัด กำหนดให้ G_t แทนฟังก์ชันที่ได้จากการเรียนรู้เมื่อรับอักขระลำดับตัวที่ t จากสายอักขระใดๆแล้วเรายอมรับว่าการเรียนรู้สำเร็จ ณ ลำดับที่ t เมื่อพบว่า

$$G_t = G_{t+1} = G_{t+2} = G_{t+3} = G_{t+4} = \ldots$$

จากการศึกษาด้วยทฤษฎีการคำนวณนั้นพบว่า ในกรณีที่เรียนรู้จากสายอักขระใดๆที่ไม่จำกัดความยาวนั้น ภาษานั้นอาจจะเป็นภาษาที่อยู่ในคลาสที่ไม่อาจจะหาเครื่องจักรมายอมรับได้ และถึงแม้หาเครื่องจักรมายอมรับได้ก็ตาม แต่เราอาจจะไม่สามารถหาฟังก์ชันนั้นได้พบ อีกทั้งจุดประสงค์ของการเรียนรู้คือนำเอาองค์ความรู้ที่ได้มาทำนายหรือจดจำเหตุการณ์ต่างเพื่อแยกว่าข้อมูลกลุ่มใดใช่หรือไม่ใช่กลุ่มเดียวกับชุดข้อมูลที่นำมาเรียนรู้ แต่เนื่องจากที่กล่าวไปว่า เราไม่

อาจจะหาฟังก์ชันที่ครอบคลุมภาษาทั้งหมดได้ เพื่อความจำเป็นในการทำนายข้อมูล ทางเลือกหนึ่งคือการนำความน่าจะเป็นเข้ามาช่วยในการทำนายโดยที่เราอาจจะไม่ต้องยอมรับสายอักขระทั้งหมด แต่เราจะเลือกยอมรับเฉพาะสายอักขระที่เกิดขึ้นซ้ำๆกันโดยอาศัยความน่าจะเป็นในการคัดเลือกสาย ในวิธีการเรียนรู้โดยใช้หลักการจำแนกภายในจำกัดดังกล่าวนี้จึงมีความยืดหยุ่นสูงและเหมาะต่อสภาพการเรียนรู้ชุดข้อมูลสายอักขระแบบเชื่อมตรงด้วยลักษณะของตัวทฤษฎีเอง ปัญหาก็คือเราจะมีวิธีเลือกหลักการในการคัดเลือกสายอักขระที่ซ้ำกันเหล่านั้นอย่างไรเพื่อให้มีประสิทธิภาพ ซึ่งเราจะกล่าวถึงเรื่องประสิทธิภาพในบทที่ 3 เพื่อเป็นหลักเกณฑ์ในการเทียบประสิทธิภาพ

2.6 เอกสารและงานวิจัยที่เกี่ยวข้อง

2.6.1 Language Identification in the limit

งานวิจัยของ โกลด์ [4] เสนอในปี 1968 เป็นการวิจัยความเป็นไปได้ในการเรียนรู้ภาษาและเสนอกระบวนทัศน์ในการเรียนรู้ภาษา โดยได้มีการจำแนกคลาสของภาษาที่เป็นไปได้ทั้งหมดพร้อมทั้งวิธีการในการนำเสนอข้อมูลตัวอย่างที่เป็นไปตามกำหนดต่อตัวการเรียนรู้ (learner) เพื่อนำข้อมูลเหล่านั้นมาเรียนรู้ ซึ่งจากการวิเคราะห์ด้วยทฤษฎีทางภาษานี้เองทำให้เกิดคำถามที่ว่าการนำเสนอข้อมูลตัวอย่างตามที่กำหนดนั้นเพียงพอหรือไม่ที่จะกำหนดการเรียนรู้ได้ว่า ภาษานั้นเป็นภาษาที่เราต้องการได้ เนื่องจากว่าการนำเสนอข้อมูลตัวอย่างนั้นสามารถกระทำได้หลายวิธีด้วยการควอนไทซ์เวลา (time quantization) เป็นช่วงๆที่จำกัด ดังนั้นจึงสามารถควอนไทซ์ได้หลายขนาดตามที่ต้องการ การเรียนรู้ด้วยกระบวนทัศน์นี้จะคำนึงว่าในแต่ละช่วงเวลาที่รับข้อมูลเข้ามานั้นตัวการเรียนรู้จะทำการเดาภาษาโดยอาศัยพื้นฐานจากข้อมูลที่เคยเรียนรู้มาก่อนหน้า ซึ่งกระบวนการนี้สามารถดำเนินการได้อย่างต่อเนื่องตลอด โดยมีคำจำกัดความที่ว่า ตัวการสามารถเรียนรู้ตามคลาสภาษาที่ต้องการได้ก็ต่อเมื่อ มีอัลกอริทึมที่มีความสามารถทำหน้าที่เป็นตัวเดาซึ่งเมื่อดำเนินการไประยะหนึ่งแล้วพบว่า ไม่มีการเปลี่ยนแปลงของสมมติฐานอีกแสดงว่า สมมติฐานนั้นครอบคลุม

2.6.2 Inductive Inference : Theory and Methods

จากงานวิจัยของ โกลด์ ที่ได้นำเสนอทฤษฎีในการเรียนรู้ภาษาออกมา ได้มีงานวิจัยต่อเนื่องมาจำนวนมากทั้งทางด้านทฤษฎีและการทดลองของระบบการเรียนรู้ด้วยการอุปนัย ซึ่งมีความพยายามในการค้นหากฎเกณฑ์ทั่วไปจากข้อมูลตัวอย่างด้วยการเรียนรู้ชนิดนี้ งานวิจัยชิ้นนี้ซึ่งเป็นงานวิจัยของ แองกลูอิน [5] เสนอในปี 1983 เป็นการรวบรวมเกี่ยวกับงานวิจัยที่เด่น รวมถึง

สรุปวิธีการในการเรียนรู้ด้วยกระบวนทัศน์แบบ โกลด์ โดยเน้นความสัมพันธ์ระหว่าง ทฤษฎีที่เกี่ยวข้อง อัลกอริทึม และ การนำเอาไปใช้งานจริง

2.6.3 Learning Deterministic Regular Grammars from stochastic samples in polynomial time

จากงานวิจัยของ คาราสโก [7] เสนอในปี 1999 ได้เสนอวิธีการจำแนกภาษาสโทแคสติกโดยอิงกระบวนทัศน์การเรียนรู้จากงานวิจัยของโกลด์ โดยที่ได้นำเสนอคลาสของอัลกอริทึมซึ่งเป็นอัลกอริทึมที่อุปนัยโครงสร้างของออโตมาตาที่มีขนาดเล็กที่สุด ซึ่งสามารถสังเคราะห์ภาษาตรงกับที่เรียนรู้ได้ โดยอาศัยหลักการผสานสถานะระหว่างสถานะที่มีความน่าจะเป็นที่เหมือนกันในการส่งผ่านทุกเหตุการณ์เข้าด้วยกัน นอกจากนี้งานวิจัยนี้ได้แสดงให้เห็นว่าเวลาที่ใช้ในการเรียนรู้นั้นสามารถดำเนินการได้รวดเร็วภายในกำหนดเวลาเชิงเส้นเทียบกับขนาดของตัวอย่างที่ใช้ในการเรียนรู้

2.6.4 Probabilistic Finite-State Machine – Part I-II

งานวิจัยได้นำเสนอโดย ไวดัล [1, 2] ปี 2004 เป็นบทความซึ่งมี 2 ส่วน โดยมีประเด็นหลักคือ การรวบรวมและสรุปงานวิจัยที่เกี่ยวกับ เครื่องจักรแบบจำกัดสถานะเชิงน่าจะเป็น (probabilistic finite state machine) ที่ใช้ในกระบวนการจดจำรูปแบบข้อมูล ในส่วนแรกจะเป็นบทความที่สำรวจและศึกษาเกี่ยวกับคุณสมบัติ และทฤษฎีที่สนับสนุนเครื่องจักรดังกล่าว ในส่วนที่สองจะเป็นการศึกษาและสำรวจงานวิจัยเกี่ยวกับตัวอย่างเครื่องจักรดังกล่าวที่มีอยู่ในปัจจุบันและเครื่องจักรอื่นๆที่สามารถสังเคราะห์สายอักขระได้ซึ่ง ได้แก่ แบบจำลองมาร์คอฟเชิงซ่อน แบบจำลองเอ็นแกรม แบบจำลองออโตมาตาเชิงน่าจะเป็น รวมถึงทฤษฎีและอัลกอริทึมในการสร้างแบบจำลองดังกล่าวในปัจจุบัน

2.6.5 On-line Learning and stochastic approximation

บทความนี้ได้นำเสนอโดย โบทู [10] ซึ่งเสนอเป็นบทความที่สรุปหลักเกณฑ์ในการเรียนรู้แบบเชื่อมตรง โดยที่ได้ทำการวิเคราะห์อัลกอริทึมต่างๆที่มีการลู่เข้าสู่ค่าความถูกต้อง โดยอาศัยทฤษฎีทางสโทแคสติก และพิสูจน์ภายใต้เงื่อนไขที่อ่อนมาก (weak condition) โดยมีการสรุปกรอบงานในการเรียนรู้แบบเชื่อมตรง ซึ่งกรอบงานนี้ได้ถูกนำมาใช้ในการเรียนรู้แบบเชื่อมตรงในปัจจุบันแทบทั้งสิ้น

2.6.6 The Power of Amnesia

งานวิจัยนี้โดย รอน [4] ในปี 1993 ซึ่งนำเสนอวิธีการเรียนรู้ทางสถิติซึ่งอาศัยหลักการของมาร์คอฟแบบใช้หน่วยความจำแปรผัน โดยนำเอาแบบจำลองต้นไม้คำเติมหลัง (suffix tree) มาช่วยในการเรียนรู้ เรียกแบบจำลองชนิดนี้ว่าต้นไม้เชิงน่าจะเป็นอิงคำเติมท้าย (prediction suffix tree: PST) จากงานวิจัยนี้ได้กล่าวถึงแบบจำลองมาร์คอฟแบบใช้หน่วยความจำที่คงที่ เช่น ลูกโซ่มาร์คอฟ ซึ่งมีเวลาในการสอนค่อนข้างช้าโดยมีฟังก์ชันเวลาในการสอนเติบโตแบบเอกโปเนเชียล อีกทั้งการจัดเก็บข้อมูลตามลำดับนั้นซ้ำซ้อน ด้วยเหตุนี้ในงานวิจัยจึงได้นำเสนอวิธีการมาร์คอฟแบบใช้หน่วยความจำแปรผันโดยใช้โครงสร้างต้นไม้ ซึ่งแนวความคิดคือ ความน่าจะเป็นในการเกิดอักขระ σ ภายใต้เงื่อนไขสายอักขระ s ใดๆจะมีค่าไม่เปลี่ยนแปลงเมื่อ s มีความยาวค่าหนึ่ง

ข้อดี

1. ฟังก์ชันการเติบโตเป็นแบบพหุนาม ซึ่งดีกว่าแบบหน่วยความจำคงที่ของแบบจำลองลูกโซ่มาร์คอฟซึ่งเติบโตแบบเอกโปเนนเชียล
2. สามารถใช้ข้อมูลตัวอย่างแบบไม่จำกัดความยาวทั้งสายอักขระเดียวต่อเนื่องหรือสายอักขระหลายสาย
3. หน่วยความจำที่เก็บข้อมูลไม่ซ้ำซ้อนเหมือนลูกโซ่มาร์คอฟแบบหน่วยความจำคงที่

งานวิจัยนี้มีคุณลักษณะความอัตโนมัติในการสอนโดยไม่ต้องคำนึงถึงรูปแบบของสายอักขระ และผลที่ได้ในการใช้งานจริงมีความสามารถสูงเทียบเท่ามาร์คอฟเชิงซ่อน (HMM)

2.6.8 Learning Probabilistic Automata with Variable Memory Length

ในปี 1996 คาราสโก ได้เสนองานวิจัยเพิ่มเติม [11] ต่อจากงานวิจัยเดิม[12] โดยได้นำเสนอวิธีการสร้างออโตมาตาจาก พีเอสที เนื่องจากพีเอสทีมีปัญหาก็คือในขั้นตอนการคำนวณความน่าจะเป็นของสายอักขระใดๆตอนใช้งานจริงจะช้าเนื่องจากต้องหาบัพที่มีชื่อตามสายอักขระที่เราต้องการคำนวณ ซึ่งกระบวนการนี้ไม่เป็นการส่งผ่านสถานะ ดังนั้นงานวิจัยนี้จึงปรับปรุงความสามารถด้วยการเสนอทฤษฎีที่ว่าพีเอสที จะมี ออโตมาตาความน่าจะเป็นที่เทียบเท่ากันโดยเสนอชนิดย่อยของออโตมาตาชนิดกำหนดได้เรียกว่า ออโตมาตาความน่าจะเป็นแบบใช้คำเติมท้าย(Probabilistic Finite Suffix Automata: PFSA) ซึ่งในการสร้างพีเอสพีเอนี้อาศัยการสร้าง PST เสียก่อนแล้วจึงนำเอาบัพต่างๆ ของ PST มาทำเป็นสถานะของ PFSA

ข้อดี ปรับปรุงการใช้งานจริงให้เป็นแบบการส่งผ่านสถานะ เพื่อความรวดเร็วในการคำนวณความน่าจะเป็นจากสายอักขระใดๆ

บทที่ 3

วิธีดำเนินการวิจัย

3.1 การวิเคราะห์และเลือกทฤษฎีการเรียนรู้

จากการศึกษาทฤษฎีการเรียนรู้ที่กล่าวไว้ในบทที่ 2 ผู้วิจัยได้เลือกทฤษฎีการเรียนรู้เพื่อกำหนดขอบเขตของความสำเร็จในการเรียนรู้ ซึ่งก็คือ ทฤษฎีการจำแนกภายในจำกัด ด้วยเหตุผลดังนี้คือ

3.1.1 การจำแนกภายในจำกัดมาจากแนวความคิดให้เครื่องจักรเรียนรู้ภาษามนุษย์ ซึ่งมีแนวความคิดเทียบกับการที่เด็กเรียนรู้ภาษา หมายความว่าการเรียนรู้ไม่มีขอบเขตของประโยค ไม่มีข้อมูลบอกว่าประโยคใดที่ถูกหรือผิด และสามารถเรียนรู้ได้ตลอดเวลา ซึ่งตรงกับแนวความคิดในการเรียนรู้ตามที่ต้องการว่า ไม่จำกัดความยาวสายอักขระ ไม่จำกัดจำนวนเหตุการณ์ และไม่จำกัดจำนวนครั้งในการเรียนรู้

3.1.2 มีความยืดหยุ่นของขอบเขตความสำเร็จสูง (Weak criterion) ซึ่งเหมาะสมกับการเรียนรู้แบบเชื่อมตรง ซึ่งต้องการความยืดหยุ่นอย่างมากในการกำหนดขอบเขตของความสำเร็จ เนื่องจากไม่สามารถทราบถึงข้อมูลที่ใช้ในการเรียนรู้ในระบบเชื่อมตรงได้ทั้งหมด จำเป็นต้องใช้หลักเกณฑ์ที่ยืดหยุ่นเพื่อให้เหมาะต่อสภาพข้อมูลในอนาคตที่ยังมาไม่ถึงได้

3.1.3 ทฤษฎีดังกล่าวเสนอขึ้นเพื่อการเรียนรู้ภาษาโดยที่ใช้ออโตมาตาเป็นเครื่องจักรที่ยอมรับภาษานั้นๆ ซึ่งสอดคล้องกับแนวความคิดที่ต้องการเรียนรู้ภาษา เพื่อให้ได้มาซึ่งแบบจำลองออโตมาตา

3.2 เกณฑ์เทียบประสิทธิภาพ

หลังจากที่ทำการเลือกทฤษฎีการเรียนรู้ด้วยการจำแนกภาษาโดยจำกัดแล้ว สิ่งที่จะต้องคำนึงถึงต่อไปก็คือ ประสิทธิภาพในการเรียนรู้ เนื่องจากการเรียนรู้ด้วยหลักการจำแนกภายในจำกัดนั้นไม่มีขอบเขตความถูกต้องที่แน่นอน เพียงแต่ยอมรับว่าสมมติฐานนั้นถูกต้องตราบใดที่สมมติฐานนั้นยังสามารถอธิบายสภาพข้อมูลได้อยู่ วิธีการในการเลือกสมมติฐานจึงสามารถมีได้หลายกรณี อย่างไรก็ตามการเลือกวิธีการตั้งสมมติฐานที่ดีต้องคำนึงถึงประสิทธิภาพต่างๆ ซึ่งหากมีการยอมรับสมมติฐานโดยไม่มีหลักเกณฑ์ประสิทธิภาพแล้ว ก็จะเป็นการสร้างสมมติฐานที่ไม่เกิดองค์ความรู้ แต่จะเป็นเพียงเสมือนฐานข้อมูลที่เก็บทุกอย่าง ทฤษฎีที่สนับสนุนหลักการในการคำนึงถึงประสิทธิภาพในการใช้งานนี้ได้ถูกเสนอขึ้นจากหลายๆงานวิจัย โดยที่สรุปไว้ใน [5] ซึ่งสรุปโดยย่อว่าประสิทธิภาพในการเรียนรู้นี้เราสามารถแบ่งออกเป็นสองแนวคือ

1) ประสิทธิภาพในการเรียนรู้เชิงเวลากับขนาดของแบบจำลอง
2) ประสิทธิภาพในการเรียนรู้เชิงความสามารถทางด้านข้อมูล

3.2.1 ประสิทธิภาพในการเรียนรู้เชิงเวลาและขนาดของแบบจำลอง

จากงานวิจัยของ [6] กล่าวว่า ลำพังเพียงการเรียนรู้ด้วยการจำแนกภายในจำกัดนั้นยังขาดประสิทธิภาพ จึงได้มีการนำเสนอหลักเกณฑ์เพื่อรับประกันประสิทธิภาพ โดยกล่าวว่าการเรียนรู้ใดๆนั้น นอกจากจะสามารถจำแนกเอกลักษณ์ได้แล้ว ยังจะต้องใช้เวลาในการเรียนรู้ไม่เกินฟังก์ชันพหุนาม รวมถึงอัตราการเติบโตของข้อมูลที่ได้จากการเรียนรู้จะต้องไม่เกินฟังก์ชันพหุนามเช่นกัน ซึ่งได้นิยามดังนี้

นิยามที่3.1 *แบบจำลองในการเรียนรู้มีประสิทธิภาพในการเรียนรู้เชิงเวลากับขนาดของแบบจำลองก็ต่อเมื่อเวลาในการเรียนรู้และขนาดของแบบจำลองที่ได้จากการเรียนรู้ จะต้องเติบโตไม่เกินฟังก์ชันพหุนาม*

เมื่อกำหนดให้ L แทนคลาสของภาษาที่ได้จากการเรียนรู้ด้วยแบบจำลอง M ให้ p( ) และ q( ) แทนฟังก์ชันพหุนามของเวลาในการเรียนรู้และการเติบโตของไวยากรณ์ และ A เป็นอัลกอริทึมที่ใช้เรียนรู้คลาสของภาษาดังกล่าวแล้ว

(1) ให้ S แทนข้อมูลตัวอย่างใดๆแล้ว จะต้องมี A ที่สามารถจะคืนค่าไวยากรณ์ g ในคลาสG โดยที่ $S \subseteq L_g$ ภายในระยะเวลา O(p(||S||))

(2) ให้ L_g เป็นภาษาใดๆ และ g คือไวยากรณ์ที่มีขนาด n แล้วโดยที่ $L_g(g)=L_g$ จะต้องมี C ที่แทนชุดข้อมูลตัวอย่างที่กำหนดขึ้นใหม่โดยที่ $C<q(n)$ เมื่อดำเนินการตาม A

ในกรณีที่ 2 อธิบายได้ว่าในระหว่างการเรียนรู้ดำเนินการด้วยอัลกอริทึม *A* หากดำเนินการจนไวยากรณ์มีขนาด *n* แล้วชุดข้อมูลตัวอย่างที่กำหนดขึ้นใหม่ในแต่ละรอบที่จะต้องพิจารณาระหว่างดำเนินการนั้นจะต้องน้อยกว่าฟังก์ชันพหุนาม ของ *n* ซึ่งเราจะเห็นภาพได้ชัดขึ้นในบทต่อไปเรื่องการออกแบบภาษา

3.2.2 ประสิทธิภาพในการเรียนรู้เชิงความสามารถทางด้านข้อมูล

นอกจากการคำนึงถึงเวลาและขนาดแล้วในงานวิจัยของ [4] ได้เสนอประสิทธิภาพเชิงความสามารถดังนี้

ความครอบคลุมของการเรียนรู้ (Scope) การครอบคลุมในการเรียนรู้นี้ถูกนิยามว่า ให้ความครอบคลุมของวิธีการ *M* คือ*คลาสของกฎเกณฑ์ที่ M ที่ใช้ในการเรียนรู้* ดังนั้น *M1* จะมีประสิทธิภาพครอบคลุมมากกว่า *M2* ก็ต่อเมื่อ *M1* มีกฎเกณฑ์ที่ใช้ในการเรียนรู้ครอบคลุม

มากกว่า *M2* ตัวอย่างเช่น *M1* มีกฎเกณฑ์ว่า อาศัยความน่าจะเป็นที่เท่ากันหรือความคลาดเคลื่อนไม่มากไปกว่าค่าคงตัวใดๆ ในการผสานสถานะ *M2* มีกฎเกณฑ์ว่าอาศัยความน่าจะเป็นที่เท่ากันระหว่างสถานะสองสถานะเท่านั้นในการผสานสถานะ จะเห็นว่ากฎเกณฑ์ของ *M1* นั้นครอบคลุมมากกว่า *M2* เนื่องจากสมาชิกในเซตตาม *M2* นั้นอยู่ในเซตของสมาชิกตาม *M1*

ประสิทธิภาพทางข้อมูล (Data Efficiency) ประสิทธิภาพทางข้อมูลนี้จะแบ่งตาม *ตำแหน่งที่ลู่เข้าสู่ค่าที่ถูกต้อง (convergent point)* สมมติฐานใดมีตำแหน่งลู่เข้าที่เร็วกว่าย่อมถือว่าดีกว่า ซึ่งคำว่าตำแหน่งการลู่เข้านี้อาจจะใช้เวลาเป็นตัววัดตำแหน่งหรือ ขนาดของข้อมูลที่ใช้ในการเรียนรู้เป็นตำแหน่งก็ได้ และอีกค่าหนึ่งคือ *ความกระชับของข้อมูล (strictly more data efficient)* ซึ่งหมายความว่าถ้าเรามีเซตภาษาหนึ่ง L_1 ซึ่งมีจำนวนสมาชิกที่น้อยกว่าแต่สามารถครอบคลุมภาษาได้เท่ากับเซตภาษา L_2 ที่มีจำนวนสมาชิกมากกว่า จากประสิทธิภาพทั้งสองตัวเรากล่าวได้ว่า L_1 มีประสิทธิภาพทางข้อมูลดีกว่า L_2 ก็ต่อเมื่อ L_1 มีความกระชับของข้อมูลที่ดีกว่า L_2 ในขณะที่มีตำแหน่งลู่เข้าที่ตำแหน่งเดียวกัน เราสามารถเปรียบเทียบความกระชับของข้อมูลเช่นเดียวกับการเข้ารหัสข้อมูลในปัจจุบัน

จำนวนครั้งของสมมติฐานที่ผิด (Number of Hypothesis) ในกรณีที่มีการเปลี่ยนแปลงสมมติฐาน จนกว่าจะพบสมมติฐานที่ถูกต้องซึ่งไม่เปลี่ยนแปลง นั้นคือตัววัดประสิทธิภาพอีกอย่างหนึ่ง ซึ่งวิธีการที่มีประสิทธิภาพที่ดีไม่ควรมีการเปลี่ยนแปลงสมมติฐานหลายครั้ง กล่าวได้ว่าวิธีการ *M1* ย่อมมีประสิทธิภาพดีกว่าวิธีการ *M2* หากว่าจำนวนครั้งในการเปลี่ยนสมมติฐานของ *M1* มีน้อยกว่า *M2* ณ ตำแหน่งที่ลู่เข้าสู่ค่าที่ถูกต้องตำแหน่งเดียวกัน

3.3 การออกแบบคลาสของภาษา

หลังจากที่เราได้เลือกกระบวนทัศน์การเรียนรู้เพื่อกำหนดขอบเขตความสำเร็จของการเรียนรู้แล้ว ประเด็นต่อไปคือเราจะพิจารณาว่า เราจะมีวิธีการเลือกสายอักขระสั้นๆอย่างไรเพื่อจะนำมานับความน่าจะเป็น ซึ่งมีผลต่อการทำนายและจดจำรูปแบบของข้อมูลให้ตรงต่อข้อมูลที่นำมาเรียนรู้ได้มากที่สุด เมื่อเราได้คำนึงถึงประสิทธิภาพตามที่ได้กล่าวไปแล้ว จะเห็นได้ว่าในการเรียนรู้แบบเชื่อมตรงนั้น ยังมีปัญหาด้านประสิทธิภาพที่จะใช้ความน่าจะเป็นเป็นหลักเกณฑ์ในการตั้งสมมติฐาน เนื่องจากว่าความน่าจะเป็นที่คำนวณได้นั้นจะมีการเปลี่ยนสมมติฐานอยู่ตลอด การระบุตำแหน่งที่ลู่เข้าสู่ค่าที่ถูกต้องได้อาจจะต้องใช้การคำนวณอย่างมาก หรือจำเป็นต้องใช้วิธีการที่สลับซับซ้อน อย่างไรก็ตามการจำแนกภาษาภายในจำกัดด้วยเฉพาะสายอักขระลำพังโดยไม่คำนึงถึงหลักความน่าจะเป็นก็ทำไม่ได้เช่นกัน เนื่องจาก ทฤษฎีของ โกลด์ ได้กล่าวไว้แล้วว่าหากภาษานั้นอยู่ในคลาสตั้งแต่ภาษาสม่ำเสมอ (Regular language) หรือสูงกว่านั้น ลำพังแต่ใช้สาย

อักขระ เราไม่อาจจะใช้เครื่องจักรในการหาไวยากรณ์ของภาษานั้นได้ภายในจำกัด จำเป็นต้องมีตัวช่วย (oracle) [4] ซึ่งในที่นี้คือความน่าจะเป็น จากงานวิจัยส่วนใหญ่พบว่า ในการออกแบบคลาสของภาษาที่ต้องการนั้น มักจะใช้ความน่าจะเป็นในการคัดเลือกสมาชิกในคลาสนั้น ตัวอย่างเช่น วิธีการผสานสถานะด้วยความน่าจะเป็น [6, 11, 13] ซึ่งเป็นวิธีการที่มีประสิทธิภาพและเหมาะสมที่สุด โดยที่อาศัยความน่าจะเป็นของการส่งผ่านที่ออกจากสถานะใดๆที่เหมือนกันมาผสานสถานะเข้าด้วยกัน

ด้วยเหตุผลเหล่านี้เองทำให้เราคิดหาว่า จะมีหลักเกณฑ์ใดหรือไม่ที่จะบ่งชี้ความน่าจะเป็นที่เหมือนกันของคู่สถานะใดๆ โดยไม่จำเป็นต้องนับความน่าจะเป็น จากการศึกษารวบรวมงานวิจัยต่างๆ เราได้แรงบันดาลใจและเห็นถึงประสิทธิภาพจากการเรียนรู้ด้วยแบบจำลองออโตมาตาความน่าจะเป็น [11, 12] ซึ่งมีการคำนวณแบบออฟไลน์ และสายอักขระที่จะนำมานับความน่าจะเป็นนั้นขึ้นอยู่กับลักษณะเด่น (Characteristic) ของข้อมูลเอง ซึ่งมีหลักการว่า ถ้าเราพบว่า

$$p(w|\, \boldsymbol{s_1}\, s_2\, s_3 \ldots\, s_n) \;=\; p(w|\, \boldsymbol{s_2}\, s_3 \ldots\, s_n)$$

โดยที่ s_n แทนอักขระใดๆ ลำดับที่ n ในสายอักขระและ $s_1 \ldots s_n$ และ w คือสายอักขระใดๆที่เกิดขึ้นจริงบนข้อมูลที่นำมาเรียนรู้แล้ว เรากล่าวได้ว่าความน่าจะเป็นในการเกิด w ภายใต้เงื่อนไข $\boldsymbol{s_2}s_3 \ldots s_n$ ก็เพียงพอแล้ว ซึ่งไม่จำเป็นต้องใช้เงื่อนไข $\boldsymbol{s_1}s_2s_3 \ldots s_n$ ดังนั้นจากประสิทธิภาพทางข้อมูลที่กล่าวว่าข้อมูลที่กระชับกว่าย่อมดีกว่าเราจึงเลือก $\boldsymbol{s_2}s_3 \ldots s_n$ ให้เป็นสมาชิกของภาษาที่เราออกแบบ ปัญหาต่อไปก็คือการคำนวณความน่าจะเป็นในการเกิด w ภายใต้เงื่อนไขใดๆเหล่านั้น ไม่สามารถคำนวณได้ในกระบวนการเชื่อมตรง เพราะเราไม่มีทางรู้ได้เลยว่าข้อมูลที่จะเกิดในอนาคตนั้นจะมีอะไรบ้าง ซึ่งอาจจะทำให้ความน่าจะเป็นในการเกิด w เปลี่ยนแปลงไปไม่ลู่เข้าสู่ค่าที่ถูกต้อง

จากหลักการเรียนรู้แบบเชื่อมตรง [10] นั้นเราสนใจในกระบวนการที่ว่าทำอย่างไรจึงจะลดความคลาดเคลื่อนให้น้อยที่สุด ณ เวลาที่ t ใดๆ หมายความว่าเราจะทำอย่างไรเพื่อให้ได้ G_t หรือฟังก์ชันการเรียนรู้เมื่อรับอักขระตัวที่ t เข้ามาแล้ว ฟังก์ชันนั้นสามารถคำนวณความน่าจะเป็นของเหตุการณ์ใดๆที่เกิดขึ้นจริงตามข้อมูลที่นำมาเรียนรู้นั้น ให้เกิดความคลาดเคลื่อนน้อยที่สุดเท่าที่เป็นไปได้ เราคำนึงถึงประเด็นที่จะตัดความคลาดเคลื่อนออกไปให้มากที่สุดโดยที่จะยังอาศัยหลักการลูกโซ่มาร์คอฟ และจากการพิจารณาจากวิธีการ ลูกโซ่มาร์คอฟเชิงความยาวแปรผัน เราได้สรุปข้อเท็จจริงดังนี้

ประพจน์ที่ 3.2 *กำหนดให้ w และ $\boldsymbol{s_1}\, s_2\, s_3 \ldots\, s_n \in \Sigma$ จะได้ว่า*

$$p(w|\, \boldsymbol{s_1}\, s_2\, s_3 \ldots\, s_n \;=\; p(w|\, \boldsymbol{s_2}\, s_3 \ldots\, s_n)$$

เป็นจริงเสมอในกรณีที่ สำหรับทุกสายอักขระ $\boldsymbol{s_2}\, s_3 \ldots\, s_n$ จะต้องมี $\boldsymbol{s_1}$ นำหน้าเสมอ

พิสูจน์ ให้ $p(w|s_1 s_2 s_3 \ldots s_n)$ มีค่าเท่ากับจำนวนครั้งในการเกิด $s_1 s_2 s_3 \ldots s_n w$ หารด้วยจำนวนครั้งที่เกิด $s_1 s_2 s_3 \ldots s_n$ และหากทุกครั้งที่เกิด $s_2 s_3 \ldots s_n$ มีเฉพาะ s_1 นำหน้าเสมอแล้ว จำนวนครั้งในการเกิด $s_2 s_3 \ldots s_n$ ย่อมเท่ากับจำนวนการเกิด $s_1 s_2 s_3 \ldots s_n$ ดังนั้นสามารถสรุปได้ว่า $p(w|s_1 s_2 s_3 \ldots s_n) = p(w|s_2 s_3 \ldots s_n)$ ∎

จากข้อเท็จจริงข้างต้นเราได้ทำการออกแบบภาษาที่ได้จากการเรียนรู้ดังนี้

3.3.1 แนวความคิดในการออกแบบคลาสของภาษาสโทแคสติก

ให้ L เป็นคลาสของภาษาที่เราต้องการออกแบบ และกำหนดให้ $L(M)$ เป็นภาษาที่เทียบเท่าภาษาที่สังเคราะห์ได้จากสโทแคสติกออโตมาตา M โดยที่ $L(M) \in L$ ซึ่งเป็นภาษาที่ได้จากการเรียนรู้ พิจารณาจากหลักการที่กล่าวมาตามประพจน์ที่ 3.2 เราได้สรุปข้อเท็จจริงดังนี้

ข้อเท็จจริงที่ 3.3 *กำหนดให้ x, u และ $y \in \Sigma^*$ และให้ $a \in \Sigma$ โดยที่ $x^{-1}L(M)$ และ $y^{-1}L(M)$ แทน ผลหารของภาษา $L(M)$ แล้ว*

3.3.1 $x^{-1}L(M) = y^{-1}L(M)$ *เป็นจริงถ้า* $ax \in suffix^*(y)$ *สำหรับ* $\forall y$ *ที่มี* $x \in suffix^*(y)$

3.3.2 $x^{-1}L(M) = y^{-1}L(M)$ *เป็นจริงถ้า* $p(u|x^{-1}L(M)) = p(u|x^{-1}L(M))$ *โดยที่*

$$p(u|x^{-1}L(M)) = \frac{p(xu|L(M))}{p(x\Sigma^*|L(M))}$$

จากข้อเท็จจริง 2 นี้เองทำให้ได้หลักการในการจำแนกภาษาภายในจำกัดด้วยความน่าจะเป็นที่ไม่ต้องนับความน่าจะเป็น นั่นคือเมื่อใดก็ตามที่เกิดเหตุการณ์ตามข้อเท็จจริง 3.3.1 แล้วเราสามารถสรุปได้ทันทีว่า $x^{-1}L(M) = y^{-1}L(M)$ จากข้อเท็จจริงดังกล่าวเราสรุปเป็นทฤษฎีดังนี้

ทฤษฎีบทที่ 3.4 *กำหนดให้ $L(M)$ เป็นสมมติฐานที่ได้จากการเรียนรู้จากสายอักขระใดๆแล้ว การนิยามสมาชิกในภาษา $L(M)$ ด้วยข้อเท็จจริง 3.3.1 สมมติฐานนั้นจะถูกต้องตามหลักการจำแนกภายในจำกัด*

พิสูจน์ สมมติฐานจะลู่เข้าสู่ค่าที่ถูกต้องนั้นก็ต่อเมื่อไม่มีการเปลี่ยนแปลงสมมติฐานอีก ตามหลักการจำแนกภายในจำกัด เนื่องจากข้อเท็จจริง 3.3.1 ยอมรับการเท่ากันของผลหารของภาษาเฉพาะในกรณีที่มีผลหารมีอักขระนำหน้าตัวเดียวเท่านั้น ดังนั้นแนวทางการเปลี่ยนแปลงสมมติฐานจะมีเพียง 2 กรณีคือ ใช่หรือไม่ ซึ่งแน่นอนว่าผลหารใดๆจะมีการเปลี่ยนแปลงสมมติฐานครั้งเดียวเท่านั้น เมื่อมีการเปลี่ยนแปลงสมมติฐานแล้วย่อมจะไม่มีการเปลี่ยนแปลงอีก ดังนั้นเราจึงสรุปว่า ภาษาที่ถูกนิยามจากข้อเท็จจริง 3.3.1 สามารถลู่เข้าสู่ค่าที่ถูกต้องได้ ∎

จากข้อเท็จจริง 2 นี้จะสามารถนำไปสู่การลดขนาดของภาษาให้น้อยลง รวมถึงการทำให้มีประสิทธิภาพตามหลักเกณฑ์ที่วางไว้ ดังข้อพิจารณาดังนี้

ประพจน์ที่ 3.5 *การจำแนกเอกลักษณ์ของสายอักขระย่อยใดๆ ด้วยข้อเท็จจริง 3.3.1 เป็นจริงเสมอในทุกๆ สายอักขระ*

พิสูจน์ กำหนดให้ b, a $\in \Sigma$ และ *x, k, s* $\in \Sigma^*$

<u>กรณีที่ 1</u> ให้สายอักขระย่อย *x* เป็นสายอักขระย่อยความยาวตั้งแต่ 2 อักขระขึ้นไปซึ่งปรากฏขึ้นหลายครั้งบนสายอักขระ *s* และทุกครั้งที่ *x* ปรากฏขึ้นพบว่ามี *ax* ทุกครั้งโดยไม่ปรากฏว่า มี *bx* เกิดขึ้นบนสายอักขระ *s* ซึ่ง a $\neq$ *b* เรากล่าวได้ว่าสายอักขระ *ax* ที่เป็นเอกลักษณ์ย่อมมีอยู่จริง เมื่อ *ax* ซึ่งเป็นเอกลักษณ์มีอยู่จริงแล้ว *x* ย่อมเป็นคำเติมท้ายที่ยาวที่สุดของ *ax* ที่มีอยู่จริง

<u>กรณีที่ 2</u> ให้ *u* เป็นสายอักขระย่อยที่ไม่เกิดขึ้นซ้ำเลยบนสายอักขระที่พิจารณาแล้ว *u* ย่อมเป็นเอกลักษณ์ เมื่อ *u* เป็นเอกลักษณ์แล้ว *u* ย่อมเป็นคำเติมท้ายที่ยาวที่สุดของ *ku* ที่มีอยู่จริงเท่านั้น ∎

ข้อพิจารณาต่อไปคือ ด้านประสิทธิภาพทางข้อมูลนั้น เรากล่าวได้ว่าการลู่เข้าสู่ค่าที่ถูกต้องของการพิจารณาแต่ข้อเท็จจริง 3.3.1 นั้นอาจจะมีตำแหน่งลู่เข้าที่มากกว่าหลักการใช้การนับความน่าจะเป็นตามข้อเท็จจริง 3.3.2 แต่ไม่เสมอไปทุกครั้งซึ่งกรณีที่ดีที่สุดคือลู่เข้า ณ ตำแหน่งเดียวกัน อย่างไรก็ตามเราสามารถยอมรับได้เนื่องจากประเด็นในการเรียนรู้ของเรานั้นไม่ได้พิจารณาแต่เรื่องของขนาดหรือระยะเวลาในการลู่เข้าของข้อมูลเพราะในที่นี้ประเด็นที่สำคัญที่สุดคือความสามารถในการเรียนรู้เชิงตลอดชีพ

ประพจน์ที่ 3.6 *ขนาดของสมมติฐานตามการพิจารณาด้วยข้อเท็จจริง 3.3.2 จะน้อยกว่าหรือเท่ากับขนาดสมมติฐานตามการพิจารณาด้วยข้อเท็จจริง 3.3.1*

พิสูจน์ ในกรณีที่ข้อเท็จจริง 3.3.2 เป็นจริงได้ 2 กรณีคือ กรณีแรก เมื่อคำนวณความน่าจะเป็นแล้ว $x^{-1}L(M) = y^{-1}L(M)$ กรณีที่สอง *x เป็น คำเติมท้ายของ y เท่านั้น* เมื่อคำนวณแล้วแน่นอนว่าต้องได้ $x^{-1}L(M) = y^{-1}L(M)$ ซึ่งในกรณีที่สองนี้ตรงกับเงื่อนไขในข้อเท็จจริง 3.3.1 เราจึงกล่าวได้ว่าการพิจารณาด้วยข้อเท็จจริง 3.3.1 ลู่เข้าสู่ค่าที่ถูกต้องได้ดีที่สุด ณ ตำแหน่งเดียวกับการพิจารณาด้วยข้อเท็จจริง 3.3.2 ∎

ในส่วนประสิทธิภาพด้านจำนวนครั้งของสมมติฐานที่ผิด เราพบว่าหากใช้หลักเกณฑ์ตามข้อเท็จจริง 3.3.2 แล้ว ในการเรียนรู้แบบเชื่อมตรงนั้นจะไม่สามารถมีหลักประกันได้ว่าข้อมูลจะลู่เข้าสู่ค่าที่ถูกต้องเนื่องจาก การนับความน่าจะเป็นที่เปลี่ยนแปลงตลอดเวลา ดังนั้นจากข้อเท็จจริง 3.3.1 นั้นสามารถรับประกันได้ว่าสมมติฐานจะลู่เข้าสู่ค่าที่ถูกต้องแน่นอน เนื่องจากพิจารณาเพียงแต่ตัวนำหน้าของสายอักขระที่พิจารณาเท่านั้น และหากพบว่าสายอักขระใดมีอักขระนำหน้า

มากกว่าหนึ่งตัวแล้วเป็นที่แน่นอนว่า สมมติฐานไม่กลับมามีหนึ่งตัวอีก ตามทฤษฎีบทที่ 3.4 ด้วยเหตุนี้ประสิทธิภาพข้อนี้จึงเป็นปัจจัยสำคัญที่สุด ที่ทำให้เราเลือกกฎเกณฑ์ตามข้อเท็จจริง 3.3.1 ในการจำแนกภาษา

3.3.3 คลาสของภาษาที่ออกแบบ

จากแนวความคิดข้างต้น เราได้ทำการกำหนดนิยามต่างๆ ในการออกแบบคลาสของภาษาสโทแคสติก การออกแบบคลาสของภาษาสโทแคสติกจะถูกแบ่งออกเป็น 2 ส่วนโดยที่ส่วนแรกจะใช้หลักการจำแนกภายในจำกัดตามข้อเท็จจริง 3.3.1 และในส่วนที่สองคือ การประมาณค่าของความน่าจะเป็นของภาษาสโทแคสติก เพื่อรับประกันว่าค่าที่ได้จากการประมาณนั้นจะลู่เข้าสู่ค่าที่ถูกต้องได้ ก่อนอื่นเรานิยามตัวสังเคราะห์ภาษาเพื่อสะดวกในการอธิบายดังนี้

ให้ $L(M)$ เป็นภาษาสโทแคสติกหนึ่งที่เทียบเท่าตัวสังเคราะห์ภาษา (Language generator) $M = (Q_M, \Sigma, \varepsilon, \tau_M, \gamma_M,)$ ตามนิยามดังนี้

$$Q_M = \{ x^{-1}L(M) \neq \{\} : x \in \Sigma^* \}$$

$$\tau_M(x^{-1}L(M),e) = xe^{-1}L(M)$$

$$\varepsilon = \varepsilon^{-1}L(M)$$

$$\frac{\gamma_M(x^{-1}{}_{L(M),e})}{\gamma_M(x^{-1}{}_{L(M),\#})} = p(e\Sigma^* \Big| x^{-1}{}_{L(M))}$$

จากหลักการตามข้อเท็จจริงข้อเท็จจริง 3.3.1 เนื่องจากเราต้องการภาษาที่ถูกเรียนรู้แบบเชื่อมตรง ภาษาที่ได้จากการเรียนรู้จะแปรเปลี่ยนไปตามเวลาต่างๆ ซึ่งเราได้นิยามภาษาที่เทียบเท่าตัวสังเคราะห์ภาษา M ณ เวลาที่ t แทนด้วย $L_t(M)$ โดยการอธิบายด้วยสโทแคสติกออโตมาตา M ณ เวลาที่ t ดังนี้

นิยามที่ 3.7 *ให้ L แทนภาษาสโทแคสติก กล่าวว่าภาษาสโทแคสติก ณ เวลาที่ t คือ ฟังก์ชันผลลัพธ์ที่ได้จากการเรียนรู้ที่ส่งคืนค่าความน่าจะเป็นของสายอักขระใดๆเมื่อการเรียนรู้ผ่านไปเป็นเวลา t แทนด้วยสัญลักษณ์* L_t

นิยามที่ 3.8 *กำหนดให้ x, y และ* $s \in \Sigma^*$ *และ* $e \in \Sigma$ *แล้ว กล่าวว่า ผลหารของภาษา* $(ye)^{-1}L_t$ *เทียบเท่ากับ* $(xe)^{-1}L_t$ *เขียนแทนด้วย* $(ye)^{-1}L_t \cong (xe)^{-1}L_t$ *ก็ต่อเมื่อ* $y = sax$ *และ* $ax \in suffix^*(y)$ *สำหรับ* $\forall y$ *ที่* $x \in suffix^*(y)$

นิยามที่ 3.9 *ให้ M คือออโตมาตาสโทแคสติกที่สังเคราะห์ภาษา L_t เขียนแทนด้วย $L_t(M)$ แล้ว เรานิยามฟังก์ชันการส่งผ่านในกรณีที่ $(ye)^{-1}L_t \cong (xe)^{-1}L_t$ ได้ว่า $\tau_M(y^{-1}L_t(M), e)=\tau_M(x^{-1}L_t(M), e)=(xe)^{-1}L_t$ เมื่อ $|y|>|x|$*

จากนิยามที่ 3.9 นี้เราสามารถผสานสถานะปลายทางของการส่งผ่านหลายตัวเข้าด้วยกัน เพื่อลดขนาดของออโตมาตา ณ เวลาการเรียนรู้ที่ *t* ปัญหาคือเมื่อเราพบว่า ณ เวลาที่ *t+k* เมื่อ *k* คือเวลาใดๆ และปรากฏว่ามี *bx* เกิดขึ้นโดยที่ b ≠a ซึ่งทำให้สมมติฐานที่ว่า *x* มี *a* เป็นอักขระนำหน้าตัวเดียวเท่านั้นไม่เป็นจริง เราจะมีวิธีการอย่างไรในการปรับเปลี่ยนสมมติฐาน L_{t+k} จากปัญหาดังกล่าวเราได้นิยามในการปรับเปลี่ยนสมมติฐานดังต่อไปนี้

นิยามที่ 3.10 *กำหนดให้ $b \in \Sigma$ และ $(bxe)^{-1}L_{t+k}$ คือผลหารภาษา ณ เวลา t+k ซึ่งทำให้เงื่อนไข ($ax \in suffix^*(y)$ สำหรับ $\forall y$ ที่ $x \in suffix^*(y)$) มีค่าเป็นเท็จ เราเรียก $(xe)^{-1}L_{t+k}(M)$ ว่า เป็นสถานะของออโตมาตาที่ต้องการเปลี่ยนสมมติฐาน*

นิยามที่ 3.11 *ให้ $(xe)^{-1}L_{t+k}(M)$ เป็นสถานะของออโตมาตาที่ต้องการเปลี่ยนสมมติฐาน เราจะเรียกเซตของสถานะที่ต้องพิจารณาในการเปลี่ยนสมมติฐาน $L_{t+k}(M)$ ว่า แคนดิเดตเซต C_n (Candidate set) โดยที่ C_n = { v โดยที่ $(ve)^{-1}L_{t+k}(M) = xe^{-1}L_{t+k}(M)$ } และ เรียก $y \in C_n$ ว่าเป็นแคนดิเดตที่สั้นที่สุด(minimize candidate: mc) เมื่อ $y = ax$ โดยที่ $a \in \Sigma$*

ในการเปลี่ยนแปลงสมมติฐานนั้น สิ่งที่เราคำนึงถึงคือการขยายความยาวสายอักขระ *x* ที่พิจารณาให้ยาวมากขึ้น เนื่องจากความยาวที่มากขึ้น จะทำให้สมมติฐานลู่เข้าสู่ความจริงมากขึ้น วิธีการเปลี่ยนสมมติฐานคือ ทำการสร้างสถานะใหม่ที่รองรับสายอักขระให้มีความยาวมากขึ้นไป โดยกำหนดให้เป็นสถานะปลายทางส่งผ่านจากสมาชิกในแคนดิเดตเซต โดยที่เราสามารถลดขนาดการสร้างสถานะใหม่ ด้วยการพิจารณาที่แคนดิเดตที่สั้นที่สุดว่ามีอักขระนำหน้าที่ต่างกันทั้งหมดกี่ตัว หากว่า อักขระนำหน้าของแคนดิเดตที่สั้นที่สุดมีเพียงตัวเดียวแล้วเราก็สามารถสร้างสถานะใหม่ที่มีอักขระนำหน้าเพียงตัวเดียวได้

นิยามที่ 3.12 *แคนดิเดตพรีฟิกเซต (candidate prefix set) CP หมายถึง {$a \in \Sigma$ โดยที่ $ay \in suffix^*(v)$ สำหรับ $\forall v \in C_n$ และ $y \in C_n$ เป็นแคนดิเดตที่สั้นที่สุด} แล้ว ให้ Lnew(M) คือ เซตของสถานะที่สร้างขึ้นใหม่โดยที่*

Lnew(M) = { $(ye)^{-1}L$} ในกรณีที่ $|CP| = 1$ เมื่อ $(xe)^{-1}L(M)$ เทียบเท่าสถานะของออโตมาตาที่ต้องการเปลี่ยนสมมติฐาน

Lnew(M) = { $(ve)^{-1}L$ โดยที่ $v \in C_n$ } ในกรณีที่ $|CP| > 1$

หลังจากที่เราทำการสร้างสถานะใหม่ขึ้นมาจากการพิจารณาจากแคนดิเดตเซต เพื่อเปลี่ยนสมมติฐานใหม่นั้นเราไม่สามารถย้อนกลับไปดูได้ว่า เมื่อเราเพิ่มความยาวของสายอักขระที่พิจารณาแล้ว สมาชิกตัวถัดไปในการส่งผ่านนั้นจะมีอะไรบ้างตามสถิติที่ผ่านมา ซึ่งแน่นอนว่าไม่มีทางที่จะมากไปกว่าตัวสมมติฐานตัวก่อนแต่อาจจะมีน้อยกว่า สิ่งที่สำคัญไม่ใช่สมาชิกตัวถัดไปว่ามีอะไรแต่เป็นการคำนวณความน่าจะเป็นของสมาชิกเหล่านั้น เนื่องจากในการประมาณค่าในการใช้งาน เรานิยมประมาณค่าของสมาชิกที่มีความน่าจะเป็นเป็นที่ไม่สามารถนิยามได้ นั่นคือสมาชิกที่ไม่มีความถี่ปรากฏมาก่อน โดยที่วิธีการแก้ปัญหานี้เรียกว่า *เทคนิคการปรับเรียบ*(Smoothing technique) เราจึงต้องอาศัยความน่าจะเป็นของสมาชิกเหล่านี้จากเงื่อนไขที่น้อยกว่า ซึ่งการประมาณค่าความถี่ของสมาชิกตัวถัดไปนี้เราจะกล่าวอีกทีในเรื่องของการประมาณค่าความถี่ *ดังนั้นสถานะใหม่ที่สร้างขึ้นนั้นจะต้องทำการโคลนจากสถานะที่ต้องการเปลี่ยนสมมติฐาน*

นิยามที่ 3.12 *เรียกสถานะ* $w^{-1}L_t(M)$ *ว่าเป็นสถานะที่ถูกทำการโคลน (cloning) จาก* $x^{-1}L_t(M)$ เขียนแทนด้วย $w^{-1}L_t(M) = clone(x^{-1}L_t(M))$ *ก็ต่อเมื่อ ทุก* $\tau_M(w^{-1}L_t(M), e)=\tau_M(x^{-1}L_t(M), e)$ *สำหรับ* $\forall e \in \Sigma$

นิยามที่ 3.13 โครงสร้างออโตมาตาภายหลังปรับเปลี่ยนสมมติฐานถูกนิยามดังนี้

$Q_{t+k} = Q_t \cup Lnew(M)$

$\tau_M(v^{-1}L_{t+k}(M), e) = w^{-1}L_{t+k}(M)$ โดยที่ $w \in Lnew(M)$ และ $v \in C_n$

$w^{-1}L_{t+k}(M) = clone(xe)^{-1}L_{t+k}(M))$ สำหรับ $\forall w \in Lnew(M)$ และ $(xe)^{-1}L_{t+k}(M))$ เป็นสถานะที่ต้องการเปลี่ยนสมมติฐาน

ทฤษฎีบทที่ 3.14 *ภาษา L(M)* ที่ถูกออกแบบตามนิยาม 3.7 ถึง 3.13 มีคุณสมบัติของการถูกจำแนกได้ภายในจำกัด

พิสูจน์ ให้ L_t เป็นสมมติฐานตัวที่ t ซึ่งตามหลักการจำแนกภายในจำกัดกล่าวว่า เมื่อพบฟังก์ชันสมมติฐานตัวที่ t แล้วสมมติฐานไม่มีการเปลี่ยนแปลงต่อไปนั่นคือ $L_t = L_{t+1}= L_{t+2}=...L_{t+n}$ สิ่งที่เราต้องพิสูจน์คือ วิธีการนี้สามารถรับประกันได้ว่า เมื่อ t มีค่าๆ หนึ่งแล้วจะไม่มีการเพิ่ม สถานะ $x^{-1}L$ ใดๆ อีกหากว่า L_t นั้นครอบคลุมข้อมูลทั้งหมดแล้ว จากประพจน์ที่ 3.5 เราบอกว่าทุกๆ สายอักขระสามารถจำแนกเอกลักษณ์ได้ตามหลักการตามข้อเท็จจริง 3.3.1 เสมอแล้ว จากภาษาที่เราออกแบบไว้ ตามข้อกำหนดที่ 3.8 ได้คำนึงถึงสมมติฐานที่มีหลักการตรงตามข้อเท็จจริง 3.3.1 ย่อมสามารถลู่เข้าสู่หลักการได้จริง และจากนิยามที่ 3.9 - 3.13 เราจะสร้างสถานะใหม่ก็ต่อเมื่อ

พบข้อมูลที่ไม่เป็นจริงตามนิยามสถานะที่ต้องการเปลี่ยนสมมติฐาน และภายหลังจากการสร้างสถานะใหม่พบว่า สมมติฐานก็ยังคงอาศัยหลักการตามข้อเท็จจริง 3.3.1 อยู่ซึ่งการเปลี่ยนสมมติฐานแต่ละครั้งจะทำให้สมมติฐานครอบคลุมต่อตัวข้อมูลมากขึ้น ตามเซตของสถานะและการส่งผ่านของออโตมาตาที่ขยายความครอบคลุมมากขึ้นตามนิยามที่ 3.13 ดังนั้นหากเราเปลี่ยนสมมติฐานจนครบและเป็นจริงเสมอโดยไม่มีการเปลี่ยนสมมติฐานแล้ว จะไม่มีการสร้างสถานะใหม่หรือ การส่งผ่านใหม่เกิดขึ้นอีก ตรงนี้เราจึงสรุปได้ว่า ภาษา *L(M)* ตามนิยามที่ 3.7 ถึง 3.13 มีคุณสมบัติของการถูกจำแนกได้ภายในจำกัด ∎

3.3.3 วิธีการประมาณความน่าจะเป็นของภาษาสโทแคสติก

หลังจากที่เรามีวิธีการระบุโครงสร้างของออโตมาตาแบบเชื่อมตรงแล้ว ปัญหาต่อไปคือเราจะมีวิธีการประมาณค่าของความน่าจะเป็นของการส่งผ่านในออโตมาตาควบคู่ไปพร้อมกับการระบุโครงสร้างออโตมาตาได้อย่างไร เพื่อให้การนับความน่าจะเป็นมีความคลาดเคลื่อนจากชุดข้อมูลตัวอย่างให้น้อยที่สุดในกรณีที่เกิดการเปลี่ยนแปลงสมมติฐาน วิธีการประมาณค่าความน่าจะเป็นตามสมมติฐานที่ไม่เปลี่ยนแปลงนั้น เราสามารถทำได้โดยการนับความถี่เพิ่มในการใช้การส่งผ่านแต่ละครั้งซึ่งปรากฏจริงในชุดข้อมูลตัวอย่าง ซึ่งสามารถนับความถี่ควบคู่กันไปกับการระบุโครงสร้าง แต่หากปรากฏว่ามีการเปลี่ยนแปลงสมมติฐานเกิดขึ้น ณ เวลาใดๆ แล้วการนับความถี่นั้นอาจจะต้องเปลี่ยนแปลง เนื่องจากวิธีการนับแบบสมมติฐานเดิมเปลี่ยนแปลงไป เพื่อลดความคลาดเคลื่อนให้น้อยที่สุดเมื่อมีการเปลี่ยนสมมติฐานใหม่เราได้นำเสนอหลักการดังนี้

การประมาณค่าความถี่

การประมาณความถี่ในที่นี้ เราเสนอหลักการอินเตอร์โพเลตอย่างง่ายเพื่อลดความคลาดเคลื่อนให้มากที่สุด จากการเปลี่ยนสมมติฐานตามข้อกำหนดที่ 3.13 เราสามารถประมาณความถี่ได้ว่า ให้ $ve \in Lnew(M)$ และ $(xe)^{-1}L_{t+k}(M))$ เป็นสถานะที่ต้องการเปลี่ยนสมมติฐาน ให้ $ve^{-1}L_{t+k}(M) = \text{clone}(xe)^{-1}L_{t+k}(M))$ และ $\forall a \in \Sigma$ เป็นเหตุการณ์ทุกเหตุการณ์ที่ได้จากการโคลนแล้วเราได้สมการประมาณความถี่ดังนี้

$$\gamma(ve^{-1}L_t(M),a) \cong \frac{\gamma(v^{-1}L_t(M),e) \bullet \gamma(xe^{-1}L_t(M),a)}{\gamma(x^{-1}L_t(M),e)} \qquad (3.1)$$

เราประมาณค่าความถี่ของสถานะที่สร้างขึ้นใหม่ ด้วยการประมาณจากความถี่ของการส่งผ่านจากสถานะเดิม ซึ่งแน่นอนว่าการประมาณเช่นนี้สามารถลดความคลาดเคลื่อนของการคำนวณความน่าจะเป็นได้ระดับหนึ่ง ณ เวลาที่ t ใดๆ อย่างไรก็ตามจากหลักการเรียนรู้ของเราซึ่ง

เป็นการเรียนรู้แบบสะสมข้อมูล สมมติฐานที่ถูกเปลี่ยนแปลงแล้วจะไม่เปลี่ยนแปลงอีก ดังนั้นความน่าจะเป็นย่อมลู่เข้าสู่ค่าที่ถูกต้องเสมอ ณ เวลาที่ $t \rightarrow \infty$ ตามทฤษฎีบทย่อยดังนี้

ทฤษฎีบทที่ 3.15 ด้วยการประมาณค่าความถี่ด้วยสมการที่ 3.1 และจากภาษาที่ได้จากนิยามที่ 3.7 – 3.13 ความน่าจะเป็นที่คำนวณได้จากความถี่ดังกล่าวจะลู่เข้าสู่ค่าที่ถูกต้องเสมอ เมื่อเวลาที่ $t \rightarrow \infty$

พิสูจน์ จากนิยามการสร้างภาษา เราเห็นได้ว่าเมื่อสถานะใดๆมีการเปลี่ยนแปลงสมมติฐานครั้งหนึ่งแล้ว จะไม่มีการเปลี่ยนแปลงสมมติฐานที่สถานะเดิมอีก ให้ $\bar{\gamma}$ แทนค่าความถี่ที่ถูกต้อง จะได้ว่า $\bar{\gamma}(ve^{-1}L_t(M),a) - \gamma(ve^{-1}L_t(M),a) = k$ เมื่อ k คือค่าคงที่ เรากำหนดให้ค่าความคลาดเคลื่อนมีค่าดังนี้

$$\xi(ve^{-1}L_t(M),w) = \frac{\bar{\gamma}((ve)^{-1}L_t(M),w) - \gamma((ve)^{-1}L_t(M),w)}{\gamma(v^{-1}L_t(M),e)}$$

ได้ว่า $$\xi((ve)^{-1}L_t(M),w) = \frac{k}{\gamma(v^{-1}L_t(M),e)}$$

เมื่อ γ เข้าใกล้ค่าอนันต์ ในกรณีที่มีการเรียนรู้ไปเรื่อยๆ แล้วจะได้ว่า

$$\xi((ve)^{-1}L_M,w) = \lim_{\gamma\to\infty}\frac{k}{\gamma} = 0$$

เราพบว่าเมื่อค่าความถี่เข้าใกล้ค่าอนันต์แล้ว ค่าความคลาดเคลื่อนของความน่าจะเป็นเข้าใกล้ศูนย์ ดังนั้นเราจึงสรุปได้ว่า ความน่าจะเป็นที่คำนวณจากความถี่จะเข้าสู่ค่าที่ถูกต้อง ∎

3.4 แบบจำลองที่ใช้เรียนรู้ภาษา

เนื่องจากงานวิจัยนี้เลือกภาษาสโทแคสติกเป็นแบบในการเรียนรู้ ดังนั้นแบบจำลองที่เหมาะสมคือ สโทแคสติกออโตมาตา แต่เนื่องจากคลาสของภาษาที่ได้ออกแบบไว้นั้น มีการสนใจในอักขระนำหน้าของแต่ละผลหารของภาษา เพื่อความสะดวกในการเรียนรู้แบบเชื่อมตรง งานวิจัยนี้ได้เสนอแบบจำลองชนิดใหม่ซึ่งเป็นแบบจำลองที่ปรับเปลี่ยนพารามิเตอร์บางตัวจากสโทแคสติกออโตมาตาแบบเดิม โดยได้นำเสนอไว้ดังนี้

3.4.1 **สโทแคสติกออโตมาตาอิงอักขระนำหน้า** (Stochastic Finite Precedent Automaton: SFPA)

ในงานวิจัยนี้จะนำเสนอแบบจำลองชนิดใหม่ซึ่งปรับเปลี่ยนจาก ออโตมาตาเชิงน่าจะเป็นเชิงกำหนดของเดิมเล็กน้อยโดยเรียกแบบจำลองชนิดนี้ว่า *สโทแคสติกออโตมาตาอิงอักขระนำหน้า* (stochastic finite precedent automaton) (SFPA) เรียกโดยย่อว่า เอสเอฟพีเอ ซึ่งสถานะแต่ละสถานะของ เอสเอฟพีเอ จะมีฉลากระบุด้วยสายอักขระที่เกิดขึ้นจริงในชุดข้อมูลตัวอย่าง เอสเอฟพีเอสามารถนิยามได้ดังนี้

นิยามที่ 3.16 *สโทแคสติกออโตมาตาอิงอักขระนำหน้าประกอบไปด้วยโครงสร้าง [Q,* $\sum$*,* ε*,* τ*,* γ *,*φ *] โดยที่*

Q คือ เซตจำกัดของสถานะ

$\sum$ *คือ เซตของอักขระที่เป็นไปได้ทั้งหมด*

ε *คือ สายอักขระว่างซึ่งเป็น สถานะเริ่มต้น*

$\tau : Q \times \sum \rightarrow Q$ *คือฟังก์ชันการส่งผ่าน*

$\gamma : Q \times \sum \rightarrow R$ *คือฟังก์ชันความถี่ในการส่งผ่านของเหตุการณ์ถัดไป*

φ *: Q -->* $\sum \cup \varepsilon \cup N$ *คือฟังก์ชันอักขระนำหน้าของสถานะใดๆซึ่งมีโดเมนเป็นสถานะใดๆ และเรนจ์เป็น* $\sum \cup \varepsilon$

จากนิยามฟังก์ชันแต่ละฟังก์ชันสามารถเขียน *ฟังก์ชันการส่งผ่าน* τ ซึ่งหมายความว่า จากสถานะ *q* ใดๆมีการส่งผ่านไปยังสถานะ *q* ได้ด้วยเหตุการณ์ *a* เขียนสัญลักษณ์แทนได้ด้วยสมการ

$$\forall q \in Q, \forall a \in \sum: \tau(q, a) = q$$

ฟังก์ชันความถี่ในการส่งผ่านของเหตุการณ์ถัดไป γ หมายความว่าจากสถานะ *q* ใดๆ มีความถี่ในการส่งผ่านด้วยเหตุการณ์ *a* เขียนสัญลักษณ์แทนได้ด้วยสมการ

$$\forall q \in Q, \forall a \in \sum, n \in \mathbf{N}: \gamma(q, a) = n$$

ฟังก์ชันอักขระนำหน้าของสถานะใดๆ φ หมายถึง สถานะใดๆ จะยอมรับอักขระนำหน้าจำนวน 1 ตัวเท่านั้น

$$\forall q \in Q, \forall a \in \sum \cup \varepsilon: \varphi(q) = a$$

ซึ่งในการคำนวณความน่าจะเป็นของการเกิดสายอักขระ *x* โดยที่ $x = x_1\ x_2\ x_3 \ldots x_n$ สามารถคำนวณแบบเวียนเกิดเพื่อหาสถานะถัดไปจากสมการ

$$\tau(q, x_1\, x_2\, x_3 \ldots x_n) = \tau(\tau(q, x_1\, x_2\, x_3 \ldots x_{n-1}), x_n)$$

และเพื่อหาความน่าจะเป็นรวมในการส่งผ่านทั้งหมดคำนวณได้จากสมการ

$$p(x) = \prod_{i=1}^{n} p(q_{i-1}, x_i)$$

โดยที่ p นิยามความน่าจะเป็นในการส่งผ่านจากสถานะ q_i ด้วยเหตุการณ์ x_i สามารถคำนวณได้จากสมการ

$$p(q_i, x_i) = \frac{\gamma(q_i, x_i)}{\gamma(q_i, \#)}$$

โดยนิยาม q_0 คือสถานะตั้งต้นในที่นี้ได้แก่ ε ซึ่งมีความน่าจะเป็นในการเกิดของสถานะนี้ เท่ากับ 1 และ $\gamma(q_i, \#)$ แทนความถี่รวมในการส่งผ่านทุกเหตุการณ์จากสถานะ q_i

3.5 **อัลกอริทึมในการเรียนรู้**

ในหัวข้อนี้จะเสนออัลกอริทึมที่ใช้ในการเรียนรู้ด้วยวิธีการแบบเชื่อมตรง โดยอาศัยหลักการจากภาษา *L(M)* ที่ออกแบบไว้ตามนิยามที่กล่าวไว้ข้างต้น โดยอัลกอริทึมนี้มีชื่อว่า (Dynamic On-line Inductive Inference) เขียนโดยย่อว่า DO^{II} (ดู-ทู) โดยที่สิ่งที่จะได้ภายหลังการเรียนรู้คือ สโทแคสติกออโตมาตาอิงอักขระนำหน้า โดยที่แต่ละสถานะจะถูกฉลากชื่อไว้ด้วยสายอักขระที่แสดงผลหารภาษาเทียบกับออโตมาตา และสโทแคสติกออโตมาตาจะมีสถานะเริ่มต้นเพียงสถานะเดียวซึ่งถูกฉลากชื่อด้วย ε เราได้นิยามคำจำกัดความต่างๆ และนำเสนอรหัสเทียมของอัลกอริทึมดังต่อไปนี้

นิยามที่ 3.16 *กำหนดให้ อักขระที่ใช้ในการเรียนรู้จากสายอักขระใด ณ ตำแหน่งที่ t แทนได้ด้วย* s_t *และให้เซตของสถานะที่จะต้องพิจารณา ณ เวลา t ใดๆเรียกว่า สถานะแอคทิฟ (active state) แทนได้ด้วย* $ac_t = \{ (xe)^{-1}L_{t-1}$ *โดยที่* $e \in \Sigma$ *และ* $e = s_t$ *และ* $x \in ac_{t-1} \} \cup \{\varepsilon^{-1}L\}$ *โดยที่* $\{\varepsilon^{-1}L\} = ac_0$

นิยามที่ 3.17 *กำหนดให้* $x_n^{-1}L_t \in ac_t$ *โดยที่* $|x_1|<|x_2|<|x_3|\ldots|x_n|$ *เรากำหนดให้ สถานะถัดไปที่เป็นไปได้ (possible next state) ของ* $x_n^{-1}L_t$ *เขียนแทนด้วยฟังก์ชัน pn โดยที่* $pn(x_n^{-1} L_t) = (x_{n-1}e)^{-1}L_t$ *เมื่อ* $e = s_t$ *และ* $|x_{n-1}| \geq 1$

นิยามที่ 3.18 *ให้ M เป็น เอสเอฟพีเอ ตัวหนึ่งและมีการส่งผ่านใดๆจากสถานะ q_1 ไปยังสถานะ q_2 โดยกำหนดให้ $q_1 = y^{-1}L$ โดยที่ $y = x_1 \ldots x_k \ldots x_{k+n}$ และ.$q_2 = x^{-1}L$ โดย $x = x_k x_{k+1} \ldots x_{k+n+1}$, ฟังก์ชันอักขระนำหน้าที่สั้นที่สุด ϕ ของสถานะ q_2 จาก q_1 สามารถคำนวณได้จากสมการ*

$$\phi(q_1,q_2) = \begin{cases} x_{k-1} & \text{if } k > 1 \\ \varepsilon & \text{if } k = 1 \text{ or } k = null \end{cases}$$ *k* เป็นจำนวนเต็มตั้งแต่1 ถึง *n*-1.

นิยามที่ 3.19 *กำหนดให้ $x_n^{-1}L_t \in ac_t$ การส่งผ่านจะเรียกว่า สถานะที่ถูกยอมรับจาก $x_n^{-1}L_t$ ไปยัง $pn(x_n^{-1} L_t)$ ด้วยเหตุการณ์ s_t ก็ต่อเมื่อ $\phi(x_n^{-1} L_t , pn(x_n^{-1} L_t)) = \varphi(pn(x_n^{-1} L_t))$ ในกรณีที่ $\varphi(pn(x_n^{-1} L_t)) \neq \{\}$ หรือ ยอมรับทุกครั้งในกรณีที่ $\varphi(pn(x_n^{-1} L_t)) = \{\}$*

นิยามที่ 3.20 *ในกรณีที่ $\phi(x_n^{-1} L_t , pn(x_n^{-1} L_t)) \neq \varphi(pn(x_n^{-1} L_t))$ และ $\varphi(pn(x_n^{-1} L_t)) \neq \{\}$ เรานิยามสถานะ $pn(x_n^{-1} L_t)$ นั้นว่าเป็น สถานะที่ต้องเปลี่ยนสมมติฐาน แทนได้ด้วยq_c*

จากนิยามข้างต้นเราได้เสนออัลกอริทึมในการเรียนรู้ดังต่อไปนี้

Learning Algorithm (Dynamic online identify induction) DO^{II}: Do-Two

Input : Strings

Output : SFPA

Initialization: $ac_0 = \{\varepsilon\}$, Q={ }, $\sum$={ }

WHILE s_t < > End of input stream DO

{ Add ε to ac_{t+1} ; // add ε to active state

$\sum \cup \{ s_t \}$; // add s_t to alphabet

WHILE ac_t < > { } DO

{ pick minimize $x_n^{-1} L_t$ from ac_t ;

$ac_t - \{ x_n^{-1} L_t \}$;

IF $\tau(x_n^{-1} L_t , s_t) = NULL$ AND $pn(x_n^{-1} L_t) = NULL$ THEN // there is no transition

{ Q $\cup (x_n s_t)^{-1} L_t$; // create new state

$\tau(x_n^{-1} L_t , s_t) = (x_n s_t)^{-1} L_t$; // add a new transition

$ac_{t+1} \cup (x_n s_t)^{-1} L_t$; // add a new state to active state set

$pn(x_{n+1}^{-1} L_{t+1}) = (x_n s_t)^{-1} L_t$; // set a possible next state

IF ($x_n^{-1} L_t$ = last of ac_t)

$\varphi(pn(x_n^{-1} L_t)) = \phi(x_n^{-1} L_t, pn(x_n^{-1} L_t))$; } // create a precedent function

ELSE IF $\tau(x_n^{-1} L_t, s_t) = NULL$ AND $pn(x_n^{-1} L_t) <> NULL$ THEN // there is a possible next state

{ IF $pn(x_n^{-1} L_t) <> q_c$ THEN // if a possible next state can be a destination state for some transition (see definition 13, 14)

{ $\tau(x_n^{-1} L_t, s_t) = pn(x_n^{-1} L_t)$; // add a transition to a possible next state

IF ($x_n^{-1} L_t$ = last of ac_t)

$\varphi(pn(x_n^{-1} L_t)) = \phi(x_n^{-1} L_t, pn(x_n^{-1} L_t))$; } // add precedent function

ELSE // $pn(x_n^{-1} L_t) = q_c$ a possible next state is a state need to modify

{ $Q \cup (x_n s_t)^{-1} L_t$; // create new state

$\tau(x_n^{-1} L_t, s_t) = (x_n s_t)^{-1} L_t$; // add a new transition

$ac_{t+1} \cup (x_n s_t)^{-1} L_t$;// add a new state to active state set

$ac_{t+1} \cup q_c$; // add a state need to modify to active state set

$pn(x_{n+1}^{-1} L_{t+1}) = (x_n s_t)^{-1} L_t$; // set a possible next state

MODIFY($pn(x_n^{-1} L_t)$); }

ELSE

{ $\gamma(x_n^{-1} L_t, s_t) = \gamma(x_n^{-1} L_t, s_t) + 1$; }//increment next symbol frequency function by 1

} // end while

MODIFY($pn(x_n^{-1} L_t)$)

{ $\varphi(pn(x_n^{-1} L_t)) = '\varepsilon'$; //

$Q_t \cup Lnew$; // see definition of *Lnew* in definition 8

$\tau(v^{-1} L_t, e) = w^{-1} L_t \quad \forall w \in Lnew, \ \forall v \in C_n$;

$w^{-1} L_t = \text{clone}(pn(x_n^{-1} L_t)) \ \forall w \in Lnew$;

$\varphi(\tau(pn(x_n^{-1} L_t), e)) = \phi(mc^{-1} L_t, \tau(pn(x_n^{-1} L_t), e))$;// mc= minimize candidate, edit new precedent function

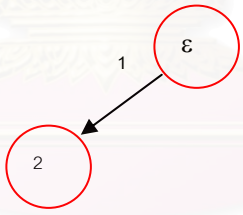
For $\forall a$ where $\tau(pn(x_n^{-1} L_t), a)$ exist $\gamma(w^{-1}L, a) \cong \frac{\gamma(v^{-1}L, e) \bullet \gamma(pn(x_n^{-1}L, a)}{\gamma(x_{n-1}^{-1}L, e)}$

}

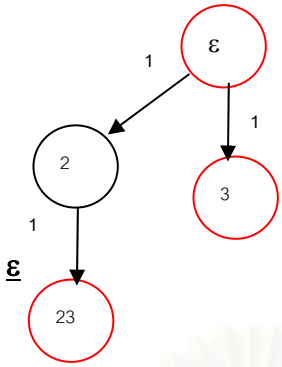
ตัวอย่างที่ 3.1 การใช้อัลกอริทึม DO[II] ในการเรียนรู้จากสายอักขระ “23512321351”

1. เริ่มที่สถานะเริ่มต้น

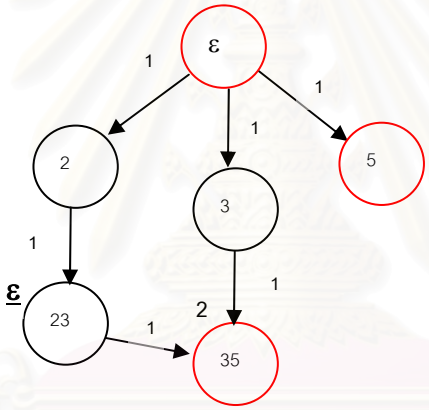
2. รับ 2 เข้ามาจากข้อมูล พิจารณาที่สถานะ ε โดยจะทำการสร้างสถานะ 2 และการส่งผ่านจาก ε ไปยัง 2 ในกรณีที่ไม่มีการส่งผ่าน พร้อมกับเก็บความถี่ 1 เพิ่มสถานะ ε และ 2 เป็นสถานะแอคทีฟ ac_t



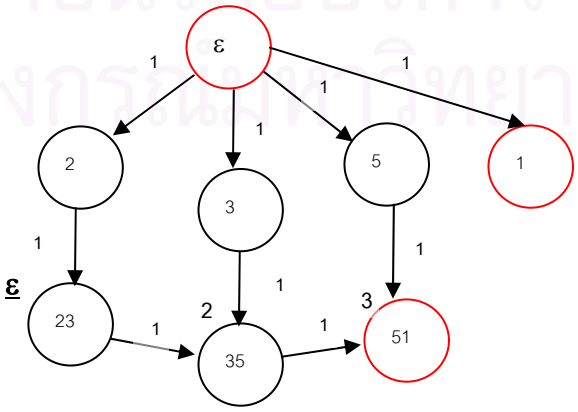
3. รับ 3 เข้ามาจากข้อมูลตัวอย่าง จะดึงสมาชิกใน ac_t คือ ε และ 2 มาพิจารณาตามลำดับ โดยจะทำการสร้างสถานะต่อจากสถานะดังกล่าว เนื่องจากยังไม่มีการส่งผ่านใดๆ และในกรณีที่พิจารณาสมาชิกตัวสุดท้ายใน ac_t ในที่นี้คือสถานะ 2 ทำการเพิ่ม $\varphi(23) = \varepsilon$ แทนด้วยตัวเลขที่ขีดเส้นใต้ เพิ่ม สถานะ ε และ 3, 23 เป็นสถานะแอคทีฟ ac_t



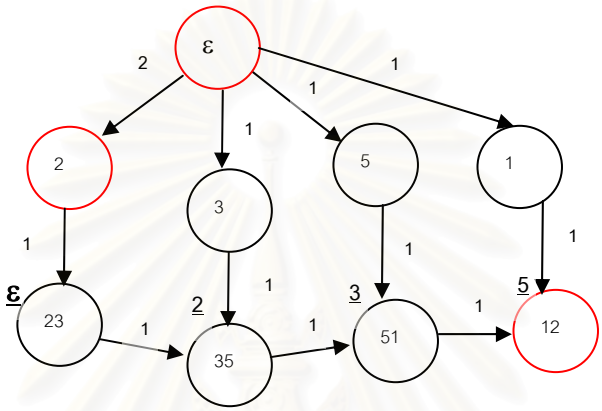
4. รับ 5 เข้ามา จะพิจารณาจากสถานะที่แอคทีฟคือ ε, 3, และ 23 โดยที่สร้างสถานะใหม่ที่ชื่อ 5 และ 35 โดยที่ 35 จะเป็นสถานะถัดไปที่เป็นไปได้ของ 23 (ดูนิยามที่ 3.17) และจะทำการเพิ่มฟังก์ชันอักขระนำหน้า $\varphi(35) = 2$ เมื่อสิ้นสุดการพิจารณาสถานะแอคทีฟตัวสุดท้าย เพิ่มสถานะ ε, 5, 35 เป็นสถานะแอคทีฟ



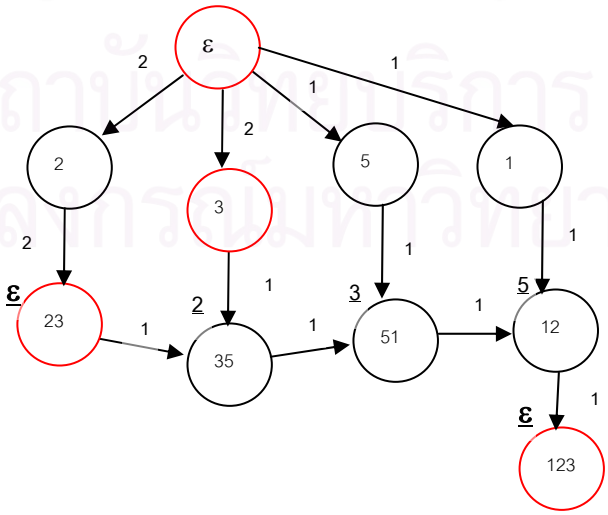
5. รับ 1 เข้า จะทำการสร้าง สถานะ 1 ใหม่พร้อมกับการส่งผ่าน จากนั้นจะสร้างสถานะ 51 ต่อจากสถานะ 5 แล้วกำหนดให้เป็นสถานะถัดไปที่เป็นไปได้ และทำการสร้างการส่งผ่านจาก 35 ไปยัง 51 พร้อมกับเก็บฟังชัน $\varphi(51) = 3$ เพิ่มสถานะ ε, 1, 51 เป็นสถานะแอคทีฟ



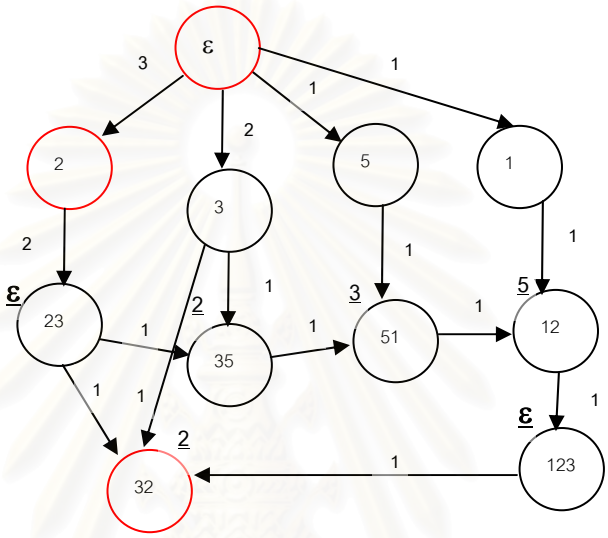
6. รับ 2 เข้า เนื่องจากเมื่อเราพิจารณาการส่งผ่านจากสถานะ ε ปรากฏว่า มีการส่งผ่านด้วย 2 อยู่แล้ว ดังนั้นจะทำการเพิ่มความถี่ที่การส่งผ่านดังกล่าว จากนั้นจะทำการสร้าง สถานะ 12 และการส่งผ่านจาก 51 ไป 12 พร้อมกับสร้าง $\varphi(12) = 5$ แล้วทำการเพิ่มสถานะε, 2, 12 เป็นสถานะแอคทีฟ



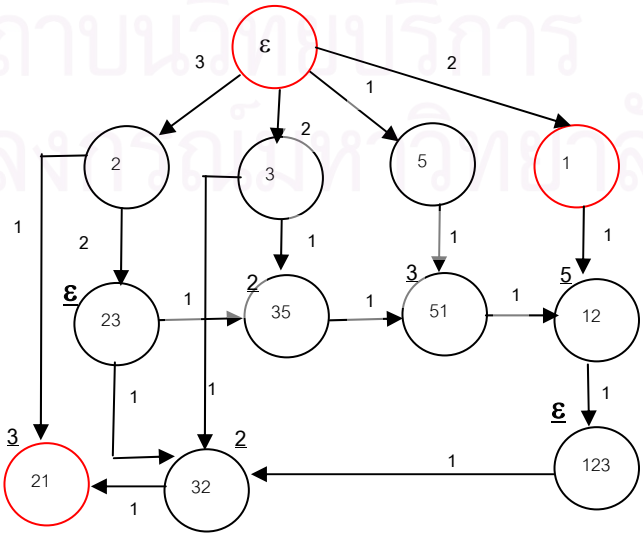
7. รับ 3 เข้า จากนั้นพิจารณาสถานะแอคทีฟแต่ละตัว ทำการเพิ่มความถี่ในกรณีที่สถานะที่พิจารณามีการส่งผ่านด้วย 3 ปรากฏว่า สถานะ 23 ซึ่งเป็นสถานะถัดไปที่เป็นไปได้นั้นไม่อาจจะยอมรับการส่งผ่านมาจากสถานะ 12 ได้ จึงทำการสร้างสถานะใหม่ 123 และเนื่องจาก 12 เป็นตัวสุดท้ายในสถานะแอคทีฟจึงทำการคำนวณ $\varphi(123) = \varepsilon$ เพิ่มสถานะε, 3, 23, 123 เป็นสถานะแอคทีฟ



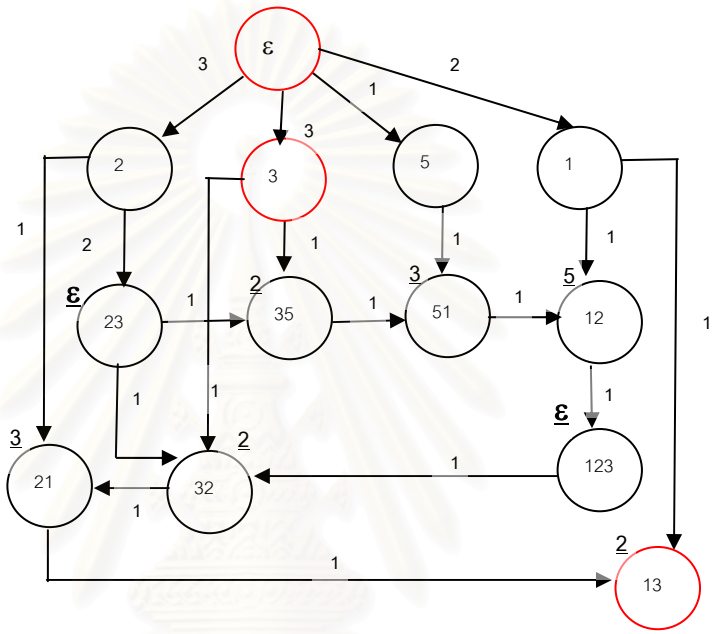
8. รับ 2 เข้า พิจารณาที่สถานะ ε ทำการเพิ่มความถี่การส่งผ่านไปยังสถานะ 2 พิจารณาที่สถานะ 3 แล้วทำการสร้างการส่งผ่านไปยังสถานะ 32 แล้วให้ 32 เป็นสถานะถัดไปที่เป็นไปได้ จากนั้นทำการสร้างสถานะจาก 23 และ 123 ไปยังสถานะ 32 แล้วทำการเพิ่ม $\varphi(32) = 2$ เพิ่มสถานะ ε, 2, 32 เป็นสถานะแอคทีฟ



9. รับ 1 เข้า ทำการเพิ่มความถี่ที่การส่งผ่านจากสถานะ ε ไปยังสถานะ 1 ทำการสร้างสถานะ 21 และการส่งผ่านจากสถานะ 2 และ 32 ไปยังสถานะ 21 แล้วเพิ่ม $\varphi(21) = 3$ เพิ่มสถานะε, 1, 21 เป็นสถานะแอคทีฟ

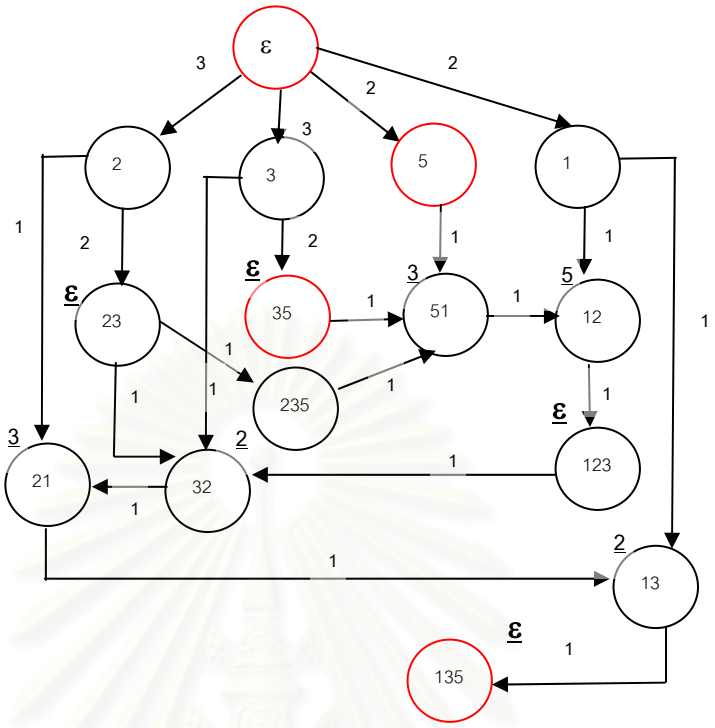


10. รับ 3 เข้า ทำการเพิ่มความถี่ที่การส่งผ่านจากสถานะ **ε** ไปยังสถานะ 3 ทำการสร้างสถานะ 13 แล้วทำการสร้างการส่งผ่านจาก 21 ไปยัง 13 ทำการสร้าง $\varphi(13)=2$ เพิ่มสถานะ **ε**, 3, 13 เป็นสถานะแอคทีฟ

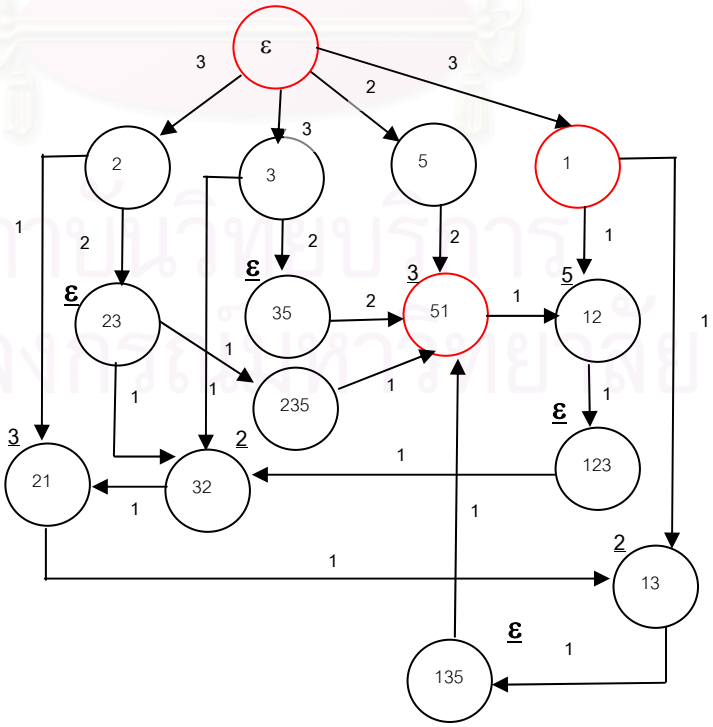


11. รับ 5 เข้าพิจารณา ทำการเพิ่มความถี่ที่การส่งผ่านจากสถานะ **ε** ไปยัง 5 แล้วทำการเพิ่มความถี่จากสถานะ 3 ไปยัง 35 ในกรณีนี้ 35 จะเป็นสถานะถัดไปที่เป็นไปได้ที่จะต้องพิจารณาร่วมกับสถานะ 13 แต่เนื่องจากเมื่อพิจารณาร่วมกันแล้วจะทำให้ 35 เป็นสถานะที่ต้องการเปลี่ยนสมมติฐาน ตามนิยาม 13 -14 เราทำการเปลี่ยนแปลงโดยกระทำการดังนี้

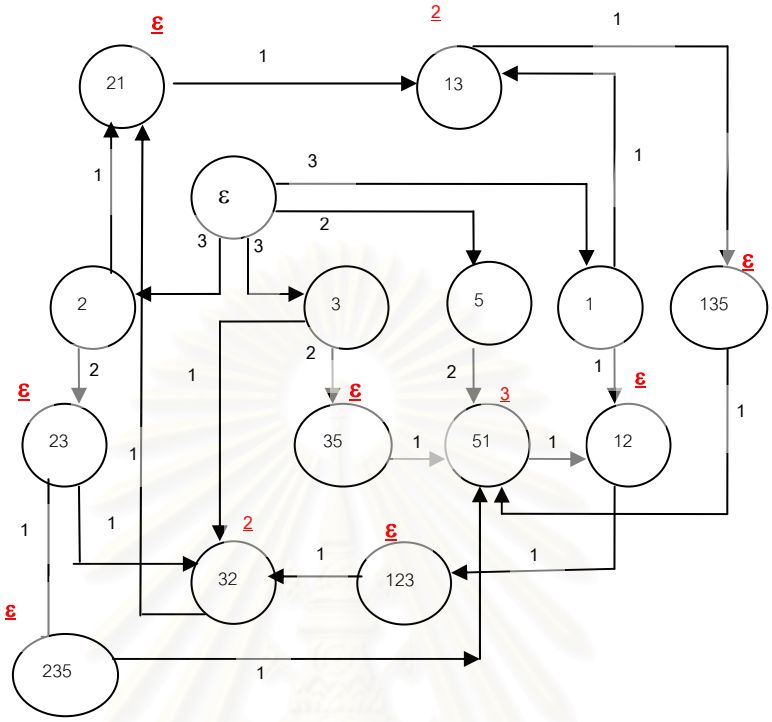
- ทำการสร้างสถานะใหม่ 135 และสร้างการส่งผ่าน
- เพิ่ม 35 และ 135 ในสถานะแอคทีฟ
- เปลี่ยนฟังก์ชันอักขระนำหน้าของ 35 ให้เป็น **ε**
- ในกรณีนี้ *Lnew* คือ 235
- ทำการโคลนสถานะ 235 กับ 35
- ประมาณความถี่การส่งผ่านจาก 235 ไปยัง 51 ในที่นี้ ยังคงเท่ากับ 1
- เพิ่มสถานะ **ε**, 5, 35, 135 เป็นสถานะแอคทิฟ



12. รับ 1 เข้ามาพิจารณา ทำการเพิ่มความถี่ที่การส่งผ่านจากสถานะ ε ไปยัง 1 เพิ่มความถี่การส่งผ่านจากสถานะ 5 ไปยัง 51 รวมทั้งจากสถานะ 35 ไปยัง 51 เนื่องจากมีการส่งผ่านอยู่ จากนั้นให้ สถานะ 51 เป็นสถานะถัดไปที่เป็นไปได้ของ 135



ซึ่งเราจะได้



รูปที่ 2 แผนภาพ*เอสเอฟพีเอ*ที่ได้จากการเรียนรู้จากสายอักขระ “23512321351”

3.6 การทำนายด้วยการคำนวณความน่าจะเป็นของสายอักขระใด ๆ

ความสำคัญอย่างหนึ่งของการเรียนรู้ก็คือความสามารถในการทำนายข้อมูล เนื่องจากการทำนายข้อมูลด้วยสโทแคสติกออโตมาตานั้นอาศัยความน่าจะเป็นในการทำนาย ดังนั้นการคำนวณความน่าจะเป็นที่เหมาะสมจึงมีส่วนอย่างมากต่อการทำนาย ถึงแม้ว่าสโทแคสติกออโตมาตาจะสามารถคำนวณหาความน่าจะเป็นได้อย่างเหมาะสมจากการคำนวณความน่าจะเป็นสะสมในการส่งผ่านแต่ละสถานะ ถึงกระนั้นก็ตาม เราก็ยังคงพบกับปัญหาในกรณีที่พบกับเหตุการณ์ที่ไม่มีการส่งผ่านต่อจากสถานะใดสถานะหนึ่ง ด้วยสถานการณ์เช่นนี้ หากคำนวณด้วยความน่าจะเป็นเท่ากับศูนย์ย่อมทำให้ความน่าจะเป็นเปลี่ยนแปลงไปอย่างมาก ซึ่งไม่เหมาะสม เนื่องจากเราไม่นิยามความน่าจะเป็นในการส่งผ่านที่ยังไม่เคยเกิดขึ้นให้เท่ากับศูนย์ เราเรียกการจัดการกับปัญหาการคำนวณการส่งผ่านที่ไม่เคยเกิดขึ้นมาก่อนอย่างที่กล่าวไปว่า *เทคนิคการปรับเรียบ* (Smoothing technique) มีงานวิจัยหลายงานที่พยายามทำเทคนิคนี้ให้มีประสิทธิภาพมากยิ่งขึ้น ส่วนใหญ่จะทำด้วยการปรับเปลี่ยนพารามิเตอร์ที่แบบจำลองด้วย*หลักการย้อนกลับ* (back off)[3] คือย้อนกลับไปหาการคำนวณความน่าจะเป็นจากเงื่อนไขที่สั้นกว่าเดิม ซึ่งหลักการนี้ก็ยังมี

การประมาณค่าหลายวิธี อย่างไรก็ตามเราไม่สามารถบอกได้ว่าวิธีใดดีกว่ากันเนื่องจากเป็นการประมาณค่าที่เหมาะสมเท่านั้น อย่างไรก็ตามเทคนิคในการปรับเปลี่ยนพารามิเตอร์ที่ตัวแบบจำลองภายหลังการเรียนรู้นี้ นิยมทำกันในแบบจำลองที่ผ่านการเรียนรู้แบบคงที่มา ซึ่งไม่ได้สนใจที่จะทำการเรียนรู้ต่อไปอีกการปรับเปลี่ยนพารามิเตอร์ที่แบบจำลองโดยตรงจึงสามารถทำได้โดยไม่มีปัญหา แต่เนื่องจากในงานวิจัยนี้มุ่งเน้นไปที่การเรียนรู้แบบเชื่อมตรง และเพื่อให้มีความสามารถต่อการโต้ตอบโดยอัตโนมัติ การปรับเปลี่ยนที่พารามิเตอร์ที่แบบจำลองจึงไม่เหมาะสมเท่าที่ควร ด้วยปัจจัยของเวลาและเพื่อความไม่ยุ่งยากต่อการเรียนรู้ในครั้งต่อไป *เราเสนอว่าหลักการคำนวณสายอักขระที่ไม่เคยเกิดขึ้นมาก่อนนั้นควรใช้วิธีการคำนวณด้วยสมการที่เหมาะสมตามพารามิเตอร์ที่มีอยู่แล้วของแบบจำลองเดิมที่ได้จากการเรียนรู้มากกว่าการที่จะปรับเปลี่ยนพารามิเตอรที่แบบจำลอง* เพื่อที่จะรองรับการเรียนรู้แบบเชื่อมตรง เราเสนอหลักการคำนวณเพื่อแก้ปัญหาการส่งผ่านที่ไม่เคยเกิดขึ้นมาก่อนอย่างง่ายโดยอาศัยหลักการ มาร์คอฟเชิงความยาวแปรผัน ซึ่งมีหลักการคำนวณว่า เมื่อใดก็ตามที่พบว่าการส่งผ่านจากสถานะใดๆด้วยเหตุการณ์หนึ่งนั้นไม่มีอยู่ เราจะคำนวณความน่าจะเป็นในการส่งผ่านจากสถานะที่มีเงื่อนไขที่น้อยลงซึ่งเราเสนอสมการดังนี้

$$p(x) = \prod_{i=1}^{n} p(q_i, x_i)$$

$$\text{โดยที่ } q_i = \begin{cases} \tau(q_{i-1}, x_{i-1}) & \text{ถ้า } \tau(q_i, x_i) \neq NULL \\ suffix(q_i) & \text{ในกรณี } \tau(q_i, x_i) = NULL \end{cases}$$

โดยที่ $q_1 = \varepsilon$

ตัวอย่างที่ 3.2 การคำนวณความน่าจะเป็นจากแบบจำลอง

จากรูปที่ 4 หากมีสายอักขระที่เราต้องการคำนวณหาความน่าจะเป็นในการเกิดดังนี้ “2321235” ซึ่งจะคำนวณได้เท่ากับ $p(2321235)=p(2|\varepsilon)\cdot p(3|2)\cdot p(2|23)\cdot p(1|32)\cdot p(2|1)\cdot p(3|12)\cdot p(5|23) = 3/11\cdot 2/3\cdot 1/2\cdot 1\cdot 1/2\cdot 1\cdot 1/2 = 1/44$ เราได้คำตอบว่า การประมาณความน่าจะเป็นในการเกิดสายอักขระดังกล่าว เท่ากับ 1/44 หรือประมาณ 2.27 เปอร์เซ็นต์

3.7 ผลการวิเคราะห์

จากการวิเคราะห์อัลกอริทึมและคลาสของภาษาที่ได้นำเสนอไปหัวข้อก่อนหน้านี้ เราได้เสนอผลวิเคราะห์สรุปเป็นทฤษฎีดังนี้

ทฤษฎีบทที่ 3.21 *อัลกอริทึม DO^{II} สามารถก่อกำเนิดเอสเอฟพีเอซึ่งเทียบเท่าภาษาสโทแคสติกตามที่ออกแบบไว้จริง*

พิสูจน์ ให้ Q_t คือ เซตของสถานะตามนิยามการออกแบบภาษา Q_{SFPA} คือ เซตของสถานะที่ได้จากอัลกอริทึม DO^{II} จากนิยามที่ 3.12

<u>กรณีแรก</u> ในกรณีที่ไม่มีการเปลี่ยนสมมติฐาน

$Q_t = Q_t \cup \{xe^{-1}L\}$ ในกรณีที่ $(ye)^{-1}L_t$ *เทียบเท่า* $(xe)^{-1}L_t$ ซึ่งจะไม่มีการสร้างสถานะ $(ye)^{-1}L_t$ เพิ่ม ในกรณีนี้ y ที่เป็นไปได้ทั้งหมดจะต้องจะต้องมีคุณสมบัติคือ $y=sax$

$Q_{SFPA}= Q_{SFPA} \cup \{y_k e^{-1}L_{SFPA}\}$ ในกรณีที่ $\forall y \in ac$ และเนื่องจาก y ทุกตัวใน ac เป็นคำเติมท้ายของสมาชิกตัวที่ยาวกว่าเสมอซึ่งมีคุณสมบัติ $y_{k+n}= say_k$ ดังนั้นในกรณีที่มีการสร้างสถานะใหม่ $(y_k e)^{-1}L_{SFPA}$ขึ้นมา เราจึงสรุปได้ว่า $(y_{k+n}e)^{-1}L_{SFPA}= (y_k e)^{-1}L_{SFPA}$

ให้ $(y_{k+n}e)^{-1}L_{SFPA}=(ye)^{-1}L_t$ และ $y_k e^{-1}L_{SFPA} = (xe)^{-1}L_t$ แล้ว

เราสรุปว่าในกรณีที่ไม่มีการเปลี่ยนสมมติฐาน $Q_t= Q_{SFPA}$

<u>กรณีที่สอง</u> ในกรณีที่มีการปรับเปลี่ยนสมมติฐาน

$Q_t = Q_t \cup Lnew(M)$ ตามนิยามที่ 3.13 เมื่อมีสถานะที่ต้องการเปลี่ยนสมมติฐาน $(xe)^{-1}L_t$ เกิดขึ้น โดยสถานะที่ต้องการเปลี่ยนสมมติฐานเกิดขึ้น เมื่อพบสถานะที่ทำให้อักขระนำหน้าของ $(xe)^{-1}L_t$ มีมากกว่า 1 ตัว

$Q_{SFPA}= Q_{SFPA} \cup Lnew(M)$ ตามฟังก์ชัน Modify เมื่อมีสถานะที่ต้องการเปลี่ยนสมมติฐานเกิดขึ้น เมื่อพบว่า ฟังก์ชันอักขระนำหน้าส่งค่าคืนไม่เท่ากับ ค่าที่คำนวณได้จากฟังก์ชันอักขระนำหน้าที่สั้นที่สุด ตามนิยามที่ 3.20 ซึ่งทำให้อักขระนำหน้าสถานะที่ต้องการเปลี่ยนสมมติฐานมีอักขระนำหน้ามากกว่า 1 ตัว ตรงกับกรณีของ Q_t

เราสรุปว่า ในกรณีที่มีการเปลี่ยนสมมติฐานเกิดขึ้นในกรณีเดียวกันและ $Q_t= Q_{SFPA}$ ∎

หลังจากที่เราได้พิสูจน์ว่าเราสามารถเรียนรู้สายอักขระใดๆด้วยอัลกอริทึม DO^{II} ได้จริง การวิเคราะห์ที่สำคัญต่อการเรียนรู้คือ ประสิทธิภาพของการจากการเรียนรู้ด้วยการคำนึงถึงประสิทธิภาพการเรียนรู้ตามนิยาม 3.1 โดยจากการวิเคราะห์ผลการทำงานของ DO^{II} เราได้เสนอทฤษฎีดังต่อไปนี้

ทฤษฎีบทที่ 3.22 *อัลกอริทึม DO^{II} สามารถก่อกำเนิดเอสเอฟพีเอซึ่งเทียบเท่าภาษาสโทแคสติกตามที่ออกแบบไว้ ด้วยอัตราการเติบโตของไวยากรณ์ไม่เกินฟังก์ชันโพลีโนเมียล และอัตราเร็วในการก่อกำเนิดเอสเอฟพีเอไม่เกินฟังก์ชันโพลีโนเมียลของความยาวรวมของสายอักขระจากข้อมูลตัวอย่าง*

พิสูจน์ เราจะทำการพิสูจน์ 2 ขั้นตอนคือ อัตราเร็วของอัลกอริทึม DO^{II} เติบโตไม่ช้าไปกว่าฟังก์ชันพหุนามของ จำนวนรวมความยาวสายอักขระที่นำมาเรียนรู้ และอัตราการเติบโตของไวยากรณ์ในแต่ละรอบที่รับข้อมูลเข้ามาเรียนรู้ โดยที่การเติบโตของไวยากรณ์ซึ่งในที่นี้คือ จำนวนสถานะของเอสเอฟพีเอในแต่ละรอบที่โตขึ้น จะต้องเติบโตไม่เกินฟังก์ชันพหุนามของจำนวนสถานะเดิมที่มีอยู่ เราได้ทำการพิสูจน์ดังนี้

กรณีที่ 1 ในส่วนของอัตราเร็วในเรียนรู้เทียบกับความยาวรวมของสายอักขระนั้น จะอาศัย DO^{II} อัลกอริทึมเป็นตัววิเคราะห์ กล่าวคือ ในกรณีที่รับ อักขระเข้ามา n ตัว อักขระแต่ละตัวที่มาจาก n อักขระนั้น จะถูกนำเอามาพิจารณาร่วมกับสถานะแอคทีฟที่ได้จากรอบก่อนหน้า ซึ่งกำหนดให้มีความยาวเฉลี่ยสูงสุดทั้งหมด *L* ตัวซึ่งเป็นกรณีเลวร้ายที่สุด และไม่อาจจะเปรียบได้ในเทอมของ n เพราะว่าค่า *L* ที่ยาวมากขึ้นจะมาจากวากยสัมพันธ์ของสายอักขระที่นำมาเรียนรู้และไม่เกี่ยวเนื่องกับค่า n และในแต่ละรอบที่พิจารณาสถานะแอคทีฟ เราพบว่าอาจจะมีการเปลี่ยนแปลงสมมติฐานโดยที่อัตราการเพิ่มของเวลาในส่วนที่เปลี่ยนแปลงสมมติฐานนี้ก็ขึ้นอยู่กับวากยสัมพันธ์ของข้อมูลเช่นกันซึ่งไม่เกี่ยวเนื่องกับค่า n แต่อย่างใด เรากำหนดให้จำนวนสมาชิกใน *Lnew* แทนด้วย *K* จากการวนลูปพิจารณาอักขระทุกตัวจำนวน n ครั้งและในแต่ละอักขระจะมีการพิจารณาสถานะมากที่สุด *L* ตัวซึ่งตั้งแต่ 1- *L* จะมีการพิจารณาการเปลี่ยนสมมติฐานซึ่งกรณีเลวร้ายสุดที่ต้องพิจารณาขึ้นอยู่กับจำนวนของ *Lnew* เราจึงสรุปว่าอัตราเร็วของอัลกอริทึมมีการเติบโตไม่เกิน O(*LKn*) ซึ่งเป็นฟังก์ชั่นเชิงเส้นเทียบกับความยาวรวมของสายอักขระที่นำมาเรียนรู้

กรณีที่ 2 ในส่วนของอัตราการเติบโตของไวยากรณ์ เราจะพิจารณาที่สถานะแอคทีฟว่าจะมีการเติบโตในแต่ละรอบเป็นเช่นไร จากการวิเคราะห์อัลกกอริทึม DO^{II} จะเห็นได้ว่าสถานะแอคทีฟนั้นมีเพิ่มจำนวนบ้างลดจำนวนบ้างไม่แน่นอน ตามแต่วากยสัมพันธ์ของสายอักขระที่นำมาเรียนรู้ดังนั้นในส่วนนี้เราจะวิเคราะห์ในกรณีที่เลวร้ายที่สุดโดยให้ *L* แทนจำนวนสถานะแอคทีฟที่ต้องพิจารณา โดยในแต่ละรอบหากมีการเพิ่มสถานะใหม่ที่มีชื่อด้วยสายอักขระที่มีความยาว 2 อักขระขึ้นไปจะมีการเพิ่มเพียงสถานะเดียว ตรงนี้จะทำให้อัตราการเพิ่มการส่งผ่านมากที่สุด *L*-(1+ k) ตัวเมื่อ k คือจำนวนรวมของสถานะเริ่มต้น สถานะไม่วนกลับ และสถานะที่ไม่มีการเพิ่มการส่งผ่านใหม่ อีกในกรณีหนึ่งคือ การเปลี่ยนสมมติฐาน หากไวยากรณ์ที่มีอยู่ทั้งหมดเทียบเป็นสถานะได้ *q* สถานะแล้ว การเพิ่มขึ้นของไวยากรณ์ที่จะสามารถเปลี่ยนแปลงได้ด้วยการโคลน และการสร้างสถานะใหม่ที่มากที่สุดก็เท่ากับขนาดของไวยากรณ์ *q* เราสรุปว่าอัตราการโตขึ้นของไวยากรณ์ในกรณีมากที่สุดไม่เกิด O(*qL*) เราสรุปได้ว่าอัตราการเติบโตของไวยากรณ์ q นั้นเติบโตเป็นฟังก์ชันเชิงเส้น

■

จากทฤษฎีที่ผ่านมาทั้งหมดทำให้เราสรุปปการเรียนรู้เพื่อวิวัฒนาการของการเรียนรู้ทางเทคโนโลยีคอมพิวเตอร์ในปัจจุบันได้ดังนี้

สรุป *สายอักขระใดๆ สามารถนำมาเรียนรู้เพื่อให้ได้มาซึ่ง สโทแคสติกออโตมาตาด้วยกระบวนการแบบเชื่อมตรงได้โดยไม่จำกัดโครงสร้างของแบบจำลองและเซตของอักขระที่เป็นไปได้ ด้วยการลู่เข้าของความน่าจะเป็นที่ถูกต้องซึ่งคำนวณได้จากสโทแคสติกออโตมาตา ณ ตำแหน่งที่อนันต์ โดยใช้อัลกอริทึม DO″ ซึ่งมีประสิทธิภาพการเรียนรู้ในเชิงเวลาและการเติบโตของไวยากรณ์ที่เติบโตไม่เกินฟังก์ชันพหุนาม*

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

สรุปผลการวิจัยและข้อเสนอแนะ

4.1 สรุปผลการวิจัย

ในงานวิจัยนี้ได้เสนอวิธีการเรียนรู้แบบใหม่ ซึ่งสนใจการเรียนรู้ด้วยกระบวนการเชื่อมตรงจากสายอักขระความยาวต่อเนื่องโดยไม่จำกัดจำนวนหรือความยาวสายอักขระ ไม่จำกัดจำนวนอักขระที่เป็นไปได้ และไม่กำหนดโครงสร้างเบื้องต้นใดๆ รวมถึงการไม่กำหนดขอบเขตจำนวนครั้งในการเรียนรู้ โดยอาศัยกระบวนทัศน์การกำหนดความสำเร็จในการเรียนรู้เพื่อให้ได้มาซึ่งโครงสร้างของแบบจำลองด้วยหลักการจำแนกภายในจำกัด และการกำหนดความถูกต้องของความน่าจะเป็นที่ได้จากการคำนวณด้วยโครงสร้างแบบจำลองนั้นกระทำด้วยกระบวนทัศน์การเรียนรู้แบบเชื่อมตรง จากการวิเคราะห์ทางทฤษฎีได้แสดงให้เห็นแล้วว่า ภาษาที่ถูกออกแบบในงานวิจัยนี้นั้นสามารถจำแนกได้ภายในจำกัดจริง รวมถึงความน่าจะเป็นที่คำนวณได้จากภาษานั้นก็สามารถลู่เข้าสู่ค่าที่ถูกต้องได้จริงตามทฤษฎี ซึ่งแสดงให้เห็นจริงตามบทพิสูจน์ อีกทั้งอัลกอริทึมที่เสนอเพื่อใช้ในการเรียนรู้นั้นมีประสิทธิภาพในการเรียนรู้เชิงเวลาเป็นฟังก์ชันเชิงเส้นเทียบกับความยาวรวมของสายอักขระที่นำมาเรียนรู้เท่านั้น เราสรุปว่าวิธีการที่นำเสนอนี้สามารถนำเอาไปเรียนรู้ด้วยการอุปนัยเพื่อใช้งานได้จริงตามทฤษฎี ในส่วนของการนำแบบจำลองที่ได้ภายหลังจากการเรียนรู้ไปใช้ทำนายข้อมูลต่างๆ สามารถนำเอาสมการที่เราได้นำเสนอไปใช้ได้จริงโดยอาศัยหลักการประมาณค่าด้วยหลักการลูกโซ่มาร์คอฟ อีกทั้งแบบจำลองที่ได้จากการเรียนรู้สามารถเรียนรู้ไปด้วยและทำนายข้อมูลไปด้วยได้ในเวลาพร้อมๆกัน เนื่องจากเวลาที่ใช้ในการเรียนรู้นั้นค่อนข้างเร็ว โดยที่ความน่าจะเป็นที่สามารถคำนวณได้ ณ เวลาใดๆ นั้นยังคงความถูกต้องด้วยการประมาณค่าเพื่อลดความคลาดเคลื่อนให้มากที่สุด ซึ่งความคลาดเคลื่อนนี้จะมากหรือน้อยนั้นขึ้นอยู่กับตัวข้อมูลที่นำมาเรียนรู้เป็นหลัก

จากทฤษฎีที่เรานำเสนอนั้นเห็นได้ว่าวิธีการของเรามีประสิทธิภาพทั้งด้านเวลาและประสิทธิภาพเชิงขนาดซึ่งพิสูจน์ให้เห็นจริงตามทฤษฎีบทที่ 3.21 โดยที่ประสิทธิภาพทางด้านเวลานั้น เราสามารถหาอัลกอริทึมที่ใช้สำหรับการเรียนรู้โดยที่ประมาณเป็นฟังก์ชันเวลาได้ภายในกำหนดไม่เกินฟังก์ชันพหุนาม $O(LKn)$ โดยที่ n คือความยาวรวมของสายอักขระข้อมูลที่ใช้ในการเรียนรู้ L คือความยาวของสายอักขระที่ยาวที่สุดซึ่งเป็นสมาชิกในภาษา K หมายถึงจำนวนสถานะที่มากที่สุดที่มีการส่งผ่านไปยังสถานะที่ต้องเปลี่ยนสมมติฐาน ทางด้านประสิทธิภาพเชิงขนาดนั้นเราพิสูจน์ให้เห็นว่า การเพิ่มขนาดของแบบจำลองนั้นเติบโตในแต่ละขั้นไม่เกิน $O(qL)$

เมื่อ q คือขนาดไวยากรณ์ที่ได้ในแต่ละรอบ L หมายถึงจำนวนสถานะที่แอคทิฟที่มากที่สุด ในส่วนของประสิทธิภาพเชิงข้อมูลที่ได้นั้น เรายังคงคำนึงถึงแต่ประสิทธิภาพทางด้านจำนวนครั้งของสมมติฐานที่เปลี่ยน โดยเราได้พบว่าในแต่ละสถานะที่ต้องถูกเปลี่ยนสมมติฐานนั้น สามารถเปลี่ยนได้เพียงครั้งเดียว ซึ่งจะทำให้สมมติฐานลู่เข้าสู่ค่าที่ถูกต้องแน่นอน สำหรับประสิทธิภาพข้อมูลในส่วนของความครอบคลุมและความกระชับของข้อมูลนั้น งานวิจัยนี้ยังไม่สามารถทำให้ดีกว่าการเรียนรู้แบบออฟไลน์ได้ เนื่องจากวิธีการเรียนรู้ที่แตกต่างกันนั้นจำกัดความสามารถในส่วนนี้

4.2 ข้อเสนอแนะ

เนื่องจากงานวิจัยนี้เป็นการเสนออัลกอริทึมในการเรียนรู้แบบใหม่ พบว่าในแต่ละส่วนยังสามารถทำการวิจัยต่อเพื่อพัฒนาให้งานวิจัยชิ้นนี้มีความสมบูรณ์มากขึ้นโดยแยกเป็นแต่ละข้อดังนี้

4.2.1 ข้อเสนอแนะในการพัฒนาด้านทฤษฎีการเรียนรู้

ในงานวิจัยนี้ ได้พบว่าการลดความคลาดเคลื่อนของโครงสร้างและความน่าจะเป็นในขณะทำการเปลี่ยนสมมติฐานนั้นอาศัยเพียงหลักการประมาณจากองค์ความรู้เดิมจากแบบจำลอง โดยที่ยังไม่มีการกำหนดขอบเขตของความคลาดเคลื่อน อีกทั้งยังไม่มีการเสนอวิธีการที่ประกันได้ว่าสามารถลดความคลาดเคลื่อนได้มากที่สุดจริง ซึ่งต้องอาศัยความรู้และเวลาจากงานวิจัยต่อไปในการพัฒนาส่วนนี้ ซึ่งหากว่าเราสามารถลดความคลาดเคลื่อนในส่วนของโครงสร้างได้จะทำให้งานวิจัยนี้สมบูรณ์ยิ่งขึ้น อีกทั้งงานวิจัยนี้ยังไม่ได้มีการพิสูจน์ว่าอัลกอริทึมที่ใช้ในการเรียนรู้นั้นเป็นอัลกอริทึมที่มีประสิทธิภาพมากที่สุด เราเสนอว่าหากมีอัลกอริทึมที่ดีกว่านี้แล้วทำการพิสูจน์ว่ามีประสิทธิภาพมากที่สุด จะทำให้งานวิจัยนี้มีความสมบูรณ์มากยิ่งขึ้น

4.2.2 ข้อเสนอแนะในการพัฒนาด้านทฤษฎีในการทำนาย

ในการเรียนรู้ด้วยแบบจำลองการส่งผ่านทุกวิธีมักจะมีปัญหาในการนำเอาไปใช้งานจริงในการทำนายนั่นก็คือ การทำนายเหตุการณ์ที่ไม่เคยเกิดขึ้นมาก่อนได้ จากวิธีการที่นำเสนอก็คือ การประมาณจากแบบจำลองเดิมด้วยสมการที่ปรับเปลี่ยน แทนที่จะปรับเปลี่ยนที่พารามิเตอร์ของแบบจำลองเดิม ด้วยวิธีการนี้ยังไม่ผ่านการวิเคราะห์ที่ลดความคลาดเคลื่อนให้มากที่สุด ซึ่งเราเสนอว่าน่าจะมีวิธีการที่ดีกว่าสมการที่นำเสนอนี้ ซึ่งจำเป็นต้องอาศัยเวลาและความรู้จากงานวิจัยต่อไปในการทำให้ส่วนนี้สมบูรณ์มากยิ่งขึ้น

4.2.3 ข้อเสนอแนะในการพัฒนาเพื่อการนำเอาไปใช้งานจริง

เนื่องจากทฤษฎีการเรียนรู้ที่นำเสนอนี้ยังไม่มีการนำเอาไปใช้งานจริง ซึ่งในการใช้งานจริงนั้นจะมีตัวแปรอื่นๆเข้ามาอีกมากมาย เราเสนอให้มีงานวิจัยต่อไปทางทฤษฎีเพื่อระบุตัวแปรเหล่านั้นให้ชัดเจนก่อนที่จะนำเอาไปใช้งานจริง ซึ่งในความเป็นจริงแล้วมุมมองในการเรียนรู้เชิงตลอดชีพนี้มีตัวแปรที่หลากหลายมาก ยกตัวอย่างเช่น การตั้งค่าพ้นขีดที่จำเป็นต่อการทำนาย วิธีการนำเอาข้อมูลที่ทำนายว่าถูกต้องกลับนำเอามาเรียนรู้อีก ระยะเวลาในการเรียนรู้ที่สามารถประกันได้ว่าสามารถตอบคำถามได้ และอื่นๆ อีกมากมาย จำเป็นต้องมีการระบุตัวแปรเหล่านี้ให้ชัดเจนรวมถึงแนวทางที่จะจัดการกับตัวแปรเหล่านี้ในแต่ละสภาพข้อมูล

สรุปสุดท้ายว่าการนำเอาทฤษฎีนี้ทำให้เกิดเป็นรูปธรรมที่ชัดเจนขึ้นมาจึงจะเป็นงานวิจัยที่สำคัญและมีประโยชน์ต่อวงการปัญญาประดิษฐ์อย่างแท้จริง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

1. Vidal, E., Thollard, F., De la Higuera, C., Casacuberta, F. and Carrasco, R. C. Probabilistic finite state automata part I. IEEE Transactions on Pattern Analysis and Machine Intelligence (2004)
2. Vidal, E., Thollard, F., De la Higuera, C., Casacuberta, F. and Carrasco, R. C. Probabilistic finite state automata part II. IEEE Transactions on Pattern Analysis and Machine Intelligence (2004)
3. Stenger, B., Ramesh, V., Paragios, N., Coetzee, F. and Buhmann, J. Topology free hidden markov models: Application to background modeling. proceedings of the 8th International Conference on Computer vision 1(2001) : 294-301.
4. Gold, E. M. Language identification in the limit. Information and Control 10(5) (1968): 447–474.
5. Angluin, D. and Smith, C. H. Inductive inference: Theory and methods. Computing Surveys 15(3) (September 1983) : 237-269.
6. Carrasco, R. and Oncina, J. Learning stochastic regular grammars by means of a state merging method. Lecture Notes in Artificial Intelligence (1994) : 139-150.
7. Carrasco, R. and Oncina, J. Learning deterministic regular grammars from stochastic samples in polynomial time. *RAIRO*(Theoretical Informatics and Applications) 33 (1999) : 1-20.
8. Ron, D., Singer, Y. and Tishby, N. The Power of Amnesia. Advance Neural Information Processing Systems 6 (1993)
9. Dupont, P., Denis, F. and Esposito, Y. Link between probabilistic automata and hidden markov models: Probability distributions, learning models and induction algorithms. Pattern Recognition (2004)
10. Bottou, L. Online learning and stochastic approximations. AT&T Labs-Research Red Bank NJ07701.
11. Ron, D., Singer, Y. and Tishby, N. The power of amnesia: Learning probabilistic automata with variable memory length. Machine learning *25* (1996): 117-149.
12. Ron, D., Singer, Y. and Tishby, N. The Power of Amnesia. Advance Neural Information Processing Systems 6 (1993)

13. Thollard, F., Dupont, P. and De la Higuera, C. Probabilistic DFA Inference using Kullback-Leibler Divergence and Minimality. 17th International Conference on Machine Learning (2000) : 975-982.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นายจิตรกร พูลโพธิ์ทอง เกิดเมื่อวันที่ 29 ตุลาคม พ.ศ. 2519 ที่จังหวัดกาญจนบุรี จบการศึกษาระดับชั้นมัธยมศึกษาตอนต้นจากโรงเรียนวีรศิลป์ อำเภอท่าม่วง จังหวัดกาญจนบุรี จบการศึกษาระดับมัธยมศึกษาตอนปลายจากโรงเรียนสารสิทธิ์พิทยาลัย อำเภอบ้านโป่ง จังหวัดราชบุรี สำเร็จการศึกษาระดับปริญญาบัณฑิตสาขาวิศวกรรมคอมพิวเตอร์จากมหาวิทยาลัยมหิดล วิทยาเขตศาลายา จังหวัดนครปฐม เข้ารับการศึกษาระดับปริญญามหาบัณฑิตสาขาวิชา วิศวกรรมคอมพิวเตอร์ที่ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2546

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย