บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

วิทยานิพนธ์นี้นำเสนอวิธีการประยุกต์ใช้ทฤษฎีการเรียนรู้ของเครื่อง (Machine Learninig) เพื่อการสร้างระบบเลียนแบบผู้สอนในการเรียนการสอนแบบสนทนาจากบันทึกการสอนในภาษาไทย โดยมีจุดประสงค์เพื่อหารูปแบบหรือปัจจัยที่ส่งผลในการเลือกใช้บทสนทนาของผู้สอน ซึ่งต้องใช้ทฤษฎีพื้นฐานที่เกียวข้องในงานวิจัยดังนี้

2.1.1 ทฤษฎีการเรียนรู้ของเครื่อง (Machine Learning)

การศึกษาด้านการเรียนรู้ของเครื่อง (Mitchell, 1997) มีจุดมุ่งหมายเพื่อทำให้โปรแกรมคอมพิวเตอร์สามารถพัฒนาความสามารถของตัวเองโดยอัตโนมัติจากตัวอย่างที่ผู้สอนป้อนให้ ตัวอย่าง เช่น โปรแกรมคอมพิวเตอร์ซึ่งสามารถเรียนรู้เพื่อเล่นหมากรุก สามารถพัฒนาความสามารถในการเล่นซึ่งวัดจากเปอร์เซ็นในการชนะคู่แข่งขัน โดยใช้ตัวอย่างซึ่งได้จากการเล่นกับตัวเอง หรือการสร้างโปรแกรมที่สามารถรู้จำภาพตัวเขียนอักษรได้ ซึ่งวัดความสามารถในการรู้จำจากเปอร์เซ็นต์ของภาพตัวอักษรที่รู้จำได้ถูกต้อง โดยใช้ตัวอย่างเป็นภาพตัวอักษรซึ่งผู้สอนป้อนให้

ได้มีการนำวิธีการเรียนรู้ของเครื่องไปใช้เป็นส่วนประกอบสำคัญในการสร้างโปรแกรมเพื่อใช้งานทั่วไป รวมทั้งมีการประยุกต์ใช้กับงานด้านต่างๆ เช่น การสร้างโปรแกรมประยุกต์เพื่อประโยชน์ในด้านธุรกิจการอนุมัติบัตรเครดิต การค้นพบความรู้จากฐานข้อมูล (Knowledge Discover from Database) การใช้งานทั่วไป เช่น การรู้จำเสียงพูด การรู้จำภาพ ฯลฯ ตัวอย่างของการประยุกต์ใช้ซึ่งนำวิธีการเรียนรู้ของเครื่องไปใช้งานได้อย่างดี ได้แก่ การรู้จำเสียงพูด (Waibel, Hanazawa, Hinton, Shikano and Lang, 1989) การทำนายอัตราการฟื้นตัวของผู้ป่วยโรคปอดอักเสบ การสืบหากรณีที่เจ้าของบัตรเครดิตชำระเงินไม่ตรงเวลา การบังคับรถโดยอัตโนมัติบนถนนหลวง (Pomerleau, 1989) หรือการสร้างกลยุทธ์ต่างๆ ในการเล่นเกม (Tesauro, 1992) ฯลฯ ในทางทฤษฎีก็เช่นกัน มีงานวิจัยจำนวนมากได้พัฒนาวิธีการเรียนรู้ของเครื่องให้มีประสิทธิภาพมากยิ่งขึ้น ตัวอย่างของการพัฒนาเหล่านี้ ได้แก่ การหาความสัมพันธ์

พื้นฐานของตัวอย่างที่ป้อนให้เพื่อทำการเรียนรู้ การหาจำนวนสมมติฐานที่เหมาะสมควรนำมาพิจารณา การคาดหมายความผิดพลาดของสมมติฐานต่างๆ การสร้างแบบจำลองของการเรียนรู้ของมนุษย์และสัตว์ รวมทั้งการทำความเข้าใจความสัมพันธ์ของการเรียนรู้ของมนุษย์และสัตว์กับขั้นตอนวิธีการเรียนรู้สำหรับเครื่องคอมพิวเตอร์ (Chi and Bassock, 1989)
ตัวอย่างของการนำวิธีการเรียนรู้ของเครื่องไปใช้งานได้อย่างมีประสิทธิภาพ ดังนี้

- การเรียนรู้เพื่อการรู้จำเสียงพูด มีการใช้วิธีการเรียนรู้ของเครื่องเป็นองค์ประกอบสำคัญ ส่งผลให้ระบบที่รู้จำเสียงพูดมีประสิทธิภาพสูงขึ้น โดยนำวิธีการเรียนรู้ของเครื่องไปใช้ในรูปแบบต่าง ๆ เช่นระบบ SPHINX (Lee, 1989) เรียนรู้เพื่อการรู้จำองค์ประกอบพื้นฐานของเสียงพูด (phonemes) และคำศัพท์ (words) จากสัญญาณเสียง รวมถึงการใช้นิวรอลเน็ตเวิร์กและจำลองฮิดเด็นมาร์คอฟ (Hidden Markov Model) (waibel, 1989) เพื่อรู้จำเสียงของผู้พูดแต่ละคน
- การเรียนรู้เพื่อบังคับพาหนะโดยอัตโนมัติ ใช้การเรียนรู้ของเครื่องเพื่อควบคุมรถยนต์ด้วยเครื่องคอมพิวเตอร์ สามารถควบคุมได้อย่างถูกต้องเมื่อขับเคลื่อนอยู่บนถนนหลายแบบ โดยเรียนรู้จากตัวอย่างที่ผู้สอนเป็นผู้ควบคุมรถ ดังตัวอย่างเช่น ระบบ ALVINN (Pomerleau, 1989) สามารถขับเคลื่อนรถยนต์ได้เองด้วยความเร็ว 70 ไมล์ต่อชั่วโมง เป็นระยะทาง 90 ไมล์ บนทางหลวง ร่วมกับรถยนต์คันอื่นซึ่งมีมนุษย์เป็นผู้บังคับ
- การเรียนรู้เพื่อจำแนกโครงสร้างใหม่ทางดาราศาสตร์ได้มีการนำวิธีการเรียนรู้ของเครื่องนำไปประยุกต์เพื่อค้นหาลักษณะพื้นฐานในระบบฐานข้อมูลขนาดใหญ่ เช่น องค์การบริหารการบินและอวกาศแห่งชาติสหรัฐอเมริกา หรือ NASA ได้ขั้นตอนวิธีการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree Learning Algorithm) เรียนรู้เพื่อจำแนกวัตถุที่อยู่บนท้องฟ้าจากภาพที่มีขนาดใหญ่มาก (Fayyad, Smyth, and Dijorgovski, 1995)
- การเรียนรู้เล่นเกมแบ็กแกมมอน (Backgammon) ซึ่งเป็นโปรแกรมคอมพิวเตอร์ที่ใช้เล่นเกมซึ่งประสบความสำเร็จมากที่สุด ได้แก่ โปรแกรมเล่นเกมแบ็กแกมมอน ซึ่งใช้วิธีการเรียนรู้ของเครื่องเป็นแนวคิดพื้นฐาน ตัวอย่างโปรแกรมเหล่านี้ได้แก่ TD-GAMMON (Tesauro, 1992) เป็นโปรแกรมที่มีความสามารถสูงมากในการเล่นเกมแบ็กแกมมอน สามารถหากลยุทธ์ในการเล่นโดยการฝึกเล่นกับตัวเองจนมีความสามารถเทียบเท่ากับมนุษย์ระดับแชมเปี้ยนโลก

2.1.2 **ระบบสอนเสริมอัจฉริยะ**(ITS : Intelligent Tutoring System)

ระบบสอนเสริมอัจฉริยะนั้นเป็นระบบที่พัฒนาต่อมาจากระบบการใช้คอมพิวเตอร์ช่วยสอน (CAI : Computer Aided Instruction) โดยระบบ CAI เดิมนั้นมีความเป็นสถิต(Static)สูง ไม่มีกลไกในการปรับเปลี่ยนกลยุทธการสอนและเนื้อหาให้เหมาะสมตามสถานการณ์และผู้เรียนได้ จึงมีการพัฒนาให้ระบบ CAI นี้มีส่วนที่เป็นกลไกในการเรียนรู้ผู้เรียนและสถานการณ์ เพื่อปรับกลยุทธ์การสอนให้เหมาะสมกับผู้เรียนและสถานการณ์ อันนำไปสู่จุดมุ่งหมายของการเรียนการสอนนั้นได้ดียิ่งขึ้น ซึ่งการเพิ่มส่วนที่มีการปรับเปลี่ยนกลยุทธ์การสอนและเนื้อหาให้เหมาะสมกับผู้เรียนและสถานการณ์ได้นี่เอง จึงได้ชื่อว่าเป็นระบบการสอนอัจฉริยะ (Intelligent Tutoring System)

Burns and Capps (1988) ได้แบ่งองค์ประกอบของระบบการสอนอัจฉริยะประกอบไปด้วย 3 องค์ประกอบ นั่นคือ

1. Domain Expert Module

   เป็นส่วนที่เกี่ยวกับความรู้ในเนื้อหาวิชาโดยตรง และความรู้นั้นได้ถูกแสดงแทน (Represent) อยู่ในระบบ เพื่อรอการดึงขยาย (extract) ไปนำเสนอแก่ผู้เรียนในรูปแบบต่างๆ ตามการตัดสินใจของระบบ ซึ่งส่วนนี้อาจเป็น Expert System ที่มีการพัฒนามาแล้วนำมาใช้ในระบบ ITS ได้โดยตรง

2. Student Module

   ส่วนนี้อาจเรียกอีกชื่อหนึ่งเป็น Student Model ได้ โดยเป็นส่วนที่คอยติดตามและแสดงสถานะของผู้เรียนในปัจจุบันแล้วแต่การออกแบบของแต่ละระบบการสอนว่าต้องการติดตามองค์ประกอบใดจากผู้เรียน เพื่อนำไปเป็นข้อมูลในการตัดสินใจใช้กลยุทธ์การสอนที่ถูกต้องต่อไป

3. Tutor Module

   เป็นส่วนของการจัดการด้านกลยุทธ์การสอนและการนำเสนอเนื้อหาให้เหมาะสมตามสถานการณ์และผู้เรียน โดยการตัดสินใจจาก Tutor Moudule นี้ได้รับข้อมูลมาจาก Student Model เป็นหลัก และมีจุดมุ่งหมายในการขยายความรู้ของ Student Model ให้ใกล้เคียงหรือเท่ากับ Expert Model มากที่สุด

ตัวอย่างของระบบการสอนแบบอัจฉริยะ

1. ระบบ GUIDON (Clancey 1983)

GUIDON ใช้ MYCIN expert system (Shortliffe 1976) ซึ่งเก็บความรู้ด้านการติดเชื้อของแบคทีเรีย เป็น Domain Expert Module โดยตรง และตัว GUIDON จะทำหน้าที่ในการถามปัญหาราวกับว่าผู้เรียนเป็น MYCIN และเปรียบเทียบคำตอบระหว่างผู้เรียนกับ MYCIN ว่ามีการตัดสินใจในแบบเดียวกันหรือไม่ โดยระบบจะไม่แทรกแซงผู้เรียนจนกว่าผู้เรียนจะขอให้ช่วย หรือคำตอบของผู้เรียนไม่ถูกต้องครบถ้วน

2. ระบบ LISP TUTOR (Anderson and Reiser 1985)

LISP TUTOR เป็น ITS ที่ใช้ในการสอนหลักการการเขียนโปรแกรมด้วยภาษา LISP โดย Expert model ของ LISP TUTOR คือชุดของ production rules ที่ถูกต้องทั้งหมด และ Student model นั้นคือ subset ของชุด production rules ที่ถูกต้องรวมไปถึง production rules ที่มักจะผิดพลาดอยู่เสมอ (Holt et al 1991). โดย LISP TUTOR ใช้หลักการเรียนรู้โดยปฏิบัติ (Learning by Doing) ให้ผู้เรียนค้นพบการสร้าง production rules ที่ถูกต้องผ่านการเผชิญกับปัญหา โดยตัวระบบจะเป็นเพียงการแนะนำแนวทางแก้ปัญหาเท่านั้น

2.1.3 **การตัดคำ** (Word Segmentation)

การตัดคำ (Word Segmentation) เป็นการแบ่งข้อความที่เรียงต่อเนื่องกันออกมาเป็นหน่วยคำเพื่อหาขอบเขตแต่ละหน่วยคำ ซึ่งวิธีการตัดคำจะขึ้นอยู่กับลักษณะของภาษานั้นๆ สำหรับลักษณะของการเขียนภาษาไทยจะมีลักษณะการเขียนที่ต่อเนื่องกันไม่สามารถหาจุดจบหรือขอบเขตของคำได้ ซึ่งจะต่างจากลักษณะการเขียนภาษาอังกฤษที่มีการเว้นวรรคคำอย่างชัดเจนและสามารถหาจุดจบของคำได้ ดังนั้นจึงได้มีงานวิจัยเกี่ยวกับการตัดคำจำนวนมาก อาทิ เช่น วิธีการตัดคำแบบยาวที่สุด (Longest Matching) (วิรัช ศรเลิศล้ำวาณิช, 2536) วิธีการตัดคำแบบสอดคล้องมากที่สุด (Maximal Matching) (วิรัช ศรเลิศล้ำวาณิช,2536) วิธีการตัดคำแบบคำนวณเชิงสถิติเพื่อหาความเป็นไปได้ (Probabilistic Model) (Asanee Kawtraku,1995) วิธีการตัดคำแบบใช้คุณลักษณะ (Feature – Based Approach) (Meknavin, Charoenpornsawat,and Kijsirikul, 1997) เป็นต้น ซึ่งในงานวิจัยนี้จะอธิบายถึง วิธีการตัดคำแบบยาวที่สุด (Longest Matching) เพื่อนำมาใช้สำหรับระบบถามตอบอัตโนมัติภาษาไทย

สำหรับงานวิจัยนี้ได้เลือกใช้โปรแกรมตัดคำSWATH (ศูนย์วิจัยและพัฒนาเทคโนโลยีสารสนเทศ, 2545) ซึ่งโปรแกรมตัดคำนี้จะใช้วิธีการตัดคำแบบยาวที่สุด (Longest Matching) สำหรับช่วยตัดคำในภาษาไทย ซึ่งการตัดคำโดยวิธีเทียบคำที่ยาวที่สุดจะอาศัยฐานข้อมูลพจนานุกรมสำหรับอ้างอิงเพื่อหาจุดจบของคำ โดยเริ่มจากตัวอักษรซ้ายสุดของข้อความนั้นไปยังตัวอักษรถัดไปจนกว่าจะพบคำที่มีอยู่ในพจนานุกรม หลังจากนั้นก็ค้นหาคำต่อไปจนกว่าจะจบข้อความ ในกรณีที่พบว่าในพจนานุกรมจากจุดเริ่มต้นเดียวกันเราจะเลือกคำที่ยาวที่สุด ตัวอย่างเช่น การแบ่งคำในประโยค “ฉันนั่งตากลมที่หน้าบ้าน” จะเริ่มจากตัวอักษร ฉ และคำแรกที่แบ่งได้คือ “ฉัน” หลังจากนั้น ก็ค้นหาตัวอักษรถัดไปและนำมาเปรียบเทียบคำในพจนานุกรม ก็จะแบ่งคำว่า “นั่ง” เป็นคำต่อไป ตัวอักษรถัดไปคือ ต จากตัวอักษรนี้ เราจะได้คำว่า “ตา” กับคำว่า “ตาก” แนวคิดนี้ให้เลือกคำที่ยาวที่สุดที่ค้นพบจึงเลือกว่า “ตาก” หลังจากนั้ก็จะค้นหาและเปรียบเทียบต่อไป ซึ่งจะได้ผลลัพธ์ของการแบ่งคำออกมาเป็นดังนี้ “ฉัน นั่ง ตาก ลม ที่ หน้า บ้าน” แต่การตัดคำโดยใช้วิธีนี้ก็ยังพบข้อผิดพลาดของการที่เลือกคำที่ยาวที่สุดตั้งแต่แรก ส่งผลกระทบให้ตัดคำอื่นๆ ผิดพลาดไปด้วย เช่น “ไปหามเหสี” สามารถแบ่งคำได้เป็น “ ไป I หาม I เห I สี “ซึ่งในกรณีนี้คำว่า “หาม” จะมีความยาวของสายอักขระมากกว่า คำว่า “หา” จึงทีทำให้การตัดคำอื่นๆ ผิดพลาดไปด้วย (นัฐวุฒิ ไชยเจริญ, 2544)

2.1.4 **การจัดกลุ่มเนื้อหา**(Text Clustering)

การจัดกลุ่มเนื้อหาข้อความ(Text Clustering) เป็นการสร้างกลุ่มให้กับเนื้อหาข้อความในแต่ละหน่วยที่มากกว่าคำ เช่น ประโยค(sentence), ย่อหน้า(paragraph), ส่วนของเอกสาร (section), บทความ (article) เพื่อแบ่งประเภทและรวมกลุ่มหน่วยหาข้อความนั้นเป็นหมวดหมู่ เพื่อการจัดการกับเนื้อหาในแต่ละหมวดหมู่ได้อย่างถูกต้องต่อไป ซึ่งขึ้นอยู่กับงานที่จะใช้

โดยทั่วไป การจัดกลุ่มเอกสารจะใช้การที่เอกสารแต่ละฉบับนั้นมีคำที่ใช้ร่วมกันมากหรือน้อยเท่าใด เพราะเราสามารถสังเกตได้ว่าเอกสารประเภทเดียวกันนั้น จะใช้คำที่คล้ายกันหรือมีคำที่เหมือนกันอยู่ในเอกสารในจำนวนที่สามารถแยกความแตกต่างได้ ดังตัวอย่างการจัดกลุ่มเอกสารดังตารางที่ 2.1

ตารางที่ 2.1

ตัวอย่างการจัดกลุ่มของบทความ

Doc	Planet	Galaxy	Nova	Film	Role	Hollywood	Diet	Habitat	Fur
A	2	1	1						
B	1	1	3						
C	2	1							
D	1	2	1						
E				1	1	1			
F				3		1			
G				1	5	2			
H				2	1	1			
I							2	3	2
J							1	1	3
K							4	1	
L							1	1	1
M	2	1	1	1	2				
N				2	1	1	1		1

จากภาพที่ 2.1 เราแยกเอกสารบทความได้เป็นบทความด้านอวกาศ บทความทางด้านภาพยนต์ และบทความด้านชีวิตสัตว์ โดยดูจากกลุ่มคำที่แตกต่างกันในแต่ละชนิดของบทความเช่น กลุ่ม ABCD นั้น เป็นบทความด้านอวกาศเนื่องจากมีคำว่า “Planet” , “Galaxy” , “Nova” อยู่ภายในเอกสารเดียวกันเป็นต้น

ข้อพิจารณาในการเลือกใช้กระบวนการจัดกลุ่ม (Clustering Taxonomy)

1. แบบลำดับชั้น (Hierarchical) หรือ แบบกั้นแบ่ง (Partitional)

   แบบลำดับชั้นสามารถมีการแบ่งกั้นย่อยได้หลายระดับในขณะที่แบบกั้นแบ่งนั้น มีได้เพียง 1 ระดับเท่านั้น

2. แบบรวมเข้า (Agglomotive) หรือ แบบแบ่งออก (Divisive)

   แบบรวมเข้าเป็นการเริ่มต้นโดยจัดทุกหน่วยเป็นกลุ่มละ 1 หน่วยก่อน แล้วจึงรวมกลุ่มเหล่านั้นที่มีความคล้ายคลึงกันเข้าด้วยกันเพิ่มขึ้นเรื่อยจนครบตาม

เงื่อนไขที่ตั้งไว้ในการจัดกลุ่ม ส่วนในแบบแบ่งออกนั้นให้รวมทุกหน่วยเป็น 1 กลุ่ม ก่อนแล้วแยกกลุ่มเป็นกลุ่มย่อย จนครบตามเงื่อนไข

3. แบบใช้ Feature เดียว (Monothetic) และหลาย Feature ในการพิจารณา

แบบ Feature เดียวคือใช้องค์ประกอบ (Attribute) เดียวในการพิจารณาจัดกลุ่ม ส่วนแบบหลายองค์ประกอบคือ การใช้หลายองค์ประกอบ (Attribute) ในการพิจารณาจัดกลุ่ม ซึ่งโดยทั่วไปแล้วจะเป็นแบบใช้หลาย Feature ในการพิจารณาจัดกลุ่ม

4. แบบ Hard หรือ แบบ Fuzzy

แบบ Hard จะเป็นการแบ่งอย่างชัดเจนว่าข้อมูลนั้นอยู่กลุ่มใด แต่แบบ Fuzzy นั้นจะใช้ค่า degree เพื่อบ่งบอกค่าความเข้ากับกลุ่มนั้นๆ ได้มากน้อยเพียงใด

2.1.5 **การเรียนรู้ด้วยนาอีฟเบย์ (Naive Bayes)**

Naive Bayes (NB) คือโมเดลการจัดกลุ่มที่ใช้หลักความน่าจะเป็นซึ่งอยู่บนพื้นฐานของ Bayes' Theorem และสมมติฐาน ที่กำหนดให้การเกิดของเหตุการณ์ต่างๆที่ใช้ในการจัดกลุ่มนั้นเป็นอิสระต่อ กัน (independence) ซึ่งการเรียนรู้จำแนกด้วยกระบวนการของ Naïve Bayes นี้ได้ถูกใช้อย่างแพร่หลายในงานวิจัยด้าน Machine Learning เหตุที่ Naïve Bayes ได้รับความนิยมเนื่องจากมีการทำงานที่ไม่ซับซ้อน แต่ทำงานได้อย่างมีประสิทธิภาพ เช่น งานวิจัยด้านการจำแนกเอกสาร (Text Categorization) ประเภทต่างๆ เป็นต้น

Bayes' Theorem

โดยกำหนดให้ P(H) ความน่าจะเป็นที่จะเกิดเหตุการณ์ H และ P(H|E) คือความน่าจะเป็นที่จะเกิดเหตุการณ์ H เมื่อเกิดเหตุการณ์ E จากตัวแปรที่กำหนดและแนวคิดของ Bayes' Theorem นั้นเราสามารถทำนายเหตุการณ์ที่พิจารณาได้จากการเกิดของเหตุการณ์ต่างๆ ได้ดังสมการ

$$P(H|E) = [P(E|H)xP(H)]/P(E)$$

ตัวอย่าง เช่น การทำนายว่าฝนจะตกเมื่อมี เหตุการณ์มีเมฆดำ กำหนดให้ H คือเหตุการณ์ที่ฝนตก และ E คือเหตุการณ์มีเมฆดำ แล้วเราจะสามารถทำนายสภาพจราจรได้ดังสมการต่อไปนี้

$$P(\text{ฝนตก} \mid \text{เมฆดำ}) = [P(\text{เมฆดำ} \mid \text{ฝนตก}) \times P(\text{ฝนตก})]/P(\text{เมฆดำ})$$

กำหนดให้

- P(เมฆดำ | ฝนตก) คือความน่าจะเป็นที่มีเมฆดำเมื่อฝนตก ซึ่งในกรณีนี้เราจะพิจารณาการเกิดเมฆดำเมื่อมีเหตุการณ์ฝนตกเท่านั้น
- P(ฝนตก) คือความน่าจะเป็นที่ฝนจะตก ความน่าจะเป็นนี้สามารถเก็บรวบรวมโดยใช้หลักการณ์ทางสถิติ เช่น การบันทึกวันที่มีฝนตกภายใน 1 ปี
- P(เมฆดำ) คือความน่าจะเป็นที่มีเมฆดำ เช่นเดียวกันความน่าจะเป็นนี้สามารถเก็บรวบรวมโดยใช้หลักการณ์ทางสถิติ แต่อย่างไรก็ตามการเกิดของเหตุการณ์ต่างๆที่ใช้ในการจัดกลุ่มซึ่งไม่ใช่ เหตุการณ์หลักที่เราพิจารณามักจะไม่ถูกบันทึกไว้ และในบางกรณีเหตุการณ์เหล่านี้ยากต่อการบันทึก เช่น P(เมฆดำ) เป็นต้น

จากตัวอย่างที่กล่าวมานั้นเราสามารถทำนายเหตุการณ์โดยสังเกตุการเกิดของเหตุการณ์บางอย่าง ซึ่งเหตุการณ์ที่นำมาใช้ในการทำนายนั้นต้องสอดคล้องกับเหตุการที่จะทำนาย เช่น ถ้าหากเราต้องการทำนายการตกของฝน เราก็จะไม่ใช้การเกิดแผ่นดินไหวมาพิจารณา เพราะการเกิดแผ่นดินไหวไม่ได้สอดคล้องกับการเกิดฝนตก

ในการจัดกลุ่ม Naive Bayes Model อาจมีการเกิดของเหตุการณ์ต่างๆที่ใช้ในการจัดกลุ่มมากกว่า 1 ชนิด และเมื่อนำมาประยุกต์ใช้กับ Bayes' Theorem จะทำให้เกิดการคำนวณที่ซับซ้อนเนื่องจากการขึ้นต่อกันของการเกิดของ เหตุการณ์ (dependence) ดังนั้น Naive Bayes Model นั้นจึงตั้งสมมติฐานให้แต่ละเหตุการณ์ที่ใช้ในการจัดกลุ่มนั้นเป็นอิสระต่อกัน ซึ่งเป็นที่มาของคำว่า Naive

เมื่อพิจารณา Bayes' Theorem ที่กล่าวมาในตอนต้นแล้วนั้นเราจะสามารถแสดงการคำนวณการจัดกลุ่มของเหตุการณ์ ที่มีการเกิดของเหตุการณ์ต่างๆที่ใช้ในการจัดกลุ่มมากกว่า 1 ชนิด ได้ดังสมการต่อไปนี้

$$P(H|E1,E2,...,En) = [P(E1,E2,....,En|H) \times P(H)]/P(E1,E2,....,En)$$

เมื่อกำหนดให้เหตุการณ์ E1, E2,....,En คือเหตุการณ์ n เหตุการณ์ที่ใช้ในการจัดกลุ่ม และจากสมมติฐานที่เรากำหนดให้แต่ละเหตุการณ์ต่างๆที่ใช้ในการจัดกลุ่มเป็น อิสระต่อกัน แล้วนั้น เราจะสามารถแสดงการคำนวณโดยใช้ Bayes' Theorem ได้ดังสมการต่อไปนี้

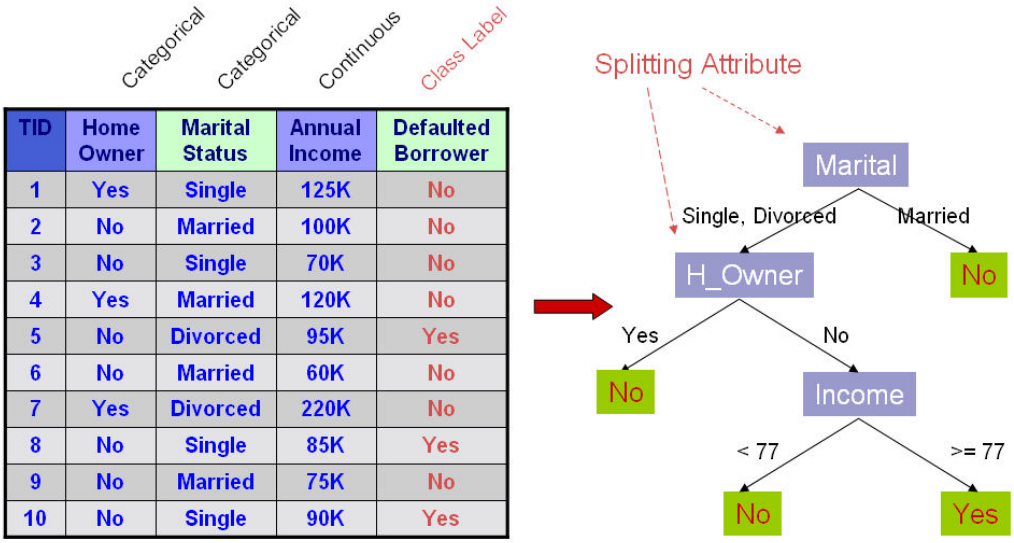
$$P(E1,E2,....,En|H) =$$
$$[P(E1|H) \times P(E2|H) \times .... \times P(En|H) \times P(H)]/P(E1,H) \times P(E2|H) \times .... \times P(En|H)$$

2.1.6 การเรียนรู้ด้วยต้นไม้ตัดสินใจ (Decision Tree)

Decision Tree หรือต้นไม้ตัดสินใจ เป็นกระบวนการเรียนรู้จำแนกโดยการแบ่งแยกข้อมูลออกเป็นแขนงสาขา ซึ่งค่อนข้างสอดคล้องกับการตัดสินใจจริงของมนุษย์และถูกแสดงเป็นภาพที่เข้าใจได้ง่าย และสามารถทำงานได้ดีกับข้อมูลที่มีความเกี่ยวพันของข้อมูลสูง และอาจมีหลายลำดับชั้น โดยมีอัลกอริทึ่มที่นิยมนำมาใช้ในการสร้าง Decision Tree อยู่หลายกระบวนการ เช่น Hunt's Algorithm, CART, ID3, C4.5 และ SLIQ เป็นต้น ซึ่งการสร้าง Decision Tree จากข้อมูลที่ต้องการทำการเรียนรู้จะต้องมีการกำหนด Attribute และการแบ่ง Node (Split) ของตัว Tree นั้น ซึ่งอาจเป็นไปตามข้อมูลที่ได้มา ดังภาพที่ 2.1

ภาพที่ 2.1

การสร้างโมเดล Decision Tree



สำหรับสิ่งที่ต้องคำนึงในการสร้าง Decision Tree นั้นประกอบไปด้วยการระบุ Attribute และชนิดของ Attribute, การ Split และ Pruning , จำนวนทางเลือกแบบ Binary หรือ Multiway เป็นต้น ส่วนการเลือกค่าที่บ่งบอกว่า attribute สมควรนำมาใช้เป็นคุณลักษณะในการแบ่งหรือไม่นั้นจะใช้ค่า Gini Index ซึ่งมีสมการดังต่อไปนี้

$$Gini(t) = 1 - \sum_{i=0} [p(i|t)]^2$$

และการคำนวณค่า Entropy หรือค่าความยุ่งเหยิงของข้อมูลนั้น มีสมการดังต่อไปนี้

$$Entropy(t) = 1 - \sum_{i=0} [p(i|t)] \log_2 p(i|t)$$

การเรียนรู้จำแนกในแบบ Decision Tree นั้น ถือได้ว่าถูกนำไปประยุกต์ใช้ในหลายปัญหาตั้งแต่ปัญหาพื้นฐานไปจนถึงปัญหาที่ซับซ้อน และแม้จะเป็นโมเดลที่ง่ายแต่ก็ได้รับการพิสูจน์ว่าสามารถทำงานได้อย่างมีประสิทธิภาพ และเป็นหนึ่งในโมเดลที่ถูกเลือกเป็นพื้นฐานในการทำงานกับข้อมูลที่มีความซับซ้อนสูง

2.2 งานวิจัยที่เกี่ยวข้อง

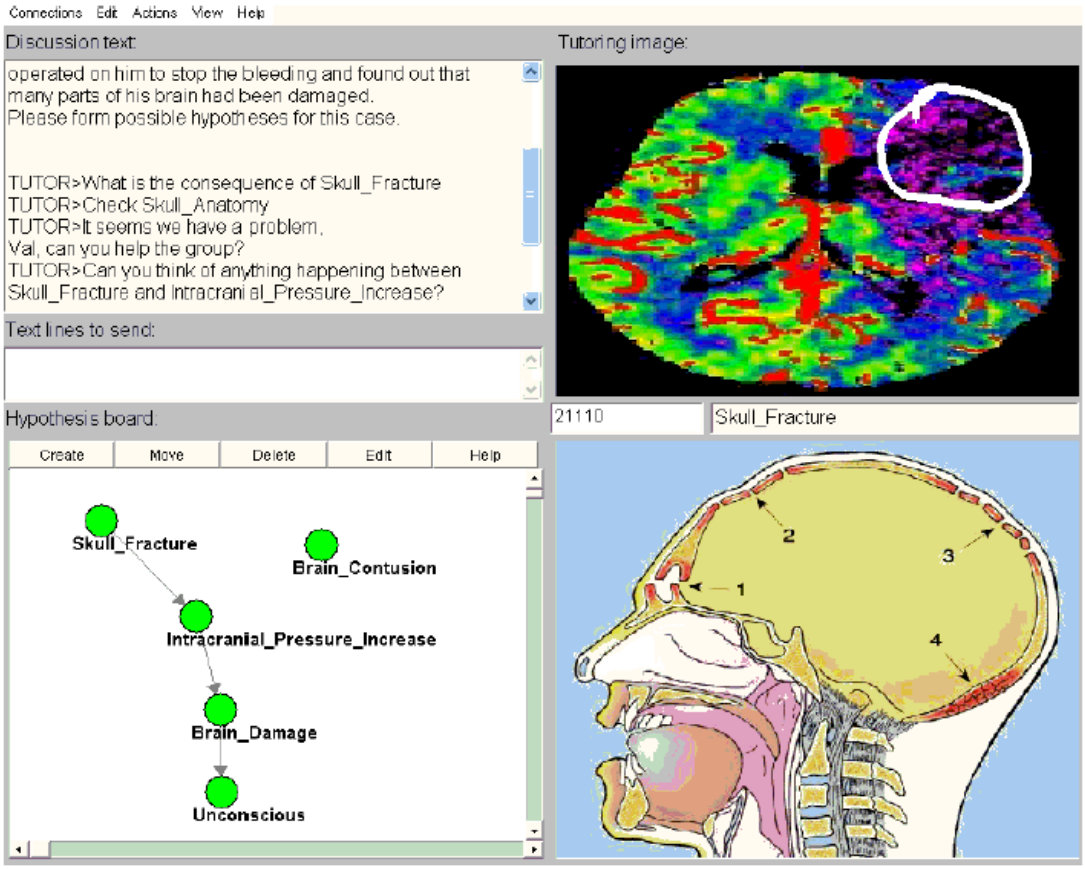
งานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบการสอนแบบสนทนานั้นได้ถูกสร้างและพัฒนาขึ้นอย่างต่อเนื่อง แม้ในแง่ของผู้พัฒนาทั่วไปจะเน้นไปที่การพัฒนา Authoring Tools หรือจัดทำ Cirriculum script โดยตรง เพื่อสร้างระบบการสอนแบบสนทนา และมุ่งเน้นไปที่การทำความเข้าใจอินพุตที่หลากหลายของผู้เรียน สำหรับในส่วนของการประยุกต์ใช้การเรียนรู้ของเครื่องกับข้อมูลการเรียนการสอนแบบสนทนานั้น ยังมีไม่มากนัก โดยเฉพาะอย่างยิ่งกับการพัฒนาระบบการสอนแบบสนทนาในภาษาไทย ในส่วนนี้จะเป็นงานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบการสอนแบบสนทนา ซึ่งมีงานวิจัยที่เกี่ยวข้องดังนี้

2.2.1 COMET

งานของ Siriwan Suebnukarn, Peter Haddawy ซึ่งนำเสนอระบบ ITS ชื่อ COMET (A Collaborative Intelligent Tutoring System for Medical Problem-Based Learning) เพื่อการสอนในระบบการเรียนการสอนแบบร่วมมือ สำหรับนักศึกษาแพทย์ ในหัวข้อการวิเคราะห์การบาดเจ็บของสมอง โดย COMET ใช้ Bayesian network ในการแทนตัว Knowledge Domain และเพื่อตรวจดู Student Model แต่ละคนด้วยว่าวิเคราะห์เหตุและผลของปัญหาได้ตาม Knowledge Domain หรือไม่ โดยการให้ผู้เรียนเสนอ Hypothesis path นำมาเทียบกับ Knowledge network ว่า Path นั้นมีความสมบูรณ์ หรือไม่สมบูรณ์อย่างไร Tutor agent ใน COMET ใช้ ประโยคสนทนาใน Discussion board ในการแทนตัวผู้สอนในการใช้กลยุทธ์การสอนต่างๆ โดย Tutor agent ซึ่งมีการจัดเตรียมไว้สำหรับสถานการณ์ต่างๆ 8 สถานการณ์ โดยระบบติดต่อกับผู้เรียนมี 4 ส่วนคือ Discussion board , Hypothesis board, Tutoring Image และ Medical Repository ดังภาพที่ 2.2 ซึ่งเป็น Interface ของระบบ COMET

ภาพที่ 2.2

ส่วนติดต่อกับผู้ใช้ของ COMET



เมื่อเริ่มต้น ผู้เรียนจะได้รับปัญหาในการวิเคราะห์โดยทุกคนสามารถร่วมกันวิเคราะห์ปัญหาผ่านทาง Discussion board และใช้ Tutoring Image ในการระบุถึงตำแหน่งต่างๆ ได้โดยการวาดลงไปในภาพ แต่ทั้ง 2 ส่วนนี้ถือเป็นส่วนอำนวยความสะดวกเท่านั้นไม่ถือเป็น input ของระบบ COMET จะรับ input ผ่านทาง hypothesis board เท่านั้น โดย hypothesis board โดยผู้เรียนใช้ Medical Repository ซึ่งมี label ของชื่ออวัยวะทั้งหมดของมนุษย์ ซึ่งผู้เรียนสามารถสร้าง Hypothesis path ได้ผ่านทางการใช้เมาส์ จับวางมาจากส่วนของ Medical Repository

งานวิจัยนี้ได้เก็บข้อมูลจากการสอนจริงในชั้นเรียน และจากการให้ข้อมูลในปัญหาของผู้สอนระหว่างการเรียนการสอนหัวข้อนี้รวมถึงวิธีแก้ไขสถานการณ์ และนำมาพัฒนาตัว Knowledge domain network และการแก้ไขสถานการณ์ไว้ 8 วิธี และใช้วิธีในการวัดผลคือเก็บข้อมูลว่าผู้สอนจริงส่วนใหญ่แก้ไขสถานการณ์อย่างไร และ Tutor agent ของ COMET ก็ได้แก้ไขสถานการณ์ด้วยวิธีเดียวกัน โดยการวัดค่าทางสถิติแล้วสรุปว่า โดยเฉลี่ยแล้ว COMET ใช้กลยุทธ์ในการแก้ปัญหาได้เหมือนกับผู้สอนจริง 74.17%

COMET ใช้การเตรียมการที่มุ่งเน้นใน Domain การสอนด้านการสร้าง Hypothesis โดยตรงทำให้ไม่มีความเป็นสาธารณะ (general) มากพอ สำหรับการประยุกต์ใช้ในระบบการสอนวิชาอื่น กล่าวคือ COMET จะต้องใช้ Domain Knowledge แบบ Bayesian Network ซึ่งในหัวข้อการสอนในรูปแบบอื่นๆ นั้นอาจจะไม่เหมาะสมกับการสร้างเป็น Bayesian Network และการรับ input ของ COMET ก็มีลักษณะปิด รวมถึงการใช้ประโยคสนทนากับผู้เรียนก็เป็นลักษณะเพื่อรองรับการแก้ไขสถานการณ์ต่างๆ เท่านั้น

2.2.2 CIRCSIM (CST)

CIRCSIM-Tutor หรือเรียกว่า CST เป็นระบบสอนเสริมอัจฉริยะสำหรับสอนเรื่องกลไกการควบคุมความดันในหลอดเลือดสำหรับนักศึกษาแพทย์ปี 1 CST แตกต่างจากระบบ ITS อื่นๆ เนื่องจากใช้ภาษาธรรมชาติโต้ตอบในการดำเนินการสอน ไม่มีการใช้รูปภาพ, ตัวเลือก หรือภาพเคลื่อนไหวใดๆ โดยตัวผู้สอนจะเป็นผู้กำหนดการสนทนา และจำลองเป็นการเรียนการสอนในรูปแบบ Socratic นั่นคือเครื่องจะเป็นผู้ตั้งคำถามให้ผู้เรียนโต้ตอบด้วยคำพูดได้อย่างอิสระเช่นเดียวกับระบบการสอนของมนุษย์

สำหรับเนื้อหาการสอนหรือ Domain knowledge นั้นจะเป็นเรื่องของความดันเลือดที่มีผลต่อตัวแปรของกลไกระบบเลือดทั้งหมด 7 ตัวแปร โดยมีจุดมุ่งหมายคือผู้เรียนควรจะสามารถสรุปถึงการเพิ่ม, ลด หรือคงที่ของตัวแปรกลไกทั้ง 7 ตัวแปรได้ในสถานการณ์ต่างๆ เพื่อให้ผู้เรียนรู้ถึงกลไกความสัมพันธ์ของตัวแปรทั้งหมด ดังนั้นระบบ CST จะเริ่มต้นด้วยคำถามซึ่งระบุถึงสถานการณ์ที่กระทบต่อระบบความดันเลือด แล้วให้ผู้เรียนคาดเดาผลกระทบที่จะเกิดขึ้นต่อตัวแปรทั้ง 7 ตัวภายใต้สถานการณ์นั้น เมื่อระบบได้รับคำตอบจากผู้เรียนแล้วจะประเมินคำตอบนั้นเพื่อเลือกประโยคโต้ตอบที่เหมาะสม โดยใช้ประโยคที่ช่วยในการแก้ไขความรู้ที่ยังไม่สมบูรณ์ของผู้เรียนตามแผนการสอนที่กำหนดไว้

เมื่อระบบรับประโยคจากผู้เรียนแล้ว จะทำการแบ่งประเภทของคำและไวยากรณ์ของคำ ตรวจสอบคำสะกดและใช้ไฟไนท์เสตท ทรานส์ดิวเซอร์ เพื่อรับรู้ความหมายของประโยคผู้เรียนและวิเคราะห์การพ้องกันของกรอบความคิด (ontology) จากนั้นจึงสรุปความหมายของประโยคนั้นโดยแยกตามประเภทที่กำหนดไว้ก่อนหน้า ซึ่งนำไปสู่การเลือกประโยคโต้ตอบที่เหมาะสมกับประโยคของผู้เรียน ซึ่งสามารถถือได้ว่าทั้งแผนการสอนและเนื้อหาการสอนนั้นถูกรวมไว้อยู่ในคลัง

ของประโยคที่เตรียมไว้ ซึ่งระบบอื่นๆ ก็มักจะใช้วิธีการในลักษณะนี้เช่นกัน แต่อาจจะแตกต่างกันในแง่ของกระบวนการประมวลผลและตีความอินพุตจากผู้เรียน

2.2.3 AUTOTUTOR

Autotutor เป็นระบบ ITS แบบสนทนาที่สอนเนื้อหาเกี่ยวกับ พื้นฐานคอมพิวเตอร์และฟิสิกส์แบบนิวโตเนี่ยน ซึ่งหมายความว่าระบบ Autotutor นี้สามารถนำไปใช้กับเนื้อหาวิชาอื่นๆ ได้อีก โดย Autotutor จะเริ่มต้นการสอนโดยตั้งคำถามที่มีลักษณะเป็นคำถามแบบเปิด และให้ผู้เรียนตอบในเชิงอธิบาย (explanation based) จากนั้น Autotutor จะนำคำตอบมาประเมินจากคำตอบในอุดมคติ เพื่อตรวจสอบส่วนที่ยังขาดหาย หรือเข้าใจผิดของผู้เรียน และใช้การสนทนาโต้ตอบหลายวิถีทาง (เช่นการ feedback และ hint หรือถามนำสู่คำที่ต้องการ เป็นต้น) เพื่อโน้มน้าวให้ผู้เรียนสร้างคำตอบที่สมบูรณ์ตามอุดมคติ ซึ่งกระบวนการนี้ถูกเรียกว่า EMT (Expectation Misconception Tutoring)

ทั้งนี้ Autotutor ต่างจากระบบอื่นในส่วนของ Interface เนื่องจากมีตัวการ์ตูนรูปมนุษย์เป็นตัวแทนของการแสดงท่าทีและเสียงพูดเพื่อโต้ตอบกับผู้เรียนโดยตรงด้วย และการประมวลผลเพื่อประเมินคำตอบของผู้เรียนใช้ LSA (Latent Semantic Analysis) ในการสรุปความรู้จากคำตอบของผู้เรียน โดยใช้การเปรียบเทียบกับบทความหรือหนังสือของผู้เชี่ยวชาญในเนื้อหาวิชานั้นเป็น Corpus ซึ่งกระบวนการ LSA จะให้คะแนนความใกล้เคียงของเนื้อหากลุ่มข้อความได้เป็นอย่างดี ดังนั้น LSA จึงเหมาะสมในการใช้กับการโต้ตอบสนทนาด้วยประโยคที่ยาว ซึ่งเหมาะกับเนื้อหาวิชาทั้งสองที่ Autotutor ใช้สอน โดยที่ตัวคำตอบแต่ละครั้งจะถูกเก็บไว้ เพื่อรวบรวมมาเทียบเคียงตามกระบวนการ LSA ว่าผู้เรียนนั้นใช้คำที่ได้แสดงความรู้ครบตามคำตอบที่ต้องการหรือไม่ นับตั้งแต่เริ่มทำการสอน

2.2.4 ATLAS-ANDES TUTOR

เป็นระบบ ITS แบบสนทนาที่สอนเนื้อหาด้าน Physic แบบนิวโตเนี่ยนเหมือนกับ Autotutor แต่ไม่ได้ใช้วิธี LSA ในการพิจารณาประโยคของผู้เรียน แต่ใช้ model ที่พัฒนาขึ้นใหม่ชื่อว่า Atlas Planning Engine หรือ APE เพื่อใช้ในการควบคุมการโต้ตอบสนทนาทั้งหมดของระบบ โดยมีจุดมุ่งหมายเพื่อให้เกิดความเป็นอิสระต่อกันทั้งในส่วนของเนื้อหาวิชาและการควบคุมการสนทนา เพื่อการสร้างระบบ ITS แบบสนทนาได้อย่างซับซ้อนยิ่งขึ้น สามารถเขียนสคริปต์ลำดับ

การสนทนาที่เป็นลำดับชั้นได้ (เช่นมีบทสนทนาย่อยเพื่อจุดประสงค์เฉพาะในระหว่างการสอนได้เป็นต้น) หรือสามารถเรียกว่าเป็นระบบวางแผนการสนทนาก็ได้ โดยยึดหลักการของ HTN (Hierarchical Task Network) หรือการกำหนดแขนงของงานออกเป็นส่วนย่อยๆ ได้ ซึ่งตัว Andes Physics Tutor เองนั้นก็เป็นตัวอย่างของระบบ ITS แบบโต้ตอบสนทนาที่ใช้ APE เป็นกลไกสำคัญในการสร้างระบบ และใน CIRCSIM รุ่นหลังๆ ก็มีการใช้ APE ด้วยเช่นกัน

ระบบ ITS แบบสนทนาโดยทั่วไป โดยเฉพาะในรูปแบบของ CIRCSIM จะใช้ Finite-State-Machine(FSM) เป็นเครื่องมือในการรับรู้ ประโยคสนทนาจากผู้เรียน ซึ่งข้อจำกัดของ FSM คือการไม่รองรับโครงสร้างการสนทนาแบบลำดับชั้นได้ ซึ่ง APE จึงถูกพัฒนามาเพื่อลบข้อจำกัดดังกล่าว อีกทั้งเพื่อการจัดโครงสร้างแบบลำดับชั้นได้ของ APE นั้นยังช่วยในรองรับการเปลี่ยนแปลง Dialogue plan ที่สามารถปรับเปลี่ยนได้ในภายหลังอย่างเป็นอิสระ

ตารางที่ 2.2
เปรียบเทียบงานวิจัยที่เกี่ยวข้อง

ผู้เขียน	นำเสนอ	ข้อดี	ข้อเสีย
Siriwan Suebnukarn, Peter Haddawy (2004)	ระบบ COMET ที่ใช้ Bayesian Network ในการแทน Domain Knowledge และใช้การสนทนาในการให้เนื้อหา โดยมีการเตรียมวิธีแก้ไขสถานการณ์การสอน 8 วิธี	- ครอบคลุมการสอนในเนื้อหาวิชา - มีกลยุทธ์ที่แน่นอนในการจัดการสถานการณ์ต่างๆ ระหว่างการสอน	- ไม่มีส่วนให้ผู้เรียนสนทนาโต้ตอบได้ - วิธีการไม่เป็นอิสระจากเนื้อหาวิชา
Woo, C.W., Evens, M.W., Freedman, R., Glass, M., Shim, L.S., Zhang, Y., Zhou, Y., Michael, J. (2006)	CIRCSIM Tutor ระบบการสอนแบบสนทนาในเชิงถาม-ตอบ ซึ่งใช้ FSM ในการวิเคราะห์คำตอบของผู้เรียน แล้วจึงเลือกบทสนทนาโต้ตอบของผู้สอน ซึ่งมีลักษณะเป็น rule-base	- การใช้ FSM ทำให้รับรู้ถึงไวยากรณ์ และทำความเข้าใจกับคำตอบผู้เรียนได้อย่างแม่นยำ	- เหมาะสมเฉพาะการสอนเนื้อหาวิชาที่ผู้เรียนใช้บทสนทนาเป็นคำตอบสั้นๆ - ขาดความยืดหยุ่นและไม่เป็นอิสระต่อเนื้อหาวิชา

ตารางที่ 2.3 (ต่อ)
เปรียบเทียบงานวิจัยที่เกี่ยวข้อง

ผู้เขียน	นำเสนอ	ข้อดี	ข้อเสีย
Arthur C. Graesser, Patrick Chipman, Brian C. Haynes, and Andrew Olney (2005)	ระบบ AUTOTUTOR ซึ่งเป็นระบบการสอนแบบสนทนาในเชิงถาม-ตอบ โดยใช้ LSA ในการตรวจสอบความเข้าใจจากคำตอบของผู้เรียน และแก้ MisConception ของผู้เรียนตามกลยุทธ์	- มีความยืดหยุ่น และสามารถประเมินคำตอบของผู้เรียนได้ดี และหลากหลาย - เป็นอิสระต่อเนื้อหาวิชามากกว่าการใช้ FSM - รองรับการตั้งคำถามจากผู้เรียน	- ต้องอาศัย Corpus ของเนื้อหาวิชาเพื่อกระบวนการทาง LSA - LSA ไม่สามารถตรวจจับลำดับของคำได้ จึงมีโอกาสเข้าใจคำตอบของผู้เรียนได้ไม่ถูกต้อง
Pamela W. Jordan and Kurt VanLehn (2002)	ระบบ ATLAS ซึ่งใช้ NLU : CARMEL ในการตรวจสอบบทสนทนาของผู้เรียนทั้งในด้าน Syntax และ Semantic เพื่อ represent เป็นระดับ Meaning เพื่อให้ตัว Planner ของระบบเลือกบทสนทนาของผู้สอนได้สอดคล้องยิ่งขึ้น และยังมีกระบวนการรองรับการสนทนาแบบลำดับชั้น (Hierachy) ได้	- เพิ่มความสมบูรณ์ของระบบการสอนแบบสนทนา โดยรองรับการสนทนาแบบมีหัวข้อย่อยเป็นลำดับชั้นได้ - เพิ่มประสิทธิภาพในการวิเคราะห์ความหมายของคำตอบผู้เรียนได้แม่นยำยิ่งขึ้น	- ต้องใช้ทรัพยากรในการพัฒนาระบบมาก เนื่องจากมีองค์ประกอบในระบบหลากหลาย - แม้จะรองรับความซับซ้อนของภาษาที่ดีขึ้น แต่ยังบทสนทนาบางรูปแบบที่ระบบไม่สามารถเข้าใจความหมายได้อย่างถูกต้อง