

ชื่อโครงการ “การพัฒนาซอฟต์แวร์ Data Mining ในการสกัดรูปแบบที่สำคัญจากฐานข้อมูล บันทึก : เพื่อค้นหาคุณสมบัติที่เหมาะสมในการได้รับคัดเลือกเข้าทำงาน”

1. นายวรวิทย์ อภิบาลคุณกุล*
2. นายศุภสิทธิ์ นามา

1) ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

บทคัดย่อ

Genetic Algorithms for Inductive Learning (GAIL) เป็นเทคนิคหนึ่งทาง Data Mining ที่ใช้ในการจำแนกข้อมูลโดยการนำ Genetic Algorithms มาผสมผสานกับแนวทาง Inductive Learning ผลลัพธ์ที่ได้จากการประยุกต์ใช้ GAIL คือกฎที่สามารถใช้จำแนกข้อมูลที่ได้ศึกษาออกเป็นกลุ่ม ๆ อย่างมีประสิทธิภาพ

โครงการวิจัยฉบับนี้นำเสนอการใช้ GAIL ในการจำแนกข้อมูลนักศึกษา และสามารถปรับปรุงประสิทธิภาพโดยใช้แนวทางเพื่อให้ได้มาซึ่งกฎเกณฑ์ที่สามารถใช้คาดการณ์การเข้ารับทำงานของผู้จบการศึกษาใน หลักสูตรวิทยาศาสตร์บัณฑิต คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี นอกจากนี้ยังได้ปรับปรุงฟังก์ชันค่าความแข็งแรงให้เหมาะสมกับคุณลักษณะของข้อมูลที่ใช้ด้วย

คำสำคัญ : การจำแนกข้อมูล / เทคโนโลยีสารสนเทศ / มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี / GAIL
Inductive Learning

การได้รับตำแหน่งหน้าที่การงานที่ดีนั้นเป็นจุดประสงค์หนึ่งที่สำคัญสำหรับการเข้ารับการศึกษาของบัณฑิต และของการพัฒนาหลักสูตรการเรียนการสอนของสถาบันอุดมศึกษา ซึ่งมีหน้าที่ผลิตบัณฑิตที่เป็นที่ต้องการของสังคม นอกเหนือไปจากนั้นการสนับสนุนและพัฒนาทางด้านกิจกรรมนับว่าเป็นสิ่งสำคัญสิ่งหนึ่งที่สถาบันอุดมศึกษาต้องตระหนักถึง เนื่องจากในปัจจุบันการคัดเลือกบุคลากรเพื่อเข้าทำงานของบางองค์กรนั้นไม่ เพียงแต่พิจารณาคุณสมบัติทางวิชาการเพียงอย่างเดียว แต่ได้มีการพิจารณาถึงคุณสมบัติอื่น ๆ ด้วย เช่น การเข้าร่วมกิจกรรม ความสามารถทางกีฬา ดนตรี เป็นต้น

Genetic Algorithms for Inductive Learning (GAIL) เป็นเทคนิคใหม่สำหรับจำแนกข้อมูล ที่นำเอา

เครื่องมือ (Tool) ในการจำแนกข้อมูลสองชนิดมาประยุกต์ใช้ร่วมกัน กล่าวคือ พัฒนาการค้นหา (Searching) ด้วย

Genetic Algorithms มาใช้ร่วมกับ Inductive Learning เนื้อหาในบทนี้กล่าวถึง ความหมายและความสำคัญของการจำแนกข้อมูล Genetic Algorithms for Inductive Learning (GAIL) และวรรณกรรมที่เกี่ยวข้อง

1.ความหมายและความสำคัญของการจำแนกข้อมูล

‘การจำแนกข้อมูล’ (Classification) และ คือ กระบวนการแบ่งข้อมูลเข้าคลาส (Class) ที่ข้อมูลนั้น ๆ ควรจะเป็นสมาชิกอยู่ในกระบวนการนี้ ทุเปิล (Tuples) ของฐานข้อมูลที่ใช้ในการปรับสอนจะถูกวิเคราะห์ค้นหา

คุณสมบัติร่วมเพื่อสร้างคำอธิบาย (Description) โมเดล หรือขั้นตอนวิธีการจำแนกข้อมูล (Classification Model or Classification Procedure) แต่ละคลาสที่สามารถนำไปใช้จำแนกข้อมูลในอนาคตได้ ทูเปิ้ลที่ใช้ปรับสอนประกอบด้วยค่าของแอตทริบิวต์หรือคุณลักษณะต่าง ๆ และชื่อคลาสที่ทูเปิ้ลนั้นเป็นสมาชิกอยู่ ซึ่งคลาสดังกล่าวต้องถูกนิยามหรือกำหนดขึ้นไว้ก่อนแล้ว การสร้างขั้นตอนวิธีการจำแนกจากกลุ่มข้อมูลซึ่งทราบคลาสที่ข้อมูลเป็นสมาชิกอยู่แล้วนี้ นอกจาก 'การจำแนกข้อมูล' แล้ว ยังสามารถเรียกแทนได้ด้วยเทอม (Term) 'การรู้จำแพทเทิร์น' (Pattern Recognition) 'การแบ่งแยก' (Discrimination) หรือ 'การเรียนรู้แบบซูปเปอร์ไวส์' (Supervised Learning)

การจำแนกข้อมูลเป็นส่วนประกอบที่สำคัญในงานการแก้ปัญหาต่าง ๆ เช่น การรู้จำ และถูกนำไปประยุกต์ใช้ในงานที่หลากหลาย เช่น การอนุมัตินเชื่อ การตลาดผลิตภัณฑ์ และการวินิจฉัยทางการแพทย์ จนอาจกล่าวได้ว่า การจำแนกข้อมูลเป็นเทคนิคทาง Data Mining ที่ถูกนำไปประยุกต์ใช้อย่างกว้างขวางมากที่สุด

2.Genetic Algorithms for Inductive Learning (GAIL)

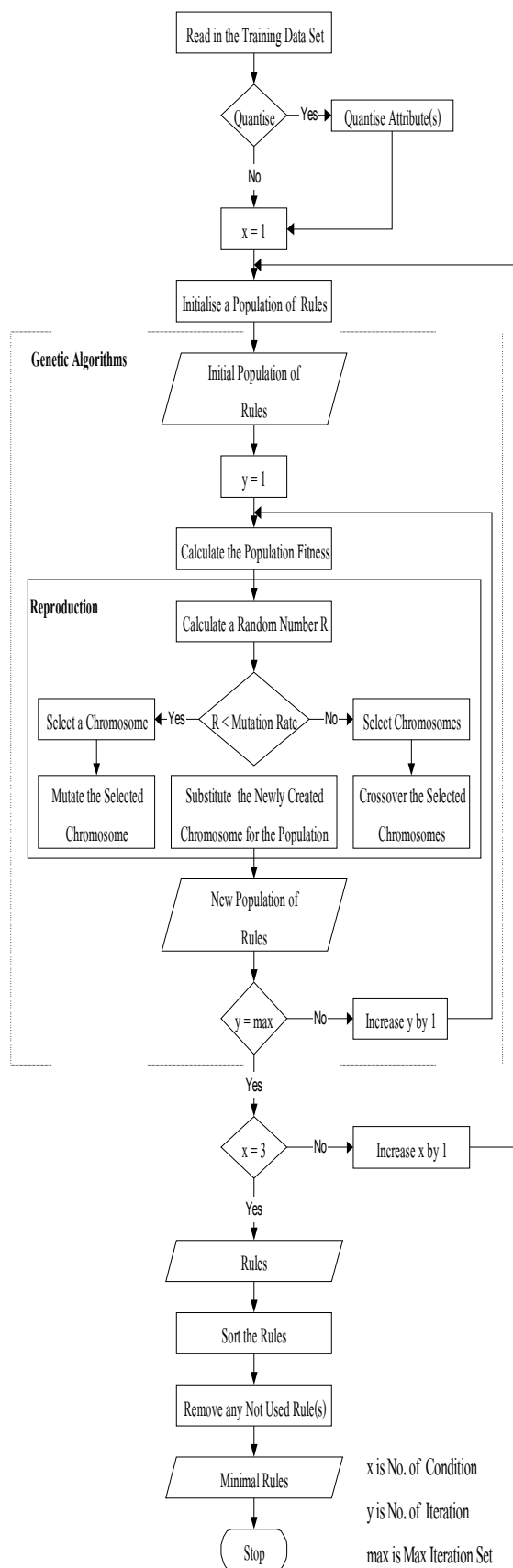
GAIL เป็นเทคนิคที่ได้รับการคิดค้นและนำไปประยุกต์ใช้แล้วว่ามีประสิทธิภาพ GAIL สามารถจำแนกกลุ่มข้อมูลดอกไอริส (Iris) ซึ่งเป็นกลุ่มข้อมูลมาตรฐานสำหรับตัดสินใจหรือวัดผลการจำแนกข้อมูล (Benchmark Data Set) ได้ถูกต้องสูงสุด 94.3% เทคนิคนี้เกิดขึ้นจากการใช้ประโยชน์ Genetic Algorithms ซึ่งเป็นอัลกอริทึมสำหรับค้นหาวิธีแก้ปัญหาที่เหมาะสม (Optimization) และเทคนิค Inductive Learning ในกระบวนการจำแนกข้อมูล ดังนั้น ผลที่ได้จากการจำแนกข้อมูลด้วย GAIL จึงเป็นกฎสำหรับจำแนกข้อมูลจำนวนเหมาะสม (Optimal) ที่สามารถจำแนกข้อมูลได้อย่างมีประสิทธิภาพ

กฎสำหรับจำแนกข้อมูลที่ GAIL สร้างขึ้นเป็นคำอธิบายคลาสที่อยู่ในรูป "IF condition(s) THEN class type" โดยในส่วนของเงื่อนไข หรือที่โดยปกติเรียกว่า ส่วนนำ (Antecedent) ของกฎระบุคุณลักษณะร่วมของข้อมูลที่เป็นสมาชิกของคลาสที่ระบุไว้ในส่วนตาม (Consequent) ของกฎ เมื่อนำกฎไปใช้ในการจำแนกข้อมูล ข้อมูลที่มีคุณลักษณะตรงตามส่วนนำจะได้รับ การจัดเข้าคลาสในส่วนตาม การอธิบายคุณลักษณะร่วมของสมาชิกในคลาสหนึ่ง ๆ ในลักษณะกฎเป็นข้อดีอีกข้อหนึ่งของ GAIL เนื่องจากกฎเป็นรูปแบบการแทนความรู้ที่ค้นพบ (Knowledge Representation) ที่เรียบง่าย เป็นธรรมชาติและเข้าใจได้โดยง่าย (Intuitive) และสามารถนำไปใช้ประโยชน์ได้ทันทีและ กฎที่ GAIL สร้างขึ้นสำหรับจำแนกกลุ่มข้อมูลดอกไอริสมีจำนวนรวม 6 ข้อ ดังนี้

1. IF (Petal Length < 2.0666) THEN class is *Iris Setosa*
2. IF (0.9 <= Petal Width < 1.3) THEN class is *Iris Versicolor*
3. IF (1.3 < Petal Width < 1.7) THEN class is *Iris Versicolor*
4. IF (Petal Width >= 2.1) THEN class is *Iris Verginica*
5. IF (Petal Length >=5.9333) THEN class is *Iris Verginica*
6. IF (1.7 <= Petal Width < 2.1) THEN class is *Iris Verginica*

3.ขั้นตอนการทำงานของ GAIL

ขั้นตอนการทำงานของ GAIL สามารถสรุปได้ดังรูปที่ 1



รูปที่ 1 : ขั้นตอนการทำงานของ GAIL

3.1 การอ่านค่ากลุ่มข้อมูลตัวอย่างสำหรับปรับ

สอน (Reading in the Training Data Set)

การอ่านค่ากลุ่มข้อมูลตัวอย่างสำหรับปรับสอนเป็นขั้นตอนแรกในการทำงานของ GAIL ในขั้นตอนนี้ข้อมูลตัวอย่างทั้งหมดถูกอ่านและจัดเก็บไว้เพื่อใช้ในการอินตักชันในขั้นตอนต่อ ๆ ไป ข้อมูลตัวอย่างชุดหนึ่ง ๆ ประกอบด้วยชื่อคลาสที่ตัวอย่างนั้น ๆ เป็นสมาชิกอยู่และกลุ่มของค่าของแอททริบิวต์ต่าง ๆ ของตัวอย่าง

3.2 การควอนไทส์ค่าของแอททริบิวต์ (Quantisation of Attributes)

การควอนไทส์ค่าของแอททริบิวต์ใน GAIL เกิดขึ้นเฉพาะกับแอททริบิวต์ที่มีค่าเป็นตัวเลข เป็นการเตรียมข้อมูลก่อนเข้าสู่กระบวนการจำแนก (Preprocessing) โดยแบ่งค่าของข้อมูลออกเป็นช่วงที่แยกจากกัน เช่น อายุ 20 – 30 ปี 31 – 45 ปี เป็นต้นขอบเขตของช่วงต่าง ๆ ถูกกำหนดขึ้นโดย GAIL เอง (โดยการสุ่ม) ผู้ใช้สามารถกำหนดให้ GAIL ควอนไทส์ (Quantise) ข้อมูลหรือไม่ก็ได้ ตามแต่ต้องการ

3.3 Genetic Algorithms (GAs)

ขั้นตอนการทำงานของ GAs ใน GAIL ไม่แตกต่างจากขั้นตอนการทำงานของ GAs โดยทั่วไป กล่าวคือ ยังคงประกอบด้วยขั้นตอนดังนี้

3.3.1 การสร้างประชากรกฎเริ่มต้น (Initialization of a Population of Rules)

เนื่องจากโครโมโซมหรือรหัสวิธีแก้ปัญหาใน GAIL คือกฎสำหรับจำแนกข้อมูล ซึ่งเมื่อถูกถอดรหัส (Decoded) จะได้เงื่อนไขและชื่อคลาสที่ข้อมูลหนึ่ง ๆ จะได้รับการจัดเข้าเป็นสมาชิกเมื่อมีค่าของแอททริบิวต์ตรงตามเงื่อนไข ดังนั้น ประชากรใน GAIL ก็คือกลุ่มของกฎนั่นเอง กฎทั้งหมดถูกเข้ารหัสเป็นสายของเลขฐาน 2 (Binary String) ทั้งนี้เพราะ GAIL ใช้การเข้ารหัสเลขฐาน 2 (Binary Encoding) เป็นแนวทางในการเข้ารหัส ประชากรกฎเริ่มต้นหรือรุ่นแรกถูกสร้างขึ้นโดยการสุ่ม

3.3.2 การคำนวณค่าความแข็งแรงของประชากร (Calculation of the Population Fitness)

GAIL คำนวณค่าความแข็งแรงของประชากรจากฟังก์ชันค่าความแข็งแรง (Fitness Function) ซึ่งมีสมการดังนี้

$$\text{Fitness Value} = \text{Right} - x(\text{Wrong})$$

(1)

โดยที่	คือจำนวนชุดข้อมูลตัวอย่างสำหรับปรับสอนที่มีค่าของแอททริบิวต์ตรงตามเงื่อนไขของโครโมโซมและชื่อคลาสในข้อมูลและโครโมโซมตรงกัน
<i>Right</i>	
<i>Wrong</i>	คือจำนวนชุดข้อมูลตัวอย่างสำหรับปรับสอนที่มีค่าของแอททริบิวต์ตรงตามเงื่อนไขของโครโมโซมแต่ชื่อคลาสในข้อมูลและโครโมโซมไม่ตรงกัน
<i>X</i>	คือค่าถ่วงน้ำหนักของ <i>Wrong</i>

ตารางที่ 1 แสดงค่า Fitness value

ในขั้นตอนนี้ ข้อมูลที่ได้จากการอ่านค่ากลุ่มข้อมูลตัวอย่างในขั้นตอนแรกจะถูกนำมาใช้เพื่อหาค่าฟิตเนส *Right* และ *Wrong* ส่งผลให้เกิดการอินตักชันขึ้นในรอบต่อ ๆ ไปของกระบวนการ ทั้งนี้เนื่องจากโครโมโซมที่มีค่าความแข็งแรงสูงเป็นโครโมโซมที่มีเงื่อนไขและส่วนตามตรงกับค่าของแอททริบิวต์และชื่อคลาสของข้อมูลตัวอย่าง ตามลำดับ ดังนั้น GAIL จึงต้องเรียนรู้ข้อมูลตัวอย่างเพื่อที่จะสามารถสร้างโครโมโซมที่มีค่าความแข็งแรงสูงได้

3.3.3 การสืบพันธุ์ (Reproduction)

การสืบพันธุ์หรือการสร้างประชากรรุ่นต่อ ๆ ไปใน GAIL ประกอบด้วยขั้นตอนใหญ่ ๆ 4 ขั้นตอน ดังนี้

1. การคำนวณตัวเลขสุ่ม R

(Calculation of a Random Number R): R คือตัวเลขที่มีค่าระหว่าง 0.0 ถึง 1.0 ถูกสร้างขึ้นอย่างสุ่มเพื่อนำไปใช้ในการตัดสินใจว่าจะใช้ตัวดำเนินการทางพันธุกรรมใดในการสืบพันธุ์ โดยที่

- ถ้า $R >$ อัตราการมิวเตชัน (Mutation Rate) ที่กำหนดไว้ GAIL จะใช้ครอสโอเวอร์เป็นตัวดำเนินการ
- ถ้า $R <$ อัตราการมิวเตชันที่กำหนดไว้ GAIL จะใช้มิวเตชันเป็นตัวดำเนินการ

2. การเลือกโครโมโซม (Selection of

Chromosome(s)): GAIL จะคัดเลือก โครโมโซมหรือกลุ่มย่อยของโครโมโซมที่แข็งแรงที่สุดจากประชากรรุ่นปัจจุบันทั้งหมดเพื่อสืบพันธุ์ จำนวนโครโมโซมที่คัดเลือกขึ้นกับตัวดำเนินการทางพันธุกรรมที่เลือกใช้ กล่าวคือ ถ้าเป็นการครอสโอเวอร์จะคัดเลือกโครโมโซม 2 ตัว แต่ถ้าเป็นมิวเตชันจะคัดเลือกโครโมโซมเพียง 1 ตัว

3. การครอสโอเวอร์หรือมิวเตชัน

(Crossover or Mutation): การครอสโอเวอร์และมิวเตชันเป็นการสร้างประชากรรุ่นใหม่ของ GAIL โดยใช้ตัวดำเนินการทางพันธุกรรม ครอสโอเวอร์และมิวเตชันตามลำดับ ซึ่งตัวดำเนินการทั้งสองมีการทำงานที่แตกต่างกันดังนี้

ครอสโอเวอร์: ครอสโอเวอร์เป็นตัวดำเนินการที่ใช้กับโครโมโซม 2 ตัว เรียกโครโมโซมทั้งสองตัวนี้ว่าโครโมโซมพ่อและโครโมโซมแม่ โดยการคัดลอกโครโมโซมบางส่วนมาจากโครโมโซมพ่อและบางส่วนมาจากโครโมโซมแม่ โครโมโซมใหม่ที่คาดว่าจะสืบทอดคุณสมบัติที่ดีจากรุ่นพ่อแม่ก็就会被สร้างขึ้น

ผลของครอสโอเวอร์คล้ายคลึงกับการจับคู่ (Mating) ในกระบวนการวิวัฒนาการทางธรรมชาติที่พ่อแม่ถ่ายทอดส่วนของโครโมโซมของตนแก่ลูก ดังนั้นลูกบางคนจึงมีคุณสมบัติเหนือกว่าพ่อแม่ ถ้าได้รับยีนส์หรือลักษณะทางพันธุกรรมเด่นจากทั้งพ่อและแม่

มิวเตชัน: มิวเตชันเป็นตัวดำเนินการที่ใช้กับโครโมโซม 1 ตัว โดยมิวเตชันจะคอมพลีเมนต์ (Complement) ยีนส์ในโครโมโซมที่เลือกไว้ ด้วย

วิธีการนี้ GAIL สามารถสร้างโครโมโซมใหม่ได้โดยอัตโนมัติ (Spontaneously)

4. การแทนที่ประชากรรุ่นปัจจุบันด้วยโครโมโซมที่สร้างขึ้น (Substitution of the Newly Created Chromosome(s) for the Population): เมื่อได้โครโมโซมใหม่จากการสืบพันธุ์ GAIL จะลบโครโมโซมเก่าที่มีค่าความแข็งแรงต่ำที่สุดในประชากรรุ่นปัจจุบันออก แล้วแทนที่ด้วยโครโมโซมใหม่นั้น เพื่อให้ได้เป็นประชากรรุ่นใหม่

การทำงานในขั้นตอนการสืบพันธุ์เป็นกระบวนการวนซ้ำ เมื่อสิ้นสุดการแทนที่ประชากรประชากรรุ่นใหม่ที่ได้จะถูกประเมินค่าความแข็งแรงและสืบพันธุ์สร้างประชากรรุ่นต่อไป เนื่องจาก GAIL ใช้จำนวนรอบเป็นตัวพิจารณาการสิ้นสุดการวนซ้ำ ดังนั้นการสร้างประชากรรุ่นใหม่จึงเกิดขึ้นจนกระทั่งถึงจำนวนรอบการทำงานสูงสุดที่กำหนดไว้

การทำงานในขั้นตอน GAs ของ GAIL จะเกิดขึ้นซ้ำทั้งหมดรวม 3 รอบ เพื่อสร้างกฎสำหรับจำแนกข้อมูลที่มีจำนวนเงื่อนไขแตกต่างกัน โดยสร้างกฎที่มีเงื่อนไข 1 2 และ 3 ข้อ ในรอบที่ 1 2 และ 3 ตามลำดับ จำนวนเงื่อนไขในกฎที่ GAIL สร้างถูกกำหนดไว้สูงสุด 3 ข้อ เพื่อไม่ให้กฎที่ได้จำเพาะเจาะจง (Specific) เฉพาะกับข้อมูลตัวอย่างสำหรับปรับสอนมากเกินไป

3.4 การจัดลำดับกฎ (Sorting the Rules)

เมื่อสิ้นสุดการทำงานของ GAs จะได้กฎสำหรับจำแนกข้อมูลจากการถอดรหัสโครโมโซมในประชากรรุ่นสุดท้าย กฎเหล่านี้จะถูกจัดลำดับตามค่าความแข็งแรงจากมากไปหาน้อยเพื่อให้พร้อมสำหรับใช้ทดสอบในขั้นตอนต่อไป

3.5 การตัดกฎที่ไม่ได้ใช้ออก (Removing any Not Used Rules)

ขั้นตอนนี้เป็นการทำให้กฎมีจำนวนต่ำสุดเท่าที่เป็นได้ โดยการทดสอบข้อมูลตัวอย่างสำหรับปรับสอน

ที่ละชุดด้วยกฎทั้งหมดที่ได้ตามลำดับ แล้วตัดกฎที่ไม่ได้ถูกนำไปใช้จำแนกข้อมูลชุดใดเลยออก

การตัดกฎที่ไม่ได้ใช้ออกเป็นขั้นตอนสุดท้ายของ GAIL เมื่อสิ้นสุดขั้นตอนนี้ จะได้กฎสำหรับจำแนกข้อมูลที่สามารถนำไปใช้จำแนกชุดข้อมูลอื่น ๆ ที่มีลักษณะเดียวกันได้

ข้อมูลที่ใช้ในโครงการวิจัย ข้อมูลที่นำมาใช้ในโครงการวิจัยนี้เป็นกลุ่มข้อมูลของนักศึกษาหลักสูตรปริญญาวิทยา ศาสตร์บัณฑิต สาขาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี จากแฟ้มข้อมูลนักศึกษาปริญญาตรีของคณะ ตั้งแต่รุ่นที่ 2 – 7 ที่ศึกษารายวิชาบังคับและเลือกตามหลักสูตรครบ ซึ่งข้อมูลทั้งหมดได้มาจาก 3 แหล่งข้อมูล คือ จากการออกแบบสอบถาม จากฐานข้อมูลของงานทะเบียน และจากฐานข้อมูลของ ALUMNI ข้อมูลทั้งหมดถูกแบ่งออกเป็น 2 กลุ่ม คือกลุ่มข้อมูลตัวอย่างสำหรับปรับสอน และกลุ่มข้อมูลสำหรับทดสอบประสิทธิภาพของกฎ

การวิเคราะห์ข้อมูลที่ใช้ในโครงการวิจัย จากการศึกษาวิเคราะห์กลุ่มข้อมูลนักศึกษา ได้คลาสและแอททริบิวต์ที่ใช้ในการจำแนก และพบข้อจำกัดบางประการของข้อมูล ดังนี้

1. คลาสที่ใช้จำแนกข้อมูลในโครงการวิจัย

โครงการวิจัยนี้ใช้แอททริบิวต์ต่างๆ ของข้อมูลนักศึกษาเป็นแอททริบิวต์สำหรับจำแนกข้อมูล จึงนิยามคลาสที่ใช้ด้วยช่วงของ การรับงานทำและตรงสาขา การได้รับทำงานและไม่ตรงสาขา และยังไม่มีการในแต่ละคลาสที่ใช้จำแนกข้อมูลในโครงการวิจัยนี้มีนิยามดังแสดงในตาราง

คลาส	นิยาม
1	ได้งานทำและตรงสาขา
2	ได้งานทำและไม่ตรงสาขา
3	ไม่มีงานทำ

ตารางที่ 2 : นิยามของคลาสที่ใช้จำแนกข้อมูลในโครงการวิจัย

โครงการวิจัยนี้ใช้นิยามคลาสด้วยช่วงของ GPA จำนวน 3 ช่วง ก็เพื่อให้ผู้ที่ได้จากการจำแนกข้อมูล นักศึกษาโดยใช้คลาสดังกล่าวมีความหมายชัดเจน เข้าใจได้ง่าย

2. แอททริบิวต์ของข้อมูลที่ใช้ในโครงการวิจัย

โครงการวิจัยนี้ทำการคัดเลือกแอททริบิวต์ของข้อมูลที่ใช้ในการจำแนกจากแอททริบิวต์ทั้งหมดในแฟ้มข้อมูลนักศึกษาปริญญาโท การคัดเลือกเริ่มจากพิจารณาแอททริบิวต์ทั้งหมดว่า แอททริบิวต์ใดบ้างที่อาจมีผลต่อการได้รับงานทำ ที่นักศึกษาจะศึกษาได้ ซึ่งจากการพิจารณาอย่างรอบคอบแล้ว แอททริบิวต์ที่คาดว่าจะมีผลต่อความสามารถในการศึกษาในหลักสูตรของนักศึกษาหรือบุคคลหนึ่ง ๆ คือ

1. สถานะภาพการทำงานในปัจจุบัน
2. GPA
3. ความสามารถเฉพาะทางคอมพิวเตอร์
4. ประสบการณ์ในการทำงานก่อนและ/หรือในระหว่างที่คุณกำลังศึกษาในระดับปริญญาตรีของคณะฯ
5. ส่วนร่วมกิจกรรมของทางคณะเทคโนโลยีสารสนเทศและทาง มจร.
6. ตำแหน่งใน สมาคม/ชมรม
7. ระดับความสามารถทางภาษาอังกฤษ
8. ระดับความสามารถในภาษาอื่น ๆ
9. เพศ

ทั้งนี้ค่าของแอททริบิวต์ทั้งหมดได้มาจากแฟ้มข้อมูลโดยตรง ยกเว้น 'อายุ' ซึ่งสามารถคำนวณค่าจากระยะเวลาเป็นปีระหว่างค่าของแอททริบิวต์ 'วันเดือน ปี เกิด' ที่จัดเก็บไว้ในแฟ้มข้อมูลกับเดือน ปี หรือภาคการศึกษาที่นักศึกษาเจ้าของข้อมูลเข้าศึกษา

3. ข้อจำกัดของข้อมูลที่ใช้ในโครงการวิจัย

กลุ่มข้อมูลนักศึกษาที่นำมาใช้ในโครงการวิจัยนี้มีข้อจำกัดบางอย่างซึ่งอาจส่งผลให้ประสิทธิภาพการจำแนกข้อมูลด้วยเทคนิค GAIL ลดต่ำลงกว่าเมื่อข้อมูลไม่มีข้อจำกัด ข้อจำกัดนี้ ได้แก่ ข้อจำกัดด้านปริมาณและความชัดเจนของข้อมูล

3.1 กลุ่มข้อมูลนักศึกษามีปริมาณจำกัด กล่าวคือข้อมูลตัวอย่างสำหรับปรับสอนในบางคลาส มีจำนวนน้อยเมื่อเทียบกับจำนวนข้อมูลในคลาสอื่น ๆ ข้อมูลนักศึกษาในคลาสต่าง ๆ มีจำนวนดังแสดงในตารางที่ 3

คลาส	จำนวน (ชุด)		
	ข้อมูลปรับสอน	ข้อมูลทดสอบ	รวม
1	180	82	262
2	56	27	83
3	22	10	32
รวม	258	119	377

ตารางที่ 3 : จำนวนข้อมูลนักศึกษาในคลาสต่าง ๆ

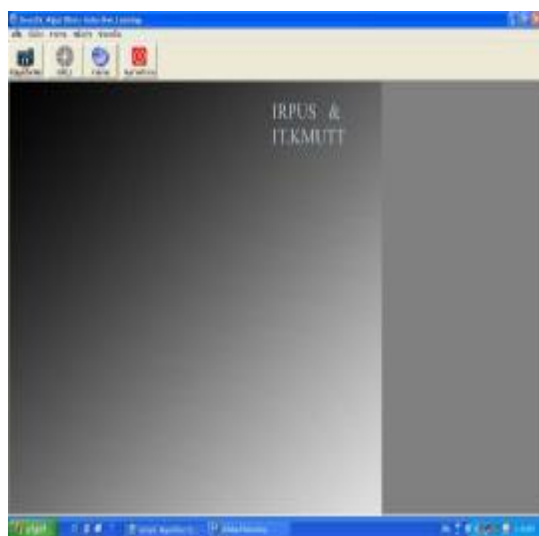
จากตารางที่ 3 ข้อมูลปรับสอนมีจำนวนรวม 258 ชุด และในคลาส 3 มีจำนวน 32 ชุด คิดเป็น 12.40% ของข้อมูลปรับสอนทั้งหมด ซึ่งนับว่ามีจำนวนน้อยมากเมื่อเทียบกับจำนวนข้อมูลในคลาส 1 และ 2 และแม้ว่าจำนวนข้อมูลในคลาส 2 จะมากกว่าในคลาส 3 แต่ก็มีเพียง 83 ชุด น้อยกว่าในคลาส 1 ที่มี 262 ชุด คิดเป็น 31.67% ของจำนวนข้อมูลในคลาสน้อยจะทำให้กฎสำหรับจำแนกข้อมูลในคลาสนั้นมีค่าความแข็งแกร่งต่ำ แนวโน้มที่ GAIL จะสร้างกฎสำหรับจำแนกข้อมูลในคลาส 3 จึงต่ำกว่าคลาส 2 และคลาส 2 ต่ำกว่า

คลาส 1 เมื่อรอบการวนซ้ำในขั้นตอนการสืบพันธุ์มีจำนวนมากอาจส่งผลให้กฎสำหรับจำแนกข้อมูลในคลาส 2-3 ที่มีจำนวนข้อมูลต่ำที่สุดอาจไม่สามารถถูกสร้างขึ้นได้

ข้อจำกัดด้านปริมาณของกลุ่มข้อมูลนักศึกษาเป็นสิ่งที่หลีกเลี่ยงไม่ได้ เนื่องจากในขณะดำเนินงานโครงการวิจัยมีนักศึกษาที่ศึกษารายวิชาบังคับและเลือกตามแบบสอบถามที่ได้รับกลับมาเพียง 377 ฉบับเท่านั้น

ลักษณะโดยทั่วไปของโปรแกรมจะเป็นการนำเสนอการวิจัยของ นักศึกษาปริญญาตรีคณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ซึ่งข้อมูลนั้นมีทั้งส่วนที่เป็นการนำเสนอแบบที่มีตัวอักษร และมีข้อมูลในส่วนที่เป็นแบบฟอร์มให้ลูกค้ากรอก ดังนั้นการนำเสนอข้อมูลต่างที่ให้กรอกทางแบบฟอร์มนั้นจึงสำคัญต่อการรันGAILเป็นอย่างมากเพื่อช่วยในความสะดวกสบายแล้วยังมีความรวดเร็วกว่าการทำงานแบบที่ไม่มีหน้าจอให้กรอกแบบฟอร์ม อีกทั้งเพื่อเพิ่มความสวยงามเพื่อเป็นการดึงดูดความสนใจ ของผู้ที่สนใจในโปรแกรมเกล(GAIL) นี้

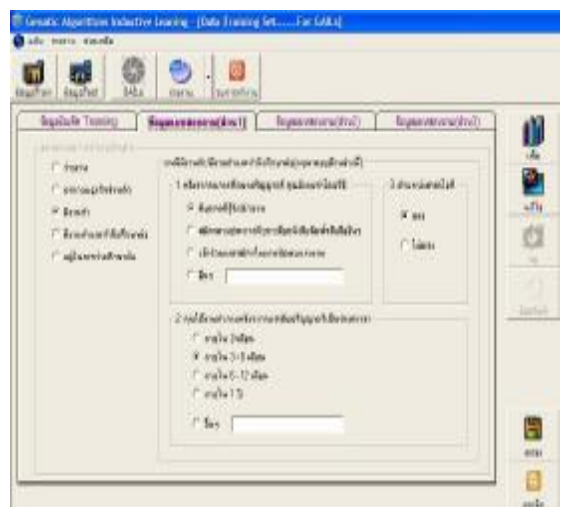
โครงสร้างโดยทั่วไป



รูปที่ 2 : รูปแสดงหน้าต่างการทำงานหลัก



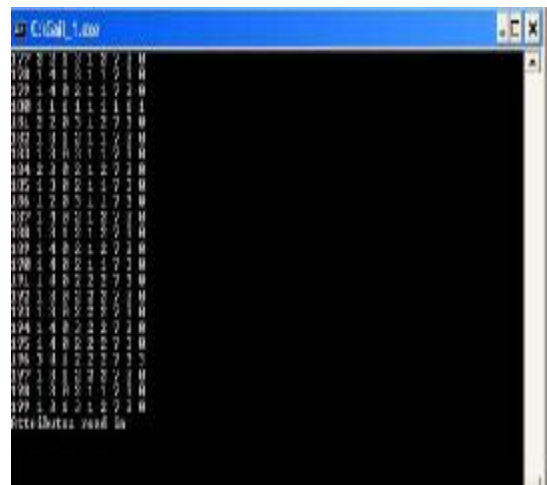
รูปที่ 3 : หน้าต่างแสดงการป้อนข้อมูลเบื้องต้นของบันทึก



รูปที่ 4 : หน้าต่างแสดงการป้อนข้อมูลแบบสอบถาม (ส่วนที่1)



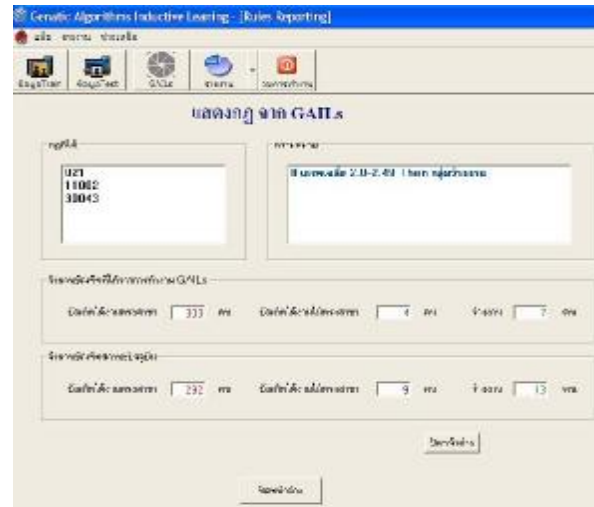
รูปที่ 5 : หน้าต่างแสดงการป้อนข้อมูลแบบสอบถาม (ส่วนที่2)



รูปที่ 7 : แสดงรูปการจำแนกข้อมูลของ Train



รูปที่ 6 : หน้าต่างแสดงการป้อนข้อมูลแบบสอบถาม (ส่วนที่3)



รูปที่ 8 : หน้าต่างรายงานกฎที่ได้จากการทำงานของ
เกล

หน้าต่างแสดงผลที่ได้จากการทำงาน เกล พร้อมทั้งบอกสถานะปัจจุบันของบัณฑิต ที่ได้จากแบบสอบถามทั้งหมด มีงานทำ ว่างงาน รวมทั้งบอกด้วยว่าเปอร์เซ็นต์ค่าเฉลี่ยดังรูปที่ 8

กิตติกรรมประกาศ

การวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดี ผู้วิจัยขอกราบขอบพระคุณ อาจารย์รัชณี กัลยาวิทย์ และดร. กิตติชัย ล้วนยานนท์ อาจารย์ที่ปรึกษาโครงการวิจัยนี้ ที่กรุณาให้คำปรึกษา ช่วยแก้ปัญหา ให้ความรู้ ข้อเสนอแนะ แนวคิด และความคิดเห็น อันเป็นประโยชน์อย่างยิ่งต่อการดำเนินโครงการวิจัย ขอขอบพระคุณคณะกรรมการทุกท่านที่ให้ความกรุณาตรวจสอบโครงการวิจัยนี้ขอขอบคุณ คุณจรรยารัตน์ พุกขานันท์ ที่ให้ความช่วยเหลือมาโดยตลอด และตอบคำถามเกี่ยวกับเทคนิค GAIL รวมทั้งให้ข้อคิดเห็นแก่ผู้วิจัย และฝ่ายงานทะเบียน ที่จัดเตรียมข้อมูลให้นำมาใช้ในโครงการวิจัย และขอขอบคุณคณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ที่ได้ให้การสนับสนุนด้านอุปกรณ์และทุนการศึกษาและขอขอบคุณโครงการอุตสาหกรรมสำหรับนักศึกษาปริญญาตรี (IPUS) ของสำนักกองทุนสนับสนุนการวิจัย (สกว.) ซึ่งให้ทุนสนับสนุนโครงการในครั้งนี้

เอกสารอ้างอิง

1. ระเบียบสถาบันเทคโนโลยีพระจอมเกล้าธนบุรี ว่าด้วยการศึกษาระดับบัณฑิตศึกษา พ.ศ. 2540, หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี พ.ศ. 2539 (ฉบับปรับปรุงครั้งที่ 1). หน้า 40-42.
2. Alcock, R.J. and Manolopoulos, Y., 1999, "Using Genetic Algorithms for Inductive Learning," 3rd IMACS Inter. Multiconference on Circuits, Systems, Communications, and Computer (CSCC'99), Athens, Greece, July 4-7.
3. Michalski, R.S., Bratko, I. and Kubat, M. (Eds.), 1998, *Machine Learning and Data Mining: Methods and Applications*, Chichester, John Wiley, pp. 59-75.
4. Rich, E. and Knight, K., 1991, *Artificial Intelligence*, 2nd ed., Singapore, McGraw-Hill, pp. 457-458.
5. Michie, D., Spiegelhalter, D.J. and Taylor, C.C. (Eds.), 1994, *Machine Learning, Neural and Statistical Classification*, Chichester, Ellis Horwood, pp. 1-6.
6. Kamber, M., Winstone, L., Gong, W., Cheng, S. and Han, J., 1997, "Generalization and Decision Tree Induction: Efficient Classification in Data Mining," *Proceedings Seventh International Workshop on Research Issues in Data Engineering*, pp. 111-120.
7. *Iris Plants Database*, 2000, <ftp://ftp.ics.uci.edu/machine-learning-databases/>
8. Sarabjot, A.S. and Hughes, J.G., 2000, "Brief Overview of Data Mining,"

- <http://inchinn.infj.ulst.ac.uk/htdocs/.%5cwhite.html>
9. Goldberg, D.E., 1989, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Reading, Addison-Wesley, pp. 1-10.
 10. Jang, J-S. R., Sun, C-T. and Mizutani, E., 1997, *Neuro-Fuzzy and Soft Computing: Computational Approach to Learning and Machine Intelligence*, New Jersey, Prentice-Hall, pp. 175-178.
 11. Zhang, D. and Currim F., 2000, "Data Mining," <http://misdb.bpa.arizona.edu/~mis96g/Reports/DataMining/report1.htm>
 12. Jushi, K.P., 1997, "Analysis of Data Mining Algorithms," http://www.gl.umbc.edu/~kjoshi1/data-mine/proj_rpt.htm
 13. Russell, S.J. and Norvig, P., 1995, *Artificial Intelligence A Modern Approach*, New Jersey, Prentice-Hall, p. 529.
 14. Oliver, J., 1994, "Finding Decision Rules with Genetic Algorithms," *AI Expert*, Vol. 9, No. 3, Mar., pp. 33-39.
 15. Boggess, J.E., 1995, "Using a Genetic Algorithm to Evolve a Set of Rules for Classifying Roads in Satellite Images," <http://www2.pitt.edu/~csna/csna95abst.html>
 16. Pei, M., Ding, Y., Punch III, W.F. and Goodman, E.D., 1995, "Genetic Algorithms for Classification and Feature Extraction," <http://www2.pitt.edu/~csna/csna95abst.html>
 17. Zuzek, M., Friedrich, B., Cestnik, B., Karalic, A. and Zimerman, A., 1995, "Medium Optimization by Genetic Algorithms and Inductive Learning," <http://www-ai.ijs.si/AramKaralic/bibliography/1995a.html>
 18. Salem, Z., 1999, "RULES Family," <http://intell-lab.engi.cf.ac.uk/Ziad/Rules.html>
 19. Pham, D.T. and Alcock, R.J., *Synergistic Classification Systems for Wood Defect Identification*, London, Springer-Verlag, pp. 1-27.
 20. Sagioglu, S., 1995, *Synergistic Neural Networks in Orientation Determination of Inertia Sensor*, Ph.D. Thesis, University of Wales Cardiff, pp. 115-118.