

รายงานการวิจัย

เรื่อง

ระบบอัจฉริยะสหวิทยาการสำหรับวิเคราะห์การศึกษาขั้นสูงโดยทฤษฎีเซต,  
สถิติและระเบียบวิธีเชิงตัวเลข

Multidisciplinary Intelligent Advance Instructional System by Rough Set  
Theory, Statistics and Numerical Method



โดย  
ดร. พรรณทิพย์ ภัทรอินทากร

นายพรชัย ชัยสนิท

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

RCH

QA

8.7

พ ๒๖๒๙

เลขหมู่

116861

เลขทะเบียน

16 ส.ย. 2554

วัน,เดือน,ปี

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัยเพื่อพัฒนาเศรษฐกิจและสังคมด้วย

วิทยาศาสตร์และเทคโนโลยี

สำนักงานคณะกรรมการวิจัยแห่งชาติ

ประจำปีงบประมาณ พ.ศ. 2551

b.12329009

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ระบบอัจฉริยะสหวิทยาการสำหรับวิเคราะห์การศึกษาระดับสูงโดยทฤษฎีเซต,  
สถิติและระเบียบวิธีเชิงตัวเลข**

**Multidisciplinary Intelligent Advance Instructional System by Rough Set  
Theory, Statistics and Numerical Method**



**งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัยเพื่อพัฒนาเศรษฐกิจและสังคมด้วย  
วิทยาศาสตร์และเทคโนโลยี  
สำนักงานคณะกรรมการวิจัยแห่งชาติ  
ประจำปีงบประมาณ พ.ศ. 2551**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## กิตติกรรมประกาศ

รายงานการวิจัยเรื่อง “ระบบอัจฉริยะสหวิทยาการสำหรับวิเคราะห์การศึกษาขั้นสูงโดยทฤษฎีรีฟเซต, สถิติและระเบียบวิธีเชิงตัวเลข” (Multidisciplinary Intelligent Advance Instructional System by Rough Set Theory, Statistics and Numerical Method) เป็นโครงการวิจัยที่จัดทำขึ้นโดยได้รับการสนับสนุนจากสำนักงานคณะกรรมการวิจัยแห่งชาติ เพื่อนำข้อมูลไปใช้ประกอบการจัดทำแนวทางการศึกษาวิจัย สำหรับเพื่อพัฒนาเศรษฐกิจและสังคมด้วยวิทยาศาสตร์และเทคโนโลยี ประเภทการวิจัยประยุกต์ สาขาวิจัยวิทยาศาสตร์กายภาพและคณิตศาสตร์และสาขาวิจัยการศึกษา

คณะผู้วิจัยขอขอบคุณคณะกรรมการสำนักงานคณะกรรมการวิจัยแห่งชาติ รวมถึงคณะทำงานอำนวยการโครงการวิจัย บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังซึ่งมี รศ.ดร.รวิวรรณ ชินะตระกูล คณบดีบัณฑิตวิทยาลัยเป็นประธาน ที่ได้ให้ข้อคิดเห็น ข้อเสนอแนะและให้ความร่วมมือแก่นักวิจัยอย่างดียิ่งและขอขอบคุณ รศ.ไพโรบลย์ พันธรักษ์พงษ์ หัวหน้าภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์และ รศ.ดร.ธีรวัฒน์ มงคลอัครรัตน์ คณบดีคณะวิทยาศาสตร์ที่ช่วยติดตามดูแลงานวิจัย โดยเฉพาะอย่างยิ่งขอขอบคุณมูลนิธิส่งเสริมโอลิมปิกวิชาการ และพัฒนามาตรฐานวิทยาศาสตร์ศึกษา ในพระอุปถัมภ์สมเด็จพระเจ้าพี่นางเธอ เจ้าฟ้ากัลยาณิวัฒนา กรมหลวงนราธิวาสราชนครินทร์ (สอวน.) และ รศ.เย็นใจ สมวิเชียร กรรมการและเหรัญญิกมูลนิธิ สอวน. ตลอดจนหน่วยงานที่เกี่ยวข้องที่เปิดโอกาสให้สัมภาษณ์ เก็บข้อมูลและให้ข้อมูลที่เป็นประโยชน์ยิ่ง

ทั้งนี้ ขอขอบคุณคณะกรรมการจากสำนักงานคณะกรรมการวิจัยแห่งชาติที่ได้อนุมัติงบประมาณอุดหนุนวิจัยเป็นค่าใช้จ่ายของโครงการวิจัยนี้ และยิ่งขอขอบคุณฝ่ายเลขานุการคณะทำงานที่ให้ความร่วมมือแก่นักวิจัยอย่างดียิ่ง

ดร.พรรณทิพย์ ภัทรอินทการ

นายพรชัย ชัยสนิท

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

2551

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทคัดย่อ

การใช้คณิตศาสตร์แขนงเก่า (hard computing) เช่น การใช้ตัวแบบทางคณิตศาสตร์ การวิเคราะห์เชิงตัวเลขในการวิเคราะห์ข้อมูลต่าง ๆ นั้น จะต้องมีการวิเคราะห์ความผิดพลาด (error analysis) ด้วยเสมอว่าในการสร้างตัวแบบนั้นจะให้ความผิดพลาดไม่เกินขอบเขตตามที่ได้พิสูจน์ไว้ แต่อย่างไรก็ดี สำหรับข้อมูลจริงที่มีขนาดใหญ่และมีค่าคลาดเคลื่อนในข้อมูลอยู่แล้วเป็นจำนวนมาก จะให้ค่าความผิดพลาดของตัวแบบดังกล่าวนั้น แบบเส้นโค้งเลขชี้กำลัง ทำให้การใช้คณิตศาสตร์แขนงเก่าดังกล่าวจึงไม่สามารถวิเคราะห์ได้ถูกต้องเพียงพอ

ในปี 2523, Zdzislaw Pawlak ได้คิดค้นคณิตศาสตร์แขนงใหม่ขึ้นจากแนวคิดเรื่องเซต นั่นคือ กราฟสายงาน (flow graph) ซึ่งเป็นคณิตศาสตร์แขนงใหม่ชนิด soft computing โดยจะคำนวณค่าคลาดเคลื่อนโดยตรงและ ไม่ทำให้เกิดความผิดพลาดแบบเส้นโค้งเลขชี้กำลังสำหรับข้อมูลจริงขนาดใหญ่มาก จึงให้ความถูกต้องมากขึ้น อีกทั้งยังสามารถคำนวณด้วยวิธีการประมวลผลแบบขนานได้ จึงทำให้คำนวณได้เร็วขึ้นและใช้ความจุหน่วยเก็บน้อยลง ด้วยประสิทธิภาพดังกล่าว ทำให้วิธีนี้ได้ถูกนำไปใช้ในศาสตร์แขนงต่าง ๆ อย่างกว้างขวาง อาทิเช่น preference analysis, conflict analysis, rule analysis, propositional calculus, data mining, granular computing

จากข้อดีทั้งการใช้ความจุหน่วยเก็บน้อยลง ความเร็วในการคำนวณและความถูกต้องของวิธีนี้ในงานวิจัยเบื้องต้นดังกล่าว ทางผู้วิจัยจึงใช้กราฟสายงานเพื่อวิเคราะห์ข้อมูลคะแนนคณิตศาสตร์โดยคำนวณสัมประสิทธิ์ดังต่อไปนี้ *strength*, *certainty* และ *coverage* เพื่อวิเคราะห์ความสัมพันธ์ของปัจจัยต่าง ๆ รวมถึงสร้างกฎการจำแนกคะแนนของนักเรียนจากความสัมพันธ์ของปัจจัยนั้น การนำความรู้หลาย ๆ สาขามารวมกันเพื่อวิจัยและพัฒนาเทคโนโลยีใหม่ ๆ หรือการวิจัยแบบสหวิทยาการ (multidisciplinary) กำลังเป็นที่ได้รับความสนใจในวงการวิจัยทั้งระดับนานาชาติและเป็นที่ตระหนักรู้ว่าคณิตศาสตร์ โดยเฉพาะอย่างยิ่งคณิตศาสตร์แขนงใหม่ (modern mathematics) มีบทบาทสำคัญยิ่งในการศึกษาวิจัยเพื่อค้นพบองค์ความรู้ใหม่ที่มีคุณภาพสูง ดังนั้น การวิจัยแบบบูรณาการซึ่งเป็นการร่วมมือกันของนักวิชาการ ผู้เชี่ยวชาญและนักวิจัยจากหลากหลายสาขาในประเทศไทยนั้น จะนำไปสู่การพึ่งพาตนเองซึ่งเป็นสิ่งสำคัญยิ่งต่อการพัฒนาประเทศไปอย่างสมดุลและยั่งยืน สู่สังคมแห่งภูมิปัญญาและการเรียนรู้อันจะเป็นรากฐานที่มั่นคงของประเทศไทยสืบต่อไป

ผู้วิจัยจึงนำเสนอโครงการวิจัยเรื่อง "ระบบอัจฉริยะบูรณาการสำหรับวิเคราะห์การศึกษาระดับสูง โดยทฤษฎีกราฟเซต, สถิติและระเบียบวิธีเชิงตัวเลข" โดยงานวิจัยชิ้นนี้ได้บูรณาการเทคโนโลยีใหม่ ทางด้านคอมพิวเตอร์ของทศวรรษนี้ที่ชื่อว่า ปัญญาประดิษฐ์ (artificial intelligence) เข้ากับ

ทฤษฎีทางคณิตศาสตร์แขนงใหม่คือ รัฟเซต (rough set theory) และสถิติ (statistics) เพื่อการวิเคราะห์ข้อมูลทางด้านการศึกษาระดับสูงอันจะเป็นการก้าวกระโดดจากการคำนวณแบบ hard computing สู่อารมณ์คำนวณแบบ soft computing และเป็นการเน้นใช้เทคโนโลยีอันทันสมัย เข้าวิเคราะห์คุณภาพคนและการศึกษา ที่มีใช้การวิจัยเชิงสังคมศาสตร์ แต่เป็นเชิงวิทยาศาสตร์ประยุกต์ เพื่อตอบสนองประเด็นยุทธศาสตร์ ของแผนการพัฒนาประเทศตามแผนพัฒนาเศรษฐกิจและสังคมแห่งชาติฉบับที่ 10 (พ.ศ. 2550-2554) อันเป็นแนวทางการวิจัยรูปแบบที่หาได้ยากในการแผนวิจัยทั่วไป

ซึ่งการวิจัยเชิงลึกนี้มิได้เน้นการแข่งขัน หรือการเปรียบเทียบกับระบบการศึกษาในระดับนานาชาติ แต่มุ่งเน้นจะทำให้เข้าใจกลไกของระบบการศึกษาไทยอย่างถ่องแท้ในมุมมองใหม่ที่ตอบสนองนโยบาย “ความเป็นอยู่อย่างพอเพียง” โดยเน้นการเข้าใจและพึ่งพาตนเองของระบบการศึกษาและนักเรียน อันจะนำมาซึ่งรากฐานอันเข้มแข็งของคุณภาพคน การศึกษา และสังคมไทย กรณีศึกษาที่พิจารณาในงานวิจัยนี้คือ นักเรียนที่มีความสามารถพิเศษทางคณิตศาสตร์จากมูลนิธิส่งเสริมโอลิมปิกวิชาการและพัฒนามาตรฐานวิทยาศาสตร์ศึกษา ในพระอุปถัมภ์สมเด็จพระเจ้าพี่นางเธอ เจ้าฟ้ากัลยาณิวัฒนา กรมหลวงนราธิวาสราชนครินทร์ (สอวน.) ที่ศูนย์โรงเรียนสวนกุหลาบวิทยาลัย

## ABSTRACT

---

Mathematical rough set theory has attracted both practical and theoretical researchers. A significant extension of rough set theory is called *flow graphs*, invented by Zdzislaw Pawlak (1980). It is a knowledge representation in the form of information flow. Mathematical flow graphs are in soft computing area which provides less complexity, times and resources. Thus, flow graph is a promising approach to analyze data flow, decision trees, decision rules, probability learning, etc.

In this research, we present their connections to pertinent techniques and propose a new extension. We discuss some important properties of flow graphs. A case study considered in this research is from the Promotion of Academic Olympiads and Development of Science Education Foundation (POSN). Several coefficients: strength, certainty and coverage are analyzed to students' score successfully.

## รายงานวิจัยฉบับสมบูรณ์

**โครงการ** ระบบอัจฉริยะสหวิทยาการสำหรับวิเคราะห์การศึกษาชั้นสูงโดยทฤษฎีรีฟเซต, สถิติและระเบียบวิธีเชิงตัวเลข  
Multidisciplinary Intelligent Advance Instructional System by Rough Set Theory, Statistics and Numerical Method

**ประเภทการวิจัย:**

การวิจัยประยุกต์

**สาขาการวิจัย:**

สาขาวิจัยวิทยาศาสตร์กายภาพและคณิตศาสตร์และสาขาวิจัยการศึกษา

**ระยะเวลาโครงการ:** 1 ปี

**การดำเนินงาน**  ได้ดำเนินงานตามแผนที่วางไว้  
 ได้ดำเนินงานล่าช้ากว่าแผนที่วางไว้

**คำสำคัญ (Keyword) ของโครงการวิจัย:** Rough set theory (RST), Hybrid intelligent system, Statistics, Survival analysis, Hard computing, Soft computing, Computer science, Applied Mathematics, Artificial intelligence (AI), Decision table, Indiscernibility relation, Score analysis, Decision rule, Artificial intelligence (AI), Accuracy

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญเรื่อง

เรื่อง	หน้า
กิตติกรรมประกาศ	I
บทคัดย่อ	II
Abstract	IV
รายงานวิจัยฉบับสมบูรณ์	V
บทนำ	1
หลักการและเหตุผล	1
คณิตศาสตร์โอลิมปิกสทวน. (POSN Mathematical Olympiad POSN-MO)	7
ส่วนที่ 1 (Part I)	
ผลการวิจัยส่วนที่ 1	1-1
บรรณานุกรม	1-7
ส่วนที่ 2 (Part II)	
1. Rough sets, flow graphs and their extension to association rules	2-1
1.1 Rough Set Theory	2-1
1.2 Flow Graphs	2-2
Flow Graphs	2-2
Normalized Flow Graphs	2-4
Paths and Connections	2-5
Dependencies	2-6
Decision Algorithm	2-7
Association Rules	2-8
1.3 Illustrative Example	2-9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญเรื่อง (ต่อ)

เรื่อง	หน้า
1.4 Experimental Results of the POSN Data Set	2-11
2 Rough Sets, Flow Graphs, Fuzzy Sets and their Extension to Decision Algorithm	2-13
2.1 Fuzzy Set Theory	2-13
2.2 Decision Rules, Decision Algorithm and Flow Graph	2-14
2.3 New Certainty and Coverage Coefficients for Decision Rules from Relaxed Mutually Exclusive	2-15
2.4 Flow Graphs with Fuzzy Attributes	2-19
2.5 Illustrative Example	2-21
2.6 Inverse Flow Graphs	2-25
2.7 Experimental Results	2-27
3 Suggestions and Extensions of Flow Graphs' Research	3-32
4 Conclusion	3-33
References	3-34
ภาคผนวก ประโยชน์ทางการประยุกต์และการเผยแพร่ผลงานวิจัย	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญตาราง

ตาราง	หน้า
ตารางที่ 1 แสดงการดำเนินงานในภาพรวมของสวอน. และเหรียญรางวัลจากการแข่งขันโอลิมปิกวิชาการระหว่างประเทศ พ.ศ. 2543 – ปัจจุบัน (คณิตศาสตร์ คอมพิวเตอร์ เคมี ชีววิทยาและฟิสิกส์)	5
<b>ส่วนที่ 2 (Part II)</b>	
Table 1. Decision table with fuzzy attributes	2-14
Table 2. An example decision table	2-21
Table 3. Decision table with fuzzy attributes	2-22
Table 4. Strength $\sigma(N_i, A_j)$ between condition and decision attribute	2-22
Table 5. Strength $\sigma(A_j, D_k)$ between condition and decision attribute	2-22
Table 6. Certainty and Coverage coefficients of $N_i$ and $A_j$	2-23
Table 7. Certainty and Coverage coefficients of $A_j$ and $D_k$	2-23
Table 8. Value of normalized strength between condition and decision attributes	2-23
Table 9. Certainty and Coverage coefficients of $N_i$ , $A_j$ and $D_k$	2-24
Table 10. Certainty and Coverage coefficients of $N_i$ and $D_k$	2-24
Table 11. Certainty and Coverage coefficients of $N_i \vee A_j$ and $D_k$	2-24

# สารบัญรูป

รูป	หน้า
<b>ส่วนที่ 1 (Part I)</b>	
รูปที่ 1.1 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามเพศ	1-1
รูปที่ 1.2 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามอายุ	1-1
รูปที่ 1.3 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามระดับ	1-2
รูปที่ 1.4 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามการอ่านหนังสือทบทวนระหว่าง เทอมโดยเฉลี่ยในแต่ละสัปดาห์	1-3
รูปที่ 1.5 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามการอ่านหนังสือเพื่อเตรียมสอบ โดยเฉลี่ยในแต่ละสัปดาห์	1-3
รูปที่ 1.6 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามวิชาที่ตีพิมพ์	1-4
รูปที่ 1.7 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามสาขาวิชาคณิตศาสตร์ที่ชอบ	1-5
รูปที่ 1.8 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามสาขาวิชาคณิตศาสตร์ที่คิดว่า เข้าใจ	1-5
รูปที่ 1.9 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามสาขาวิชาคณิตศาสตร์ที่ฝึก ทำโจทย์	1-6

## ส่วนที่ 2 (Part II)

Fig. 1. Flow graph initial data of voting problem (adapted from [9]).	2-3
Fig. 2. Social class, age group and party votes flow graph.	2-10
Fig. 3. Combined of flow graph.	2-10
Fig. 4. Flow graph of initial Olympiad mathematics' students score from POSN.	2-11
Fig. 5. Comparison of <i>sup</i> , <i>conf</i> and <i>accuracy/100</i> from four cases of POSN data set.	2-13
Fig. 6. Flow graph G.	2-27
Fig. 7. Inverse flow graph $G^{-1}$ .	2-27

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญรูป (ต่อ)

รูป

หน้า

Fig. 8. Comparison of *cer*, *cov* of the POSN data.

2-31



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทนำ

### หลักการและเหตุผล

ในอดีตที่ผ่านมา ระบบการศึกษาและการวิจัยของไทย ยังไม่สามารถพัฒนาศักยภาพของผู้ที่มีความสามารถพิเศษได้ดีเท่าที่ควร อันเป็นสาเหตุให้เราสูญเสียบุคลากรชั้นแนวหน้าที่มีความเป็นเลิศในสาขาต่างๆ ไป อันได้แก่ เด็กและเยาวชนที่มีความสามารถพิเศษซึ่งมีอยู่เพียงประมาณร้อยละ 3 ของแต่ละด้าน/สาขา ดังนั้น ในช่วง 10 ปี ที่ผ่านมาประเทศไทยเริ่มตระหนักถึงความสำคัญของเด็กที่มีความสามารถพิเศษ โดยกำหนดนโยบายและทิศทางการดำเนินงานไว้ในแผนพัฒนาการศึกษาสำหรับเด็กและเยาวชนผู้มีความสามารถพิเศษ พ.ศ. 2541 โดยให้มีมาตรการที่จะทำให้เกิดการกระจายโอกาสครอบคลุมไปถึงกลุ่มเด็กและเยาวชนที่เป็นเป้าหมายทั่วประเทศ รวมถึงให้มีมาตรการที่จะทำให้การปฏิบัติตามนโยบายและแผนเป็นไปในทิศทางเดียวกัน และเกิดผลอย่างเป็นรูปธรรมชัดเจน

หลังจากนั้น ประเทศไทยกำหนดนโยบายต่างๆ ทางด้านการศึกษาขึ้นเพื่อการพัฒนาเด็กและเยาวชนที่มีความสามารถพิเศษ ซึ่งมีหลากหลายสาขาวิชาหนึ่งในจำนวนนั้น คือ สาขาวิชาคณิตศาสตร์เนื่องจากเป็นวิชาพื้นฐานที่สำคัญยิ่งของการประยุกต์อื่น ๆ และนำไปบูรณาการกับสาขาอื่นได้ทุกวิชา ประเทศใดมีเยาวชนที่มีความสามารถพิเศษทางด้านคณิตศาสตร์มาก ประเทศนั้นจะมีโอกาสที่จะพัฒนาเทคโนโลยีเหนือประเทศต่างๆ (สำนักงานคณะกรรมการการศึกษาแห่งชาติ, 2542)

แต่การจัดการศึกษาสำหรับเด็กที่มีความสามารถพิเศษทางคณิตศาสตร์ สิ่งแรกที่จะต้องพิจารณาคือการคัดเลือกเด็กที่มีความสามารถพิเศษทางคณิตศาสตร์อย่างแท้จริงเข้ารับการศึกษาศึกษาพิเศษ โดยการคัดเลือกมีหลายวิธีด้วยกัน สำหรับประเทศไทยได้มีการประเมินผลการจัดการศึกษาสำหรับเด็กที่มีความสามารถพิเศษของสำนักงานนโยบายและแผนการศึกษาขั้นพื้นฐาน สำนักงานคณะกรรมการการศึกษาแห่งชาติและสำนักนายกรัฐมนตรี ได้พบปัญหาในการดำเนินงานเนื่องจาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อผู้ขาดเห็นไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขาดเครื่องมือตรวจสอบศักยภาพของเด็ก ซึ่งถ้าเครื่องมือไม่มีคุณภาพดีพอ ผลกระทบจะตกอยู่กับเด็ก (สำนักงานนโยบายและแผนการศึกษาขั้นพื้นฐาน, 2536)

ด้วยเหตุนี้การที่สามารถระบุถึงตัวบ่งชี้เด็กที่มีความสามารถพิเศษทางคณิตศาสตร์ จึงนับว่าเป็นขั้นตอนหนึ่งที่สำคัญแต่มีความสลับซับซ้อนและมีลักษณะเฉพาะตัวที่แตกต่างกับสาขาอื่น ๆ หรือเด็กและเยาวชนในประเทศอื่น ๆ การวิจัยในประเทศไทยสำหรับเด็กและเยาวชนไทยเกี่ยวกับตัวบ่งชี้นี้จึงเป็นสิ่งที่ขาดเสียไม่ได้ และถือเป็นอีกขั้นตอนหนึ่ง เพื่อยังประโยชน์ในการพัฒนาการเรียนการสอนสำหรับผู้มีความสามารถพิเศษทางคณิตศาสตร์ให้ได้รับการพัฒนาจนบรรลุศักยภาพสูงสุดของแต่ละบุคคลต่อไป

บุญทัน อบุญญม [1] ได้อธิบายถึงลักษณะของเด็กที่เรียนเก่งในวิชาคณิตศาสตร์ไว้ดังนี้

- 1) มีสติปัญญาดีระดับ I.Q. 120 หรือสูงกว่า สามารถเรียนรู้และเข้าใจสิ่งที่เป็นนามธรรม และความคิดรวบยอดทางคณิตศาสตร์เป็นอย่างดี
- 2) มีความจำดีมาก ความรู้พื้นฐานเดิมในทางคณิตศาสตร์ดี จึงสามารถเรียนรู้ความคิดรวบยอดใหม่ได้รวดเร็ว
- 3) มีความกระตือรือร้น และสนใจในสิ่งต่างๆ รอบตัว ชอบค้นคว้าหาความรู้อยู่เสมอ จึงมีความสนุกสนานในการเรียนรู้สิ่งใหม่
- 4) มีความสามารถในด้านภาษา การอ่าน การเขียน สามารถอ่านคำสั่งใจหายได้ และแปลความได้รวดเร็ว จึงทำให้สามารถทำโจทย์ปัญหาทางคณิตศาสตร์ที่ซับซ้อนได้
- 5) สามารถทำงานที่ทำท่ายได้เป็นระยะเวลาานาน
- 6) สามารถหาวิธีแก้ปัญหาทางคณิตศาสตร์ได้หลายๆ วิธีและสามารถเลือกวิธีที่ดีที่สุด
- 7) สามารถถ่ายโยงสิ่งที่เรียนแล้วให้เข้ากับสถานการณ์ใหม่ได้
- 8) เป็นผู้ที่มีแรงจูงใจภายใน เพราะชอบวิชาคณิตศาสตร์จึงทำให้มีความสนใจ และเรียนได้ดี

ศักดิ์ดา บุญโต [2] ได้กล่าวว่านักเรียนที่มีความสามารถพิเศษทางคณิตศาสตร์นั้นจะเป็นเด็กที่มีลักษณะดังนี้ คือ จะเป็นเด็กที่มีความสนใจด้านจำนวน ตัวเลข การคำนวณ และสัญลักษณ์อย่างต่อเนื่อง มีกลยุทธ์การคิดทางคณิตศาสตร์ได้อย่างพลิกแพลง แยกแยะ สร้างสรรค์และสมเหตุสมผล เด็กจะมองเห็นความสำคัญของมิติได้ มีความสามารถทางคณิตศาสตร์ด้านใดด้านหนึ่งหรือหลายๆ ด้านรวมกันอย่างโดดเด่น มากกว่าเด็กในวัยเดียวกัน รวมทั้งเด็กสามารถนำความรู้ทางคณิตศาสตร์ไปใช้ได้ไปในหลายๆ รูปแบบ มีการคิดวิเคราะห์ มีกระบวนการแก้ปัญหาที่แปลกใหม่ และมีเอกลักษณ์เฉพาะตน

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพชรลดา สีหะวงศ์ [3] ได้ศึกษาข้อมูลของนักเรียน และได้สรุปคุณลักษณะที่โดดเด่นของเด็กที่มีความสามารถพิเศษทางคณิตศาสตร์ ซึ่งมีคุณลักษณะด้านต่าง ๆ สูงโดดเด่นกว่าเด็กทั่วไปดังต่อไปนี้

- 1) องค์ประกอบด้านเชาวน์ปัญญา
- 2) องค์ประกอบด้านการแก้ปัญหาอย่างสร้างสรรค์ทางคณิตศาสตร์ ซึ่งสอดคล้องกับแนวคิดของนักวิจัยท่านอื่น ๆ ที่กล่าวถึงแบบทดสอบวัดการแก้ปัญหาอย่างสร้างสรรค์ทางด้านคณิตศาสตร์ไว้ว่า ควรมีเนื้อหาที่น่าสนใจทางด้านทฤษฎีจำนวน พีชคณิต และเรขาคณิต
- 3) องค์ประกอบด้านความถนัดทางคณิตศาสตร์ มี 3 ตัวบ่งชี้คือสมรรถภาพด้านจำนวน สมรรถภาพด้านเหตุผล และสมรรถภาพด้านมิติสัมพันธ์

ถัดไปคือรายละเอียดจาก [4], ที่ผ่านมา การเปิดโอกาสให้เด็กและเยาวชนผู้มีความสามารถพิเศษด้านคณิตศาสตร์ได้แสดงความสามารถในระดับสากลที่สำคัญ คือ การแข่งขันคณิตศาสตร์โอลิมปิกระหว่างประเทศ โดยจากข้อมูลของมูลนิธิส่งเสริมโอลิมปิกวิชาการและพัฒนามาตรฐานวิทยาศาสตร์ศึกษา ในพระอุปถัมภ์สมเด็จพระเจ้าพี่นางเธอ เจ้าฟ้ากัลยาณิวัฒนา กรมหลวงนราธิวาสราชนครินทร์ (สอวน.) การที่ประเทศไทยส่งนักเรียนเข้าแข่งขันคณิตศาสตร์โอลิมปิกระหว่างประเทศมีเป้าหมายที่จะวัดมาตรฐานมัธยมศึกษาของประเทศไทยกับสากล และนำความรู้จากการอบรมแข่งขัน มาช่วยยกระดับมาตรฐานการศึกษาวิทยาศาสตร์และคณิตศาสตร์ของประเทศไทย แต่การดำเนินการส่งเยาวชนไทยแข่งขันโอลิมปิกวิชาการตั้งแต่ พ.ศ. 2532 นับถึงปัจจุบันเป็นเวลากว่า 20 ปี แล้ว ยังไม่บรรลุเป้าหมาย เพราะไม่ได้มีการนำความรู้และวิธีที่ได้จากการอบรมเพิ่ม รวมทั้งแนวข้อสอบ วิธีการสอบและสูตรความสำเร็จของประเทศที่ชนะการแข่งขันมาใช้ประโยชน์ในการปรับปรุงหลักสูตรการเรียนการสอนวิทยาศาสตร์ คณิตศาสตร์ของประเทศไทย ให้ขยับใกล้มาตรฐานสากล

นอกจากนี้ การแข่งขันโอลิมปิกวิชาการเป็นมาตรฐานมัธยมศึกษาระดับสากล ซึ่งแตกต่างจากมาตรฐานมัธยมศึกษาของประเทศไทย ข้อสอบซับซ้อน ต้องอาศัยความรู้ทั้งหมดที่เรียนมาและกระบวนการศึกษาทางวิทยาศาสตร์เพื่อแก้ปัญหา รวมทั้งเน้นภาคปฏิบัติที่เด็กและเยาวชนไทยถึงแม้จะมีความสามารถทัดเทียมกับนานาชาติ แต่เด็กกลับขาดทักษะดังกล่าวอย่างมาก นอกจากนี้ นักเรียนที่ได้ประโยชน์จากการเข้าค่ายอบรมเพิ่มในโครงการโอลิมปิกวิชาการในช่วงแรกมีเพียง 125 คนต่อปี ซึ่งน้อยมาก ถึงแม้ว่าการยกระดับมาตรฐานด้านวิทยาศาสตร์และ

คณิตศาสตร์ในระดับมัธยมศึกษาให้เทียบเท่าสากลจะกระทำได้ยากในระยะเวลานั้น แต่การ  
แข่งขันโอลิมปิกวิชาการยังสามารถนำมาใช้ประโยชน์เป็นตัวเร่งนอกระบบได้

สมเด็จพระเจ้าพี่นางเธอเจ้าฟ้ากัลยาณิวัฒนา กรมหลวงนราธิวาสราชนครินทร์ ได้ทรง  
เล็งเห็นถึงความสำคัญและปัญหาในการพัฒนาการศึกษาวิทยาศาสตร์ คณิตศาสตร์ เพื่อให้ได้  
ระดับมาตรฐานสากลจึงได้ทรงพระราชทานเงินส่วนพระองค์ให้ ศาสตราจารย์ศักดิ์ ศิริพันธ์  
(นายกสมาคมวิทยาศาสตร์แห่งประเทศไทยในพระบรมราชูปถัมภ์ ในขณะนั้น) นำไปดำเนินการ  
จัดตั้ง “มูลนิธิส่งเสริมโอลิมปิกวิชาการและพัฒนามาตรฐานวิทยาศาสตร์ศึกษา ในพระอุปถัมภ์  
สมเด็จพระเจ้าพี่นางเธอ เจ้าฟ้ากัลยาณิวัฒนา กรมหลวงนราธิวาสราชนครินทร์ (สอวน.)” และทรง  
รับเป็นองค์ประธานมูลนิธิด้วยทรงเห็นว่า การดำเนินการปรับมาตรฐานการศึกษาให้เทียบเท่าสากล  
โดยใช้วิธีการเร่งนอกระบบจำเป็นต้องอาศัยองค์การอิสระที่มีความสามารถ และความคล่องตัวใน  
การทำงานไม่ต้องยึดติดกับระเบียบของทางราชการและมูลนิธิ สอวน. มีคณะกรรมการที่มีความ  
แน่นแฟ้นและมีความสามารถที่จะช่วยผลักดันโครงการที่ยากยิ่งแต่สำคัญมากนี้ ให้บรรลุเป้าหมาย  
ได้นับเป็นพระมหากรุณาธิคุณเป็นล้นพ้นหาที่สุดมิได้ที่ทรงมีต่อวงการวิทยาศาสตร์ คณิตศาสตร์  
และวงการการศึกษาของประเทศไทยโดยรวม

เพื่อสนองพระดำริขององค์ประธานมูลนิธิ สอวน. และเพื่อประโยชน์ของการศึกษา  
วิทยาศาสตร์และคณิตศาสตร์ขอประเทศไทยในระยะยาว มูลนิธิ สอวน. จึงได้จัดทำโครงการ  
ส่งเสริมโอลิมปิกวิชาการและพัฒนามาตรฐานวิทยาศาสตร์ คณิตศาสตร์ศึกษาขึ้น โดยมุ่งเน้นไปที่  
กลุ่มนักเรียนมัธยมศึกษาตอนปลายทั่วประเทศที่มีความพร้อมที่จะเรียนรู้ก่อน รวมทั้งพัฒนา  
ความรู้ วิธีการสอนและทัศนคติของครูไปพร้อมๆ กัน โดยขอความร่วมมือจากคณะวิทยาศาสตร์  
มหาวิทยาลัยของรัฐทั่วประเทศหลังจากนั้นจึงจะขยายผลต่อไป

กลุ่มเป้าหมายของสอวน.

1. นักเรียนที่มีความพร้อมในด้านวิทยาศาสตร์ คณิตศาสตร์ และคอมพิวเตอร์ ไม่เกินระดับ  
มัธยมศึกษาปีที่ 5 ทั่วประเทศ
2. อาจารย์ผู้สอนวิชาคณิตศาสตร์ คอมพิวเตอร์ เคมี ชีววิทยา และฟิสิกส์ที่มีความพร้อม
3. โรงเรียนในระดับมัธยมศึกษา ที่มีความพร้อมในด้านการเรียนการสอนวิชาวิทยาศาสตร์  
คณิตศาสตร์ และคอมพิวเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลในตารางที่ 1 คือผลการดำเนินงานในภาพรวมของสอวน. และเหรียญรางวัลจากการแข่งขันโอลิมปิกวิชาการระหว่างประเทศในปี พ.ศ. 2543 – ปัจจุบัน ในสาขาวิชา คณิตศาสตร์ คอมพิวเตอร์ เคมี ชีววิทยาและฟิสิกส์

ตารางที่ 1 แสดงการดำเนินงานในภาพรวมของสอวน. และเหรียญรางวัลจากการแข่งขันโอลิมปิกวิชาการระหว่างประเทศ พ.ศ. 2543 – ปัจจุบัน  
(คณิตศาสตร์ คอมพิวเตอร์ เคมี ชีววิทยาและฟิสิกส์)<sup>1</sup>

รุ่น (ปี)	ผู้สมัคร (ปี)	ค่ายสอ วน.1 (ปี)	ค่ายสอ วน.2 (ปี)	ผู้แทน ประเทศ ไทย (คน)	เหรียญรางวัลโอลิมปิก				
					ทอง	เงิน	ทองแดง	เกียรติคุณ ประกาศ	รวม
1 (2543)	11,893 (2543)	2,611 (2543)	1,394 (2544)	4	1	3	-	-	4
2 (2544)	14,755 (2544)	2,536 (2544)	1,203 (2545)	16	8	3	2	3	16
3 (2545)	18,326 (2545)	2609 (2545)	1,359 (2546)	21	2	6	11	2	21
4 (2546)	23,786 (2546)	3,060 (2546)	1,134 (2547)	23	8	12	3	-	23
5 (2547)	29,625 (2547)	2,927 (2547)	1,457 (2548)	22	6	10	5	-	21
6 (2548)	36,251 (2548)	3,180 (2548)	1,647 (2549)	23	6	13	4	-	23
7 (2549)	44,318 (2549)	3,134 (2549)	1,780 (2550)	23	11	10	2	-	23
8 (2550)	45,689 (2550)	3,092 (2550)	1,686 (2551)	พ.ศ. 2552	ก.ค. 2552				

<sup>1</sup> ข้อมูลจาก <http://www.posn.or.th/>

วิธีการดำเนินการในโครงการนี้ เป็นการส่งเสริมและพัฒนาศักยภาพทางปัญญาของนักเรียนชั้นมัธยมศึกษาตอนปลายที่มีความพร้อมทั่วประเทศ และเป็นตัวป้อนนักเรียนที่มีศักยภาพสูงทางปัญญา จำนวนหนึ่งเข้าสู่ระบบของโครงการโอลิมปิกวิชาการ มูลนิธิ สอวน. ได้ร่วมมือกับกระทรวงศึกษาธิการและคณะวิทยาศาสตร์ มหาวิทยาลัยของรัฐจัดตั้งศูนย์ สอวน. ขึ้นทั่วประเทศ เพื่อเป็นศูนย์พัฒนาศักยภาพของนักเรียน (และครู) ทางด้านวิทยาศาสตร์และคณิตศาสตร์ศึกษา จำนวน 16 ศูนย์ และจะดำเนินการในช่วงปิดภาคเรียน ดังนี้

1. ปิดภาคต้น เข้าค่ายที่ 1 จำนวน 2,660 คน ใช้เวลาประมาณ 20 วัน
2. ปิดภาคปลาย เข้าค่ายที่ 2 จำนวน 1,550 คน ใช้เวลาประมาณ 20 วัน

ศูนย์สอวน. ประกอบด้วยศูนย์โรงเรียนและศูนย์มหาวิทยาลัย แบ่งออกได้เป็น 2 ประเภท ดังนี้

1. ศูนย์ภูมิภาค มีทั้งหมด 15 ศูนย์ กระจายอยู่ตามส่วนภูมิภาคต่างๆ ทั่วประเทศ มีคณะวิทยาศาสตร์ มหาวิทยาลัยของรัฐ 13 แห่ง รับผิดชอบในการพัฒนานักเรียนและครูในจังหวัดต่าง ๆ ที่ได้รับมอบหมาย โดยมีโรงเรียนเครือข่ายอย่างน้อยจังหวัดละ 1 โรงเรียน ช่วยในการดำเนินการ มี 6 สาขาวิชา คือ คณิตศาสตร์ คอมพิวเตอร์ เคมี ชีววิทยา ฟิสิกส์ และดาราศาสตร์ รับนักเรียนเข้าอบรมค่าย 1 จำนวน 35 คน และคัดเลือกเข้าค่าย 2 จำนวน 20 คน
2. ศูนย์กทม. เป็นศูนย์วิชาทั้งหมด 6 ศูนย์ มีโรงเรียนเป็นศูนย์หลัก ประกอบด้วยโรงเรียนต่าง ๆ ในเครือข่าย โดยมีมหาวิทยาลัย 7 แห่ง เป็นศูนย์ที่เลี้ยงทางด้านวิชาการ แต่ละศูนย์วิชารับนักเรียนจำนวน 175 คน

ศ.ดร.สุรินทร์ พงศ์ศุภสมิทธิ ผู้อำนวยการสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) กล่าวว่า ผู้แทนประเทศไทยเข้าร่วมแข่งขันโอลิมปิกฯ ทั้ง 23 คน จะได้รับสิทธิพิเศษทุนเรียนต่อต่างประเทศระดับปริญญาตรี-โท-เอก ใน 3 สาขา คือ วิทยาศาสตร์ คณิตศาสตร์ และคอมพิวเตอร์ เพราะ 3 วิชานี้เป็นพื้นฐานทางแพทย์ เกสซ์ วิศววะ ซึ่งขณะนี้ไม่มีเด็กที่ได้รับทุนนี้เรียนต่อต่างประเทศ จะสำเร็จการศึกษากลับมารุ่นแรกในปี 2552 เด็กเหล่านี้เมื่อกลับประเทศไทยก็มาเป็นนักวิจัยชั้นนำระดับชาติ เพื่อให้พัฒนาองค์ความรู้ใหม่ ๆ สนองการพัฒนาประเทศ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ศ.ดร.ณรงค์ บัณนิม<sup>2</sup> อาจารย์สาขาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประธานคณะกรรมการผู้ดูแลวิชาคณิตศาสตร์ มูลนิธิส่งเสริมโอลิมปิกวิชาการและพัฒนามาตรฐานวิทยาศาสตร์ศึกษาในพระอุปถัมภ์สมเด็จพระเจ้าพี่นางเธอเจ้าฟ้ากัลยาณิวัฒนา กรมหลวงนราธิวาสราชนครินทร์ (สอวน.) เปิดเผยข้อมูลจากสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) ว่าเด็กไทยได้คะแนนวิชาคณิตศาสตร์และวิทยาศาสตร์ต่ำกว่าคะแนนเฉลี่ยนานาชาติ ทั้งนี้ผลสัมฤทธิ์ทางการเรียนทั้ง 2 วิชา ยังต่ำกว่าช่วงปี 2542 – 2546 ว่าเป็นเรื่องจริง และเป็นปัญหาที่หมักหมมอยู่ในระบบการเรียนการสอนในสังคมไทยมานานมาก สืบเนื่องมาจากครูไทยใช้ระบบข้อสอบปรนัยมาวัดผลเด็ก นอกจากการวัดผลแล้วการฝึกเด็กก็ยังใช้ข้อสอบปรนัย ซึ่งเป็นข้อสอบที่ไม่ได้พัฒนาระบบคิด การใช้เหตุผล ข้อสอบปรนัยเป็นข้อสอบที่ทำให้ลายเด็กไทยอย่างมาก คนที่มาเรียนครูก็ถูกมอมเมาด้วยข้อสอบปรนัย มาเป็นครูก็ใช้ข้อสอบปรนัย เด็กบางคนทำข้อสอบโดยไม่ต้องอ่านคำถามใช้วิธีสุ่มเดาก็สามารถสอบผ่านได้

### คณิตศาสตร์โอลิมปิกสอวน.

#### (POSN Mathematical Olympiad POSN-MO)

ในหัวข้อนี้กล่าวถึงข้อมูลเกี่ยวกับคณิตศาสตร์โอลิมปิกสอวน. โดยเฉพาะอย่างยิ่งที่ศูนย์สาขาวิชาคณิตศาสตร์ ศูนย์โรงเรียนสวนกุหลาบวิทยาลัย - ศูนย์มหาวิทยาลัย มหาวิทยาลัยรามคำแหง มหาวิทยาลัยศรีนครินทรวิโรฒ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังและมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

เนื้อหาคณิตศาสตร์ที่ทำการเรียนการสอนจะประกอบด้วย 7 สาขาหลัก ดังนี้

1. ตรรกศาสตร์และการพิสูจน์ (Logic and proof)
2. เรขาคณิต (Geometry)
3. พีชคณิต (Algebra)
4. ทฤษฎีจำนวน (Number Theory)
5. คอมบินาทอริก (Combinatorics)
6. อสมการ (Inequalities)
7. สมการเชิงฟังก์ชัน (Functional Equations)

<sup>2</sup> ข้อมูลจาก <http://www.manager.co.th/>

การจัดสอบแข่งขันแบ่งเป็น 2 ภาค คะแนนภาคปริญญ์ต่อคะแนนภาคอัตรันัยเป็น 1:2

- ภาคปริญญ์ ข้อสอบครอบคลุมทุกเรื่อง ใช้เวลา 3 ชั่วโมง นักเรียนต้องใช้ความสามารถในการคิดแก้ปัญหาอย่างรวดเร็ว
- ภาคอัตรันัย ข้อสอบจะเป็นแบบบูรณาการครอบคลุมทุกเรื่อง ใช้เวลาสอบ 3 ชั่วโมง นักเรียนต้องใช้ความสามารถในการแก้ปัญหาอย่างเป็นระบบ

นักเรียนที่มีคะแนนลำดับต้นจะได้รับการคัดเลือกไปแข่งขัน ในโครงการจัดส่งผู้แทนประเทศไทยไปแข่งขันโอลิมปิกวิชาการระหว่างประเทศที่สสวท. ต่อไป

เพชรลดดา สีหะวงศ์ [3] ได้ให้ข้อเสนอแนะในการนำผลการวิจัยไปใช้ที่สำคัญดังต่อไปนี้

1. จากผลการพัฒนาตัวบ่งชี้คุณลักษณะเด็กที่มีความสามารถพิเศษทางคณิตศาสตร์ ครูและผู้ที่เกี่ยวข้องกับเด็กที่มีความสามารถพิเศษทางคณิตศาสตร์สามารถนำองค์ประกอบและตัวบ่งชี้ นำไปใช้เป็นแบบในการพัฒนาคุณลักษณะเด็กที่มีความสามารถพิเศษทางคณิตศาสตร์ให้สามารถครอบคลุมทุกด้านต่อไป โดยเฉพาะอย่างยิ่งองค์ประกอบด้านการแก้ปัญหาอย่างสร้างสรรค์ทางคณิตศาสตร์ ซึ่งเป็นความสามารถในการนำความรู้หรือประสบการณ์ที่มีอยู่ทางคณิตศาสตร์ เพื่อมาประยุกต์ใช้กับสถานการณ์ซึ่งนำไปสู่วิธีการแก้ปัญหาที่แปลกใหม่ ริเริ่ม และมองเห็นผลผลิตในรูปแบบใหม่ในการหาคำตอบของปัญหาที่ซับซ้อนทางคณิตศาสตร์ โดยเป็นทักษะที่ครูสามารถนำไปพัฒนาในชั้นเรียนได้
2. จากผลการวิจัย พบว่า องค์ประกอบคุณลักษณะเด็กที่มีความสามารถพิเศษทางคณิตศาสตร์ ด้านเชาวน์ปัญญา ตัวบ่งชี้ความสามารถในการย้อนคิดมีน้ำหนักความสำคัญมากที่สุด ส่วนด้านการแก้ปัญหาอย่างสร้างสรรค์ทางคณิตศาสตร์ ตัวบ่งชี้ความสามารถด้านจำนวน มีน้ำหนักความสำคัญมากที่สุด และด้านความถนัดทางคณิตศาสตร์ ตัวบ่งชี้สมรรถภาพด้านจำนวน มีน้ำหนักความสำคัญมากที่สุด ซึ่งทางโครงการต่าง ๆ ที่ส่งเสริมการจัดการศึกษาสำหรับเด็กที่มีความสามารถพิเศษทางคณิตศาสตร์ สามารถนำตัวบ่งชี้ในแต่ละองค์ประกอบที่ได้จากผลการวิจัย ในครั้งนี้ไปใช้ใน การพัฒนา ส่งเสริมและวัดประเมินผลคุณภาพคุณลักษณะเด็กที่มีความสามารถพิเศษทางคณิตศาสตร์ได้ และเพื่อปรับปรุงแก้ไขทำให้การเสาะแสวงหาเด็กที่มีความสามารถพิเศษทาง คณิตศาสตร์มีประสิทธิภาพมากยิ่งขึ้น ข้อเสนอแนะในการวิจัยครั้งต่อไป

ผู้วิจัยได้เก็บรวบรวมข้อมูลดังกล่าวจากนักเรียนจากโครงการ สอวน. สาขาคณิตศาสตร์ ศูนย์โรงเรียนสวนกุหลาบวิทยาลัย การวิจัยครั้งต่อไปควรมีการรวบรวมข้อมูลจากศูนย์อื่น ๆ อีกทั่ว เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์เพื่อการวิจัยเพื่อการศึกษาเท่านั้น เมื่อนำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประเทศด้วย เพื่อตรวจสอบความไม่แปรเปลี่ยนของโมเดลที่ได้จากการวิจัย รวมทั้งนักเรียนในระดับชั้นอื่น ๆ หรือโครงการอื่น ๆ ที่จัดขึ้นส่งเสริมการจัดการศึกษาสำหรับเด็กที่มีความสามารถพิเศษ แล้วนำมาเปรียบเทียบกัน เพื่อศึกษาร่วมมีตัวแปรใดที่แตกต่างกันบ้าง เพื่อเป็นประโยชน์ในการนำผลการวิจัยไปใช้ในการพัฒนาตัวบ่งชี้คุณลักษณะเด็ก ที่มีความสามารถพิเศษทางคณิตศาสตร์ต่อไป

3. ผู้วิจัยได้ยึดแนวทางการวิจัยจากการวิเคราะห์สังเคราะห์ องค์ประกอบและตัวบ่งชี้จากเอกสาร และงานวิจัยที่ได้รับการตรวจสอบตัวบ่งชี้จากผู้เชี่ยวชาญ คือ องค์ประกอบด้านเชาวน์ปัญญา องค์ประกอบด้านการแก้ปัญหาอย่างสร้างสรรค์ทางคณิตศาสตร์ และองค์ประกอบด้านความถนัดทางคณิตศาสตร์ ซึ่งโดยรวมแล้วเป็นเพียงองค์ประกอบของคุณลักษณะด้านสติปัญญาของเด็กที่มีความสามารถพิเศษเท่านั้น โดยยังไม่ครอบคลุมคุณลักษณะด้านอารมณ์ของเด็กที่มีความสามารถพิเศษทางคณิตศาสตร์ ในการวิจัยครั้งต่อไปจึงควรศึกษาแนวทางการวิจัยให้ครอบคลุมทั้งคุณลักษณะด้าน สติปัญญาและอารมณ์ เพื่อความสมบูรณ์ของตัวบ่งชี้

ซึ่งการวิจัยที่ผ่านมาของเด็กที่มีความสามารถพิเศษทางคณิตศาสตร์อาจยังไม่ครอบคลุมถึงแนวคิดและทัศนคติต่าง ๆ จากกลุ่มเด็กเองและข้อมูลเกี่ยวกับเด็กอื่น ๆ เช่น วิชาคณิตศาสตร์ที่เด็กคิดว่าชอบและเข้าใจได้ดี หรือระยะเวลาที่เด็กใช้เตรียมตัวอ่านหนังสือหรือทำโจทย์ทบทวน ฯลฯ ซึ่งเป็นข้อมูลด้านต่าง ๆ ที่จะได้จากนักเรียนเอง ในการวิจัยครั้งนี้ผู้วิจัยจึงได้รวบรวมข้อมูลในลักษณะดังกล่าวโดยประชากรที่ใช้ในการวิจัยครั้งนี้ คือ นักเรียนโครงการส่งเสริมโอลิมปิกวิชาการและพัฒนามาตรฐานวิทยาศาสตร์ คณิตศาสตร์ศึกษา (สอวน.) สาขาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 - 4 ทั่วประเทศ กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือ นักเรียนในโครงการส่งเสริมโอลิมปิกวิชาการและพัฒนามาตรฐานวิทยาศาสตร์ คณิตศาสตร์ศึกษา (สอวน.) สาขาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 - 4 ในเขตกรุงเทพมหานคร ซึ่งได้รับการอบรมเข้าค่ายที่ศูนย์โรงเรียนสวนกุหลาบวิทยาลัย สาขาคณิตศาสตร์ รุ่นที่ 7 เท่านั้น

รายงานวิจัยฉบับนี้จะประกอบด้วยสองส่วนหลักนั่นคือ ส่วนที่ 1 จะประกอบด้วยผลจากการทำแบบสำรวจความคิดเห็นนักเรียนจำนวน 75 คน ด้วยแบบสอบถาม และผลทางสถิติที่ได้ ส่วนในส่วนที่ 2 จะกล่าวถึงผลการวิจัยทางคณิตศาสตร์ที่ได้ค้นพบทฤษฎีทางคณิตศาสตร์ใหม่ ๆ รวมถึงใช้คุณสมบัติทางคณิตศาสตร์ดังกล่าวนั้นกับข้อมูลคะแนนของนักเรียนเป็นกรณีศึกษาได้อย่างถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# ส่วนที่ 1

## Part I

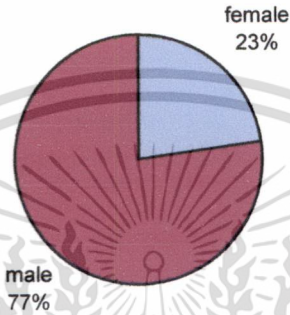


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# ผลการวิจัยส่วนที่ 1

ส่วนที่ 1 ประกอบด้วยผลจากการทำแบบสำรวจความคิดเห็นนักเรียนจำนวน 75 คน ด้วยแบบสอบถาม และผลลัพธ์ทางสถิติที่ได้ แสดงได้ดังต่อไปนี้

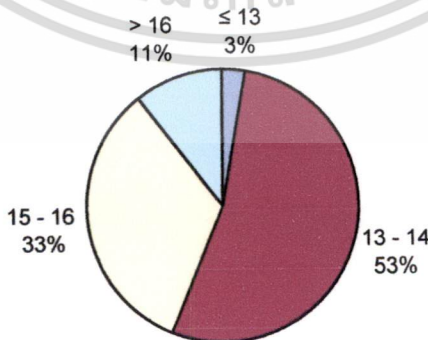
Gender



รูปที่ 1.1 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามเพศ

จากรูปที่ 1.1 พบว่า นักเรียนเป็นเพศชายรวม 58 คน คิดเป็น 77% และเป็นเพศหญิงรวม 17 คน คิดเป็น 23 % ซึ่งมีความแตกต่างกันค่อนข้างมาก จึงควรเพิ่มการรณรงค์และสนับสนุนให้นักเรียนเพศหญิงหันมาสนใจในวิชาคณิตศาสตร์มากขึ้น เพื่อเพิ่มอัตราส่วนของผู้มีความสามารถพิเศษทางคณิตศาสตร์ในเพศหญิงให้มากขึ้นได้อีกทางหนึ่ง

Age

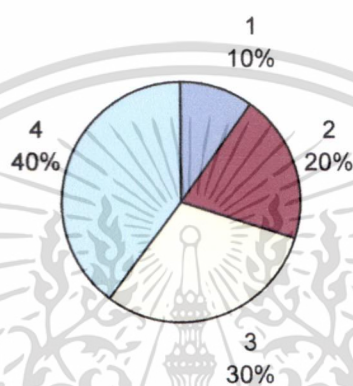


รูปที่ 1.2 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามอายุ

เอกสารนี้เป็นเอกสารสงวนลิขสิทธิ์ไว้เพื่อใช้ในการศึกษาเท่านั้น เมื่อผู้ใดเห็นประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 1.2 จากนักเรียนที่ทำแบบสอบถามทั้งหมด มีนักเรียนอายุน้อยกว่า 13 ปี 2 คน คิดเป็น 3% นักเรียนอายุระหว่าง 13 – 14 ปี รวม 40 คน คิดเป็น 53 % นักเรียนอายุระหว่าง 15 – 16 ปี รวม 25 คน คิดเป็น 33 % และนักเรียนอายุมากกว่า 16 ปี 8 คน คิดเป็น 11% ซึ่งการมีนักเรียนอายุต่ำกว่า 15 ปี ถึง 56% บ่งบอกถึงความกระตือรือร้นของนักเรียน ในการเรียนวิชาคณิตศาสตร์ ตั้งแต่อายุน้อย ซึ่งถือเป็นผลดีของวงการการศึกษา

Level



รูปที่ 1.3 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามระดับ

จากรูปที่ 1.3 พบว่านักเรียนศึกษาอยู่ระดับมัธยมศึกษาปีที่ 1 จำนวน 11 คน คิดเป็น 10% นักเรียนศึกษาอยู่ระดับมัธยมศึกษาปีที่ 2 จำนวน 26 คน คิดเป็น 20 % นักเรียนศึกษาอยู่ระดับมัธยมศึกษาปีที่ 3 จำนวน 13 คน คิดเป็น 30 % และนักเรียนศึกษาอยู่ระดับมัธยมศึกษาปีที่ 4 จำนวน 25 คน คิดเป็น 40% ซึ่งเห็นได้ชัดว่านักเรียนให้ความสนใจเพิ่มขึ้นตามลำดับ ทั้งในระดับมัธยมต้นและมัธยมปลาย ทั้งนี้ อาจรวมถึงนักเรียนเข้ามาเรียนที่ศูนย์อย่างต่อเนื่องตั้งแต่ระดับมัธยมต้น

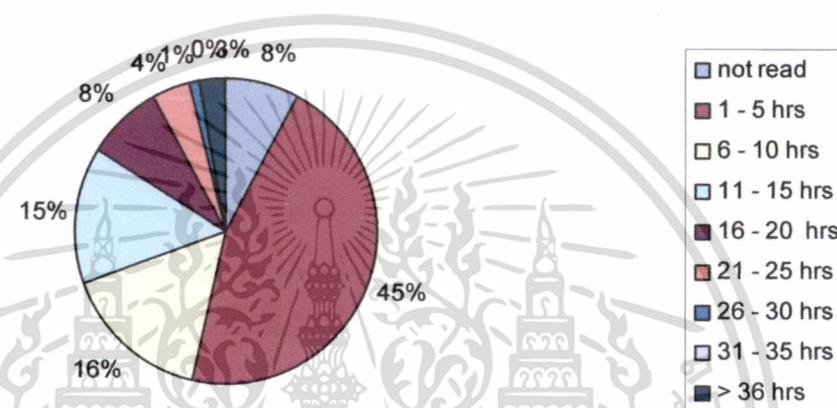
ในรูปที่ 1.4 นั้น ในแต่ละสัปดาห์ โดยเฉลี่ยแล้วนักเรียนไม่อ่านหนังสือทบทวนระหว่างเทอมเลย 6 คน คิดเป็น 8% นักเรียนอ่านหนังสือทบทวนระหว่างเทอม 1 – 5 ชั่วโมง รวม 34 คน คิดเป็น 45 % นักเรียนอ่านหนังสือทบทวนระหว่างเทอม 6 - 10 ชั่วโมง รวม 12 คน คิดเป็น 16 % นักเรียนอ่านหนังสือทบทวนระหว่างเทอม 11 - 15 ชั่วโมง รวม 11 คน คิดเป็น 15 % นักเรียนอ่านหนังสือทบทวนระหว่างเทอม 16 - 20 ชั่วโมง รวม 6 คน คิดเป็น 8 % นักเรียนอ่านหนังสือทบทวนระหว่างเทอม 21 - 25 ชั่วโมง รวม 3 คน คิดเป็น 4 % นักเรียนอ่านหนังสือทบทวนระหว่างเทอม 26 - 30 ชั่วโมง รวม 1 คน คิดเป็น 1 % นักเรียนอ่านหนังสือทบทวนระหว่างเทอม 31 - 35 ชั่วโมง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้า เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รวม 0 คน คิดเป็น 3 % และนักเรียนอ่านหนังสือทบทวนระหว่างเทอมมากกว่า 36 ชั่วโมง รวม 2 คน คิดเป็น 8 %

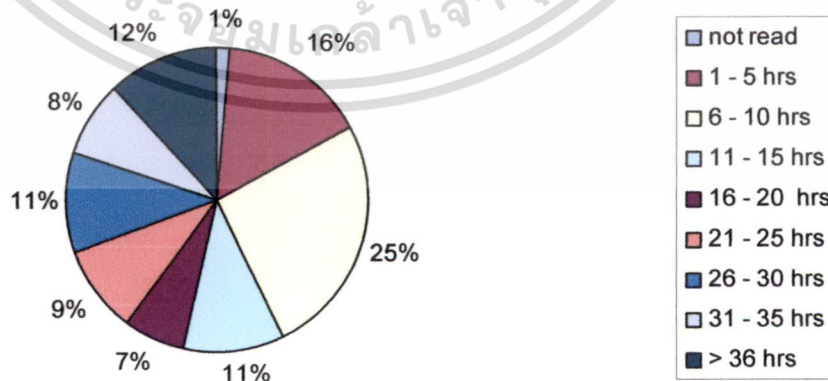
แสดงให้เห็นว่านักเรียนอ่านหนังสือเพื่อทบทวนระหว่างเทอมโดยเฉลี่ยมากกว่า 5 ชั่วโมงต่อวัน 8% แต่ส่วนใหญ่แล้วนักเรียนผู้มีความสามารถพิเศษทางด้านคณิตศาสตร์นี้ กลับอ่านหนังสือทบทวนโดยเฉลี่ยไม่ถึง 1 ชั่วโมงต่อวันมากถึง 45% หรือเกือบกึ่งหนึ่งของทั้งหมด ทั้งนี้อาจเนื่องมาจากนักเรียนได้ใช้เวลาส่วนหนึ่งในการติววิชาต่าง ๆ เพิ่มไปแล้ว

Reading hours for regular study



รูปที่ 1.4 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามการอ่านหนังสือทบทวนระหว่างเทอมโดยเฉลี่ยในแต่ละสัปดาห์

Reading hours for test

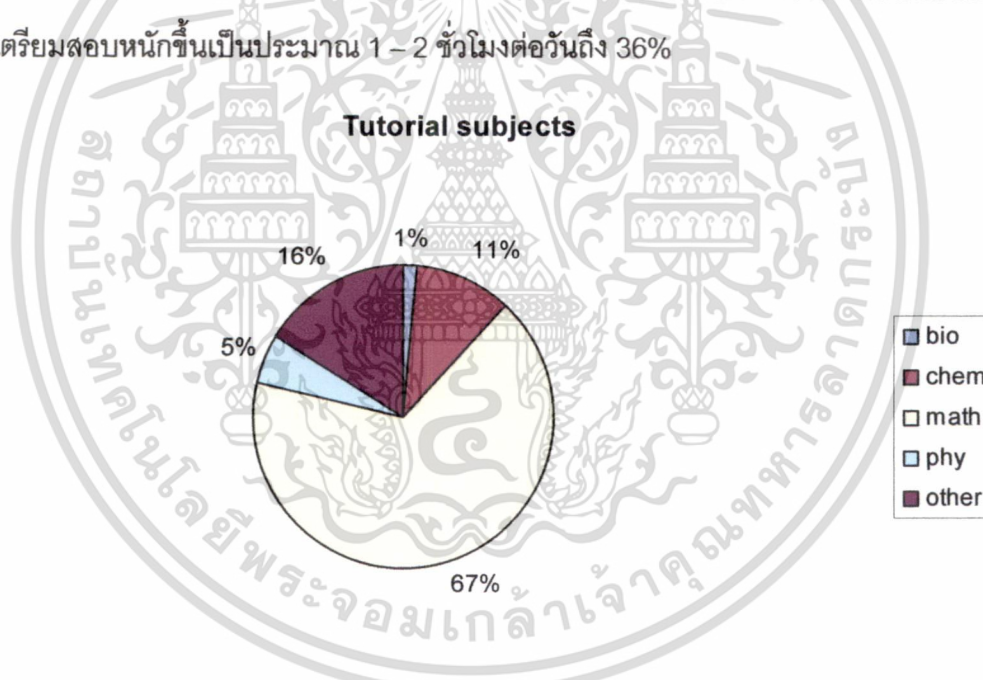


รูปที่ 1.5 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามการอ่านหนังสือเพื่อเตรียมสอบโดย

เฉลี่ยในแต่ละสัปดาห์ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 1.5 นั้นในแต่ละสัปดาห์โดยเฉลี่ยแล้วนักเรียนไม่อ่านหนังสือเตรียมสอบเลย 1 คน คิดเป็น 1% นักเรียนอ่านหนังสือเตรียมสอบ 1 – 5 ชั่วโมง รวม 12 คน คิดเป็น 16 % นักเรียนอ่านหนังสือเตรียมสอบ 6 - 10 ชั่วโมง รวม 19 คน คิดเป็น 25 % นักเรียนอ่านหนังสือเตรียมสอบ 11 - 15 ชั่วโมง รวม 8 คน คิดเป็น 11 % นักเรียนอ่านหนังสือเตรียมสอบ 16 - 20 ชั่วโมง รวม 5 คน คิดเป็น 7 % นักเรียนอ่านหนังสือเตรียมสอบ 21 - 25 ชั่วโมง รวม 7 คน คิดเป็น 9 % นักเรียนอ่านหนังสือเตรียมสอบ 26 - 30 ชั่วโมง รวม 8 คน คิดเป็น 11 % นักเรียนอ่านหนังสือเตรียมสอบ 31 - 35 ชั่วโมง รวม 6 คน คิดเป็น 8 % และนักเรียนอ่านหนังสือเตรียมสอบมากกว่า 36 ชั่วโมง รวม 9 คน คิดเป็น 12 %

แสดงให้เห็นว่า พฤติกรรมการอ่านหนังสือเพื่อเตรียมสอบในกลุ่มตัวอย่างนั้นแตกต่างกับการอ่านหนังสือทบทวนระหว่างเทอมอย่างเห็นได้ชัด โดยนักเรียนอ่านหนังสือเตรียมสอบโดยเฉลี่ยมากกว่า 5 ชั่วโมงต่อวัน เพิ่มขึ้นเป็น 12% และจากส่วนใหญ่ที่นักเรียนผู้มีความสามารถพิเศษทางด้านคณิตศาสตร์นี้อ่านหนังสือทบทวนโดยเฉลี่ยไม่ถึง 1 ชั่วโมงต่อวันมากถึง 45% ก็เพิ่มเป็นอ่านหนังสือเตรียมสอบหนักขึ้นเป็นประมาณ 1 – 2 ชั่วโมงต่อวันถึง 36%



รูปที่ 1.6 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามวิชาที่ติวเพิ่ม

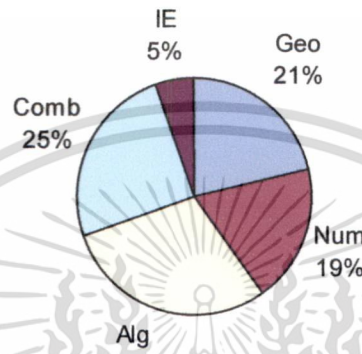
จากรูปที่ 1.6 สรุปได้ว่านักเรียนติววิชาชีววิทยาเพิ่ม 1 คน คิดเป็น 1% นักเรียนติววิชาเคมีเพิ่ม 8 คน คิดเป็น 11% นักเรียนติววิชาคณิตศาสตร์เพิ่ม 50 คน คิดเป็น 67% นักเรียนติววิชาฟิสิกส์เพิ่ม 4 คน คิดเป็น 5% และนักเรียนติววิชาอื่น ๆ เพิ่ม 12 คน คิดเป็น 16% ซึ่งแสดงให้เห็นว่านักเรียนผู้มีความสามารถพิเศษด้านคณิตศาสตร์ในกลุ่มตัวอย่างนี้ ติววิชาคณิตศาสตร์เพิ่มถึง 67%

ถัดไป จะพิจารณาเฉพาะวิชาคณิตศาสตร์ โดยมองจำแนกไปที่สาขาย่อยต่าง ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

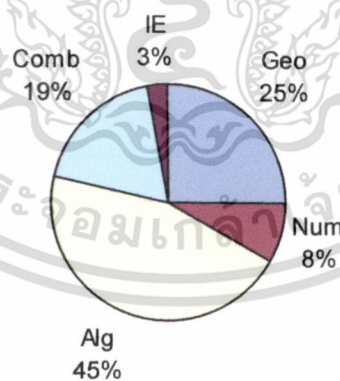
จากวิชาคณิตศาสตร์สาขาต่าง ๆ<sup>3</sup> ในรูปที่ 1.7 นักเรียนชอบเรียนวิชาเรขาคณิตมากที่สุด 16 คน คิดเป็น 21% นักเรียนชอบเรียนวิชาทฤษฎีจำนวนมากที่สุด 14 คน คิดเป็น 19% นักเรียนชอบเรียนวิชาพีชคณิตมากที่สุด 22 คน คิดเป็น 30% นักเรียนชอบเรียนวิชาคอมบินาทอริกมากที่สุด 19 คน คิดเป็น 25% และนักเรียนชอบเรียนวิชาอสมการมากที่สุด 4 คน คิดเป็น 5%

**Favourite subjects**



รูปที่ 1.7 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามสาขาวิชาคณิตศาสตร์ที่ชอบ

**Most understand subjects**

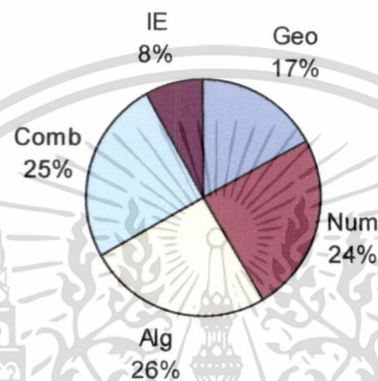


รูปที่ 1.8 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามสาขาวิชาคณิตศาสตร์ที่คิดว่าเข้าใจ

<sup>3</sup> เอกสารนี้เป็นเอกสารของบริษัทเอกชนที่ขอสงวนลิขสิทธิ์ไว้และขอสงวนสิทธิ์ในข้อมูลที่ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 1.8 จากวิชาคณิตศาสตร์สาขาต่าง ๆ นักเรียนคิดว่าเข้าใจเนื้อหาวิชาเรขาคณิตมากที่สุด 19 คน คิดเป็น 25% นักเรียนคิดว่าเข้าใจเนื้อหาวิชาทฤษฎีจำนวนมากที่สุด 6 คน คิดเป็น 8% นักเรียนคิดว่าเข้าใจเนื้อหาวิชาพีชคณิตมากที่สุด 34 คน คิดเป็น 45% นักเรียนคิดว่าเข้าใจเนื้อหาวิชาคอมบินาทอริกมากที่สุด 14 คน คิดเป็น 19% และนักเรียนคิดว่าเข้าใจเนื้อหาวิชาอสมการมากที่สุด 2 คน คิดเป็น 3%

### Most practiced subjects



รูปที่ 1.9 แผนภูมิวงกลมแสดงจำนวนของนักเรียนจำแนกตามสาขาวิชาคณิตศาสตร์ที่ฝึกทำโจทย์

จากรูปที่ 1.9 จากวิชาคณิตศาสตร์สาขาต่าง ๆ นักเรียนฝึกทำโจทย์วิชาเรขาคณิตมากที่สุด 13 คน คิดเป็น 17% นักเรียนฝึกทำโจทย์วิชาทฤษฎีจำนวนมากที่สุด 18 คน คิดเป็น 24% นักเรียนฝึกทำโจทย์วิชาพีชคณิตมากที่สุด 19 คน คิดเป็น 26% นักเรียนฝึกทำโจทย์วิชาคอมบินาทอริกมากที่สุด 19 คน คิดเป็น 25% และนักเรียนฝึกทำโจทย์วิชาอสมการมากที่สุด 6 คน คิดเป็น 8%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บรรณานุกรม

1. บุญทัน อยู่ชมบุญ, พฤติกรรมการเรียนการสอนคณิตศาสตร์: ระดับประถมศึกษา, โอเดียนสโตร์, 2529
2. คักดา บุญโต และคณะ, คณิตศาสตร์มหัศจรรย์: คู่มือการจัดการศึกษาสำหรับผู้มีความสามารถพิเศษด้านคณิตศาสตร์, กรุงเทพฯ: ศูนย์แห่งชาติเพื่อพัฒนาผู้มีความสามารถพิเศษ สำนักงานคณะกรรมการการศึกษาแห่งชาติ, 2544
3. เพชรลดา สีหะวงศ์, The Development of Indicators for the Characteristics of the Talented Children in Mathematics, วิทยานิพนธ์, จุฬาลงกรณ์มหาวิทยาลัย, 2550
4. <http://www.posn.or.th/>, 2009
5. <http://www.manager.co.th>, 2009



# ส่วนที่ 2

## Part II



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 1 Rough sets, flow graphs and their extension to association rules

In Part II, the mathematical discovery of rough sets are provided. Several case studies on the POSN data sets are also reported.

Rough set theory was introduced by Zdzislaw Pawlak in the early 1980's [1]. The crux of rough set theory is the novel use of approximations to cope with uncertainty. A significant extension of rough set theory is called *flow graphs*, such that the union of all inputs  $x$  of  $y$  are interpreted as the *upper approximation* of  $y$  and the union of all inputs  $x$  of  $y$  with  $cer(x, y) = 1$  are interpreted as the *lower approximation* of  $y$  [2]. Pawlak's flow graphs are different to *flow networks* introduced by Ford et al. that do not analyze optimal flow [3].

Flow graphs represent the model of information flow from the given data set. The branches of a flow graph can be constructed as decision rules, while the entire flow graph describes a decision algorithm [4]. With every decision rule, there are three associated coefficients: *strength*, *certainty* and *coverage*. These coefficients satisfy Bayes' theorem. We are also able to discover dependency and correlation within a data set without reference to its probabilistic nature by using flow graphs. Preference analysis [2], conflict analysis [5], rule analysis [6], propositional calculus [7, 8], data mining [9, 10] and granular computing [11] were analyzed by flow graphs efficiently. In [11, 12], the authors attached objects to the flow graphs' nodes. Hence, more precise relationships among nodes were considered. We can construct a decision tree as a unique flow graph by removing the root while its nodes are labeled by the same attribute. The author concluded that such a flow graph provides better structure of data compared to a decision tree [10]. In [11, 12], the authors proposed the extended flow graphs and algorithms to handle redundancy and consistency. The mathematical model of flow graph in the decision algorithm can build an efficient search engines for the CDDB database [13, 14].

Thus, flow graphs are a useful approach to discover hidden knowledge in the data. For these reasons, we propose data analysis based on Pawlak's flow graphs and extend them to handle association rule extraction.

### 1.1 Rough Set Theory

The following rough sets preliminary is taken from the study of Pawlak [1]. This mathematical theory formed the basis of flow graphs later in the early 2000s [2, 4, 8–10, 15].

Rough sets are based on an *information system* (*data table*). More formally, it is a pair  $S = (U, A)$ , where  $U$  is a nonempty finite set of objects called the *universe* and  $A$  is a nonempty finite set of attributes such that  $a: U \rightarrow V_a$  for every  $a \in A$ . The set  $V_a$  is called the *domain* of  $a$ . Any subset  $B$  of  $A$  determines

a binary relation  $I(B)$  on  $U$  called an *indiscernibility relation*. It is defined as  $(x, y) \in I(B)$  if and only if  $a(x) = a(y)$  for every  $a \in A$ , where  $a(x)$  denotes the attribute value of element  $x$ . If  $(x, y)$  belongs to  $I(B)$ ,  $x$  and  $y$  are called *B-indiscernible*. Equivalence classes of the relation  $I(B)$  are referred to as *B-elementary sets* (or *B-granules*, for more *granular computing* studies please refer to [2]).

If we distinguish an information system to two disjoint classes of attributes, called *condition* and *decision attributes*, then the information system will be called a *decision system*, denoted by  $S = (U, C, D)$ , where  $C$  and  $D$  are disjoint sets of condition and decision attributes, respectively.

Suppose we are given an information system  $S = (U, A)$ ,  $X \subseteq U$ , and  $B \subseteq A$ . We can approximate  $X$  using the information in  $B$  by constructing the *B-lower* and *B-upper approximations* of  $X$ , denoted by  $B_*(X)$  and  $B^*(X)$  respectively, and defined as follows:

$$B_*(X) = \bigcup_{x \in U} \{B(x) \mid B(x) \subseteq X\} \quad (1)$$

$$B^*(X) = \bigcup_{x \in U} \{B(x) \mid B(x) \cap X \neq \emptyset\}. \quad (2)$$

In addition, the set

$$BN_B(X) = B^*(X) - B_*(X) \quad (3)$$

will be referred to as *B-boundary region* of  $X$ . If the boundary region of  $X$  is an empty set ( $BN_B(X) = \emptyset$ ), then  $X$  is *crisp*. In contrast, if  $BN_B(X) \neq \emptyset$ , then  $X$  is *rough*. More studies involving rough sets are in [2, 4, 5, 7–10, 15, 16].

## 1.2 Flow Graphs

In this section we present basic definitions and some properties of flow graphs from the studies of Pawlak [2, 4, 8–10, 15, 16]. A flow graph is an alternative representation of the knowledge, more specifically the information flow, in the given data.

### Flow Graphs

A *flow graph* is a *directed, acyclic, finite graph*  $G = (N, B, \varphi)$ , where  $N$  is a set of nodes,  $B: N \times N$  is a set of *directed branches*,  $\varphi: B \rightarrow R^+$  is a *flow function* and  $R^+$  is the set of non-negative real numbers. If  $(x, y) \in B$  then  $x$  is an *input* of node  $y$  denoted by  $I(y)$  and  $y$  is an *output* of node  $x$  denoted by  $O(x)$ .

**Example 1a** (adapted from [9]) Suppose Fig. 1 depicts the flow graph of voting initial data. Nodes of the flow graph are represented by circles, labeled by  $x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2$  and  $z_3$ . A branch  $(x, y)$  is denoted by an arrow from nodes  $x$  to  $y$ . Thus, inputs of node  $y_1$  are nodes  $x_1$  and  $x_2$ , and outputs of node  $x_1$  are  $y_1$  and  $y_2$ , respectively.

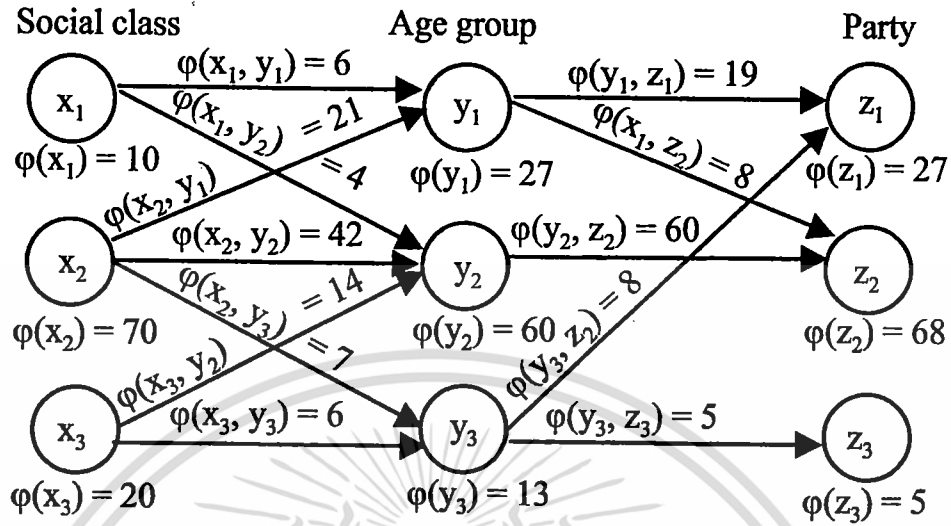


Fig. 1. Flow graph initial data of voting problem (adapted from [9]).

Next, *input* and *output* of a flow graph  $G$  are defined by  $I(G) = \{x \in N \mid I(x) = \emptyset\}$  and  $O(G) = \{x \in N \mid O(x) = \emptyset\}$ , respectively. These inputs and outputs of  $G$  are called *external nodes* of  $G$  whereas other nodes are called *internal nodes* of  $G$ . If  $(x, y) \in B$  then we call  $(x, y)$  a *throughflow* from  $x$  to  $y$ . We will assume in what follows that  $\varphi(x, y) \neq 0$  for every  $(x, y) \in B$ .

**Example 1b** In Fig. 1, inputs and outputs of this flow graph are  $x_1, x_2, x_3$  and  $z_1, z_2, z_3$  respectively. Meanwhile,  $x_1, x_2, x_3$  and  $z_1, z_2, z_3$  are external nodes and  $y_1, y_2, y_3$  are internal nodes of the flow graph  $G$ , respectively. The branch  $(x_1, y_1)$  has its throughflow  $\varphi(x_1, y_1) = 6$ .

With every node  $x$  of a flow graph  $G$ , we have its associated *inflow*:

$$\varphi_+(x) = \sum_{y \in I(x)} \varphi(x, y) \quad (4)$$

and *outflow*:

$$\varphi_-(x) = \sum_{y \in O(x)} \varphi(x, y). \quad (5)$$

Similarly, an *inflow* and an *outflow* for the flow graph  $G$  are defined as:

$$\varphi_+(G) = \sum_{y \in I(G)} \varphi(x, y) \quad (6)$$

and

$$\varphi_-(G) = \sum_{y \in O(G)} \varphi(x, y). \quad (7)$$

We assume that for any internal node  $x$ ,  $\varphi_-(x) = \varphi_+(x) = \varphi(x)$ , where  $\varphi(x)$  is a *throughflow* of node  $x$ . Similarly then,  $\varphi_-(G) = \varphi_+(G) = \varphi(G)$  is a *throughflow* of graph  $G$ . As discussed by Pawlak [2], the above equations can be considered as *flow conservation equations* [3].

**Example 1c** In Fig. 1, inflow of node  $y_1$  is  $\varphi_+(y_1) = 6 + 21 = 27$ , and outflow of node  $y_1$  is  $\varphi_-(y_1) = 19 + 8 = 27$ . Inflow of the flow graph  $G$  is  $\varphi_+(G) = 100$ , and outflow of the flow graph  $G$  is  $\varphi_-(G) = 100$ .

### Normalized Flow Graphs

In order to demonstrate interesting relationships between flow graphs and other disciplines, we next come to the normalized version of flow graphs (consider  $\sigma$  in  $[0, 1]$  instead of  $R^+$ ). A *normalized flow graph* is a *directed, acyclic, finite graph*  $G = (N, B, \sigma)$ , where  $N$  is a set of *nodes*,  $B \subseteq N \times N$  is a set of *directed branches* and  $\sigma: B \rightarrow [0, 1]$  is a *normalized flow* of  $(x, y)$ . The *strength* of  $(x, y)$  is given by

$$\sigma(x, y) = \frac{\varphi(x, y)}{\varphi(G)} \quad (8)$$

where  $0 \leq \sigma(x, y) \leq 1$ .

The *strength\_of\_branch*  $\times 100$  can be understood as the percentage of a total flow through the branch. For simplicity, from now we will use normalized flow graphs only.

With every node  $x$  of a flow graph  $G$ , associated *normalized inflow* and *outflow* are defined respectively as:

$$\varphi_+(x) = \frac{\varphi_+(x)}{\varphi(G)} = \sum_{y \in I(x)} \sigma(y, x) \quad (9)$$

and

$$\varphi_-(x) = \frac{\varphi_-(x)}{\varphi(G)} = \sum_{y \in O(x)} \sigma(x, y). \quad (10)$$

For any internal node  $x$ , it holds that  $\sigma_+(x) = \sigma_-(x) = \sigma(x)$ , where  $\sigma(x)$  is a *normalized throughflow* of  $x$ . Similarly, *normalized inflow* and *outflow* for the flow graph  $G$  are defined as:

$$\varphi_+(G) = \frac{\varphi_+(G)}{\varphi(G)} = \sum_{x \in I(G)} \sigma_-(x) \quad (11)$$

and

$$\varphi_-(G) = \frac{\varphi_-(G)}{\varphi(G)} = \sum_{x \in O(G)} \sigma_+(x). \quad (12)$$

It also holds that  $\sigma_+(G) = \sigma_-(G) = \sigma(G) = 1$ .

With every branch  $(x, y)$  of a flow graph  $G$ , the *certainty* and the *coverage* of  $(x, y)$  are defined respectively as:

$$cer(x, y) = \frac{\sigma(x, y)}{\sigma(x)} \quad (13)$$

and

$$cov(x, y) = \frac{\sigma(x, y)}{\sigma(y)}, \quad (14)$$

where  $\sigma(x) \neq 0$  and  $\sigma(y) \neq 0$ . In accordance with the previous definitions, the following are some consequence properties:

$$\sum_{y \in O(x)} cer(x, y) = 1, \quad (15)$$

$$\sum_{y \in I(x)} cov(x, y) = 1, \quad (16)$$

$$\sigma(x) = \sum_{y \in O(x)} cer(x, y)\sigma(x) = \sum_{y \in O(x)} \sigma(x, y), \quad (17)$$

$$\sigma(x) = \sum_{x \in I(x)} cov(x, y)\sigma(y) = \sum_{x \in I(x)} \sigma(x, y), \quad (18)$$

$$cer(x, y) = \frac{cov(x, y)\sigma(y)}{\sigma(x)}, \quad (19)$$

$$cov(x, y) = \frac{cer(x, y)\sigma(x)}{\sigma(y)}. \quad (20)$$

Equations (17)-(20) have some probabilistic relations. (17)-(18) have the form of a total probability theorem, while (19)-(20) are Bayes' rules [4, 8-10, 15].

### Paths and Connections

A (directed) *path* from  $x$  to  $y$  ( $x \neq y$ ) in  $G$ , denoted by  $[x \dots y]$ , is a sequence of nodes  $x_1, \dots, x_n$  such that  $x_1 = x$  and  $x_n = y$  and  $(x_i, x_{i+1}) \in B$  for every  $i$ ,  $1 \leq i \leq n-1$ . The *certainty*, *coverage* and *strength* of the path  $[x_1 \dots x_n]$  are defined respectively as:

$$cer[x_1 \dots x_n] = \prod_{i=1}^{n-1} cer(x_i, x_{i+1}), \quad (21)$$

$$cov[x_1 \dots x_n] = \prod_{i=1}^{n-1} cov(x_i, x_{i+1}) \quad (22)$$

and

$$\sigma[x \dots y] = \sigma(x)cer[x \dots y] = \sigma(y)cov[x \dots y]. \quad (23)$$

The set of all paths from  $x$  to  $y$  ( $x \neq y$ ) in  $G$ , denoted by  $\langle x, y \rangle$ , is a *connection* of  $G$  determined by nodes  $x$  and  $y$ . For every connection  $\langle x, y \rangle$ , its associate *certainty*, *coverage* and *strength of the connection*  $\langle x, y \rangle$  are defined as:

$$cer \langle x, y \rangle = \sum_{[x \dots y] \in \langle x, y \rangle} cer[x \dots y], \quad (24)$$

$$cov \langle x, y \rangle = \sum_{[x \dots y] \in \langle x, y \rangle} cov[x \dots y] \quad (25)$$

and

$$\sigma \langle x, y \rangle = \sum_{[x \dots y] \in \langle x, y \rangle} \sigma[x \dots y] = \sigma(x)cer \langle x, y \rangle = \sigma(y)cov \langle x, y \rangle. \quad (26)$$

If  $[x \dots y]$  is a path such that  $x$  and  $y$  are input and output of the graph  $G$ , respectively, then  $[x \dots y]$  will be referred to as *complete*. The set of complete paths from  $x$  to  $y$  will be called a *complete connection* from  $x$  to  $y$  in  $G$ . For simplicity, from now we will consider complete paths and connections only.

If we substitute every complete connection  $\langle x, y \rangle$  in  $G$ , where  $x$  and  $y$  are an input and an output (respectively) of a graph  $G$  with a single branch  $(x, y)$  such that  $\sigma(x, y) = \sigma \langle x, y \rangle$ ,  $cer(x, y) = cer \langle x, y \rangle$  and  $cov(x, y) = cov \langle x, y \rangle$  then we have a new flow graph with the property:  $\sigma(G) = \sigma(G')$ . This new flow graph is called a *combined flow graph*. It represents a relationship between its inputs and outputs more precisely.

### Dependencies

In this section, we show that the dependency factor of flow graphs has a close relationship to statistical measurements. For  $(x, y) \in B$ , nodes  $x$  and  $y$  are *independent* of each other if

$$\sigma(x, y) = \sigma(x)\sigma(y). \quad (27)$$

Thus, we have

$$\frac{\sigma(x, y)}{\sigma(x)} = cer(x, y) = \sigma(y). \quad (28)$$

Moreover, if

$$cer(x, y) > \sigma(y) \quad (29)$$

or

$$cov(x, y) > \sigma(x), \quad (30)$$

then  $x$  and  $y$  *depend positively* on each other. On the other hand, if

$$cer(x, y) < \sigma(y) \quad (31)$$

or

$$cov(x, y) < \sigma(x), \quad (32)$$

then  $x$  and  $y$  depend negatively on each other.

Furthermore, there is a factor that expresses the degree of dependency in the flow graph. For every  $(x, y) \in B$  we define a *dependency factor*,  $\eta(x, y)$ , as

$$\eta(x, y) = \frac{cer(x, y) - \sigma(y)}{cer(x, y) + \sigma(y)} = \frac{cov(x, y) - \sigma(x)}{cov(x, y) + \sigma(x)}. \quad (33)$$

The interpretation is: if  $\eta(x, y) = 0$ , then  $x$  and  $y$  are independent on each other, if  $-1 < \eta(x, y) < 0$ , then  $x$  and  $y$  are negatively dependent on each other and if  $0 < \eta(x, y) < 1$  then  $x$  and  $y$  are positively dependent on each other.

### Decision Algorithm

In this section, we present flow graphs that can be interpreted as a *decision algorithm*, taken from [4, 9].

Let us assume that the set of nodes of a graph is interpreted as a set of *formulas*, denoted by  $\Phi, \Psi$ . The formulas can be viewed as *propositional functions* in logic [4]. Then every branch  $(\Phi, \Psi)$  can be understood as a *decision rule*  $\Phi \rightarrow \Psi$ ,  $\Phi$  will be referred to as a *condition*, whereas  $\Psi$  is the *decision*. Such a rule is characterized by three numbers,  $\sigma(\Phi, \Psi)$ ,  $cer(\Phi, \Psi)$  and  $cov(\Phi, \Psi)$ . Thus, every *path*  $[\Phi_1 \dots \Phi_n]$  determines a sequence of decision rules  $\Phi_1 \rightarrow \Phi_2$ ,  $\Phi_2 \rightarrow \Phi_3, \dots, \Phi_{n-1} \rightarrow \Phi_n$ . As shown in [4], this sequence of decision rules can be interpreted as a single decision rule  $\Phi_1 \Phi_2 \dots \Phi_{n-1} \rightarrow \Phi_n$ , in short  $\Phi^* \rightarrow \Phi_n$ , where  $\Phi^* = \Phi_1 \wedge \Phi_2 \wedge \dots \wedge \Phi_{n-1}$ , characterized by three coefficients:

$$cer(\Phi^*, \Phi_n) = cer[\Phi_1 \dots \Phi_n], \quad (34)$$

$$cov(\Phi^*, \Phi_n) = cov[\Phi_1 \dots \Phi_n] \quad (35)$$

and

$$\sigma(\Phi^*, \Phi_n) = \sigma(\Phi_1)cer[\Phi_1 \dots \Phi_n] = \sigma(\Phi_n)cov[\Phi_1 \dots \Phi_n], \quad (36)$$

where  $\sigma(\Phi)$  can be considered as the truth value of the formula  $\Phi$  and  $\sigma(\Phi, \Psi)$  is the strength of the decision rule  $\Phi \rightarrow \Psi$ . Similarly, every *connection*  $\langle \Phi, \Psi \rangle$  can be interpreted as a single decision rule  $\Phi \rightarrow \Psi$  such that;

$$cer(\Phi, \Psi) = cer \langle \Phi, \Psi \rangle, \quad (37)$$

$$cov(\Phi, \Psi) = cov \langle \Phi, \Psi \rangle \quad (38)$$

and

$$\sigma(\Phi, \Psi) = \sigma(\Phi)cer \langle \Phi, \Psi \rangle = \sigma(\Psi)cov \langle \Phi, \Psi \rangle. \quad (39)$$

Let  $[\Phi_1 \dots \Phi_n]$  be a path such that  $\Phi_1$  is an input and  $\Phi_n$  is an output of flow graph  $G$ , respectively. Such a path and the corresponding connection  $\langle \Phi_1 \dots \Phi_n \rangle$  will be called *complete*. The set of all decision rules associated with all complete paths will be called a *decision algorithm* induced by the flow graph.

### Association Rule

An *association rule* identifies frequently occurring patterns of information in a given set of data [17, 18].

A rule is an expression  $X \rightarrow Y$ . A pattern  $X$  can be an item  $X_i$  or a conjunctive item set  $X_i \wedge \dots \wedge X_j$  where  $X_i, \dots, X_j$  are sets of data,  $1 \leq i, j \leq n, i \neq j$ .  $X$  is called the *antecedence*, while  $Y$  is called the *consequence* of the rule. A rule has two associated coefficients, called support and confidence [17, 18]. The *support* ( $sup$ ) of the rule  $X \rightarrow Y$  is the ratio of the number of data satisfy  $X$  and  $Y$  to the number of the total data. The *confidence* ( $conf$ ) of the rule is the ratio of the number of data which satisfy  $X$  and  $Y$  to the size of the set  $X$ . If we consider Example 1, a simple association rule is

(*SocialClass* = high  $\wedge$  *AgeGroup* = old)  $\rightarrow$  *VotedParty* =  $z_1$ ,  $sup = 0.04$ ,  $conf = 0.67$ .

We can interpret that when there is a high social class and old age group the voted party  $z_1$  will occur with 4 percent, while 67 percent of the high social class and old age group voted for party  $z_1$ .

**Remark a.**  $sup[xyz] = \frac{\varphi[xyz]}{\varphi(G)}$ .

**b.**  $conf[xyz] = \frac{\varphi[xyz]}{\varphi(x \wedge y)}$ .

We next come to our proposed propositions to reveal relationship between association rules and flow graphs.

**Proposition 1.** Let  $G = (N, B, \sigma)$  be a flow graph. The strength of  $(x, y)$  is given by  $\sigma(x, y) = \frac{\varphi(x, y)}{\varphi(G)}$  where  $0 \leq \sigma(x, y) \leq 1$ . A path from  $x$  to  $y$  ( $x \neq y$ ) in  $G$ , denoted by  $[x \dots y]$ , is a sequence of nodes  $x_1, \dots, x_n$  such that  $x_1 = x$  and  $x_n = y$  and  $(x_i, x_{i+1}) \in B$  for every  $i, 1 \leq i \leq n - 1$ .

For every path  $[xyz]$ ,  $cer(y, z) = 1$ , an association rule:  $(x \wedge y) \rightarrow z$ , where  $x, y$ , are antecedences and  $z$  is a consequence has

$$sup[xyz] = \sigma[xyz] \quad (40)$$

where  $0 \leq sup[xyz] \leq 1$ .

**Proof** By Remark a,  $sup[xyz] = \frac{\varphi[xyz]}{\varphi(G)}$ . Form Remark 1 and (4), we have  $sup[xyz] = \frac{\varphi(x, y)}{\varphi(G)} = \sigma(x, y)$ , respectively.

We next multiply by  $\frac{\sigma(x)}{\sigma(x)}$  and use (9) to obtain  $sup[xyz] = \frac{\sigma(x)\sigma(x, y)}{\sigma(x)} = \sigma(x)cer(x, y)$ .

If  $cer(y, z) = 1$ , then  $sup[xyz] = \sigma(x)cer(x, y)cer(y, z)$ .

Finally, (16) and (18) give us  $sup[xyz] = \sigma[xyz]$ . Q.E.D.

**Proposition 2.** Let  $G = (N, B, \sigma)$  be a flow graph. The strength of  $(x, y)$  is given by  $\sigma(x, y) = \frac{\varphi(x, y)}{\varphi(G)}$  where  $0 < \sigma(x, y) \leq 1$ . A path from  $x$  to  $y$  ( $x \neq y$ ) in

$G$ , denoted by  $[x \dots y]$ , is a sequence of nodes  $x_1, \dots, x_n$  such that  $x_1 = x$  and  $x_n = y$  and  $(x_i, x_{i+1}) \in B$  for every  $i, 1 \leq i \leq n - 1$ .

For every path  $[xyz]$ ,  $cer(y, z) = 1$ , an association rule:  $(x \wedge y) \rightarrow z$ , where  $x, y$ , are antecedences and  $z$  is a consequence has

$$conf[xyz] = 1 \quad (41)$$

where  $0 \leq conf[xyz] \leq 1$ .

**Proof** By Remarks b and (1),  $conf[xyz] = \frac{\varphi[xyz]}{\varphi(x \wedge y)} = \frac{\varphi(x \wedge y)}{\varphi(x \wedge y)} = 1$ . Q.E.D.

### 1.3 Illustrative Example

In this section we will illustrate the above ideas by means of a simple example concerning a voting analysis as depicted in Fig. 1. It is a modification of the example give in [9].

Suppose we have three disjoint age groups of voters  $y_1$  (old),  $y_2$  (middle aged) and  $y_3$  (young) belonging to three social classes  $x_1$  (high),  $x_2$  (middle) and  $x_3$  (low). The voters voted for three political parties  $z_1, z_2$  and  $z_3$ . In this population, e.g., certainties of input layer indicate that 10 percent of votets has high social class, 70 percent has middle social class and 20 percent has low social class. Similarly, certainties of hidden layer, age group, and the output layer, party, have their values as shown in Fig. 2.

We can interpret its coverages that, e.g., party  $z_1$  obtains 27 percent of total votes, which 70 percent is from old age group  $y_1$  and 30 percent is from young age group  $y_3$ . One can see that the lower approximation of the party  $z_1$  ( $cer = 1$ ) is an empty set and upper approximation of the party  $z_1$  ( $cer \neq 1$ ) is  $(x_1 \cap y_1) \cup (x_2 \cap y_1) \cup (x_2 \cap y_3) \cup (x_3 \cap y_3)$ .

Similarly, the lower approximation of the party  $z_2$  is  $(x_1 \cap y_2) \cup (x_2 \cap y_2) \cup (x_3 \cap y_2)$  and the upper approximation of the party  $z_2$  is  $(x_1 \cap y_1) \cup (x_2 \cap y_1) \cup (x_1 \cap y_2) \cup (x_2 \cap y_2) \cup (x_3 \cap y_2)$ .

Form Fig. 2, since  $cer(y_2, z_2) = 1$ , we obtain three association rules by using our presented propositions as:

$$\begin{aligned} (x_1 \wedge y_2) &\rightarrow z_2, \text{ sup} = 0.04 \text{ and } conf = 1, \\ (x_2 \wedge y_2) &\rightarrow z_2, \text{ sup} = 0.42 \text{ and } conf = 1, \\ (x_3 \wedge y_2) &\rightarrow z_2, \text{ sup} = 0.14 \text{ and } conf = 1. \end{aligned}$$

Next, if we want to know how votes are distributed between parties with respect to social classes, we can compute the combined of the corresponding graph. Employing the algorithm presented in [8, 9, 16], we obtain the results as in Fig. 3.

For example, we can see that the party  $z_3$  obtains 5 percent of total vote, of which 54 percent is from the social class  $x_2$  and 46 percent is from the social class  $x_3$ . When we consider the dependency factor, the negative dependency between high social class and party  $z_2$  is  $\eta(x_1, z_2) = -0.08$  with strength  $\sigma = 0.04$ . The negative dependency reveals that when one node is increasing, its pair is decreasing as well.

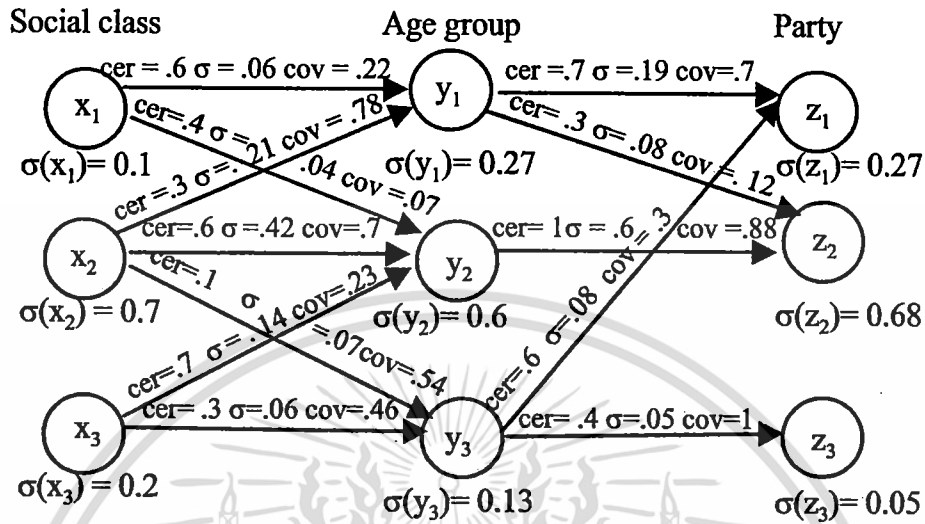


Fig. 2. Social class, age group and party votes flow graph.

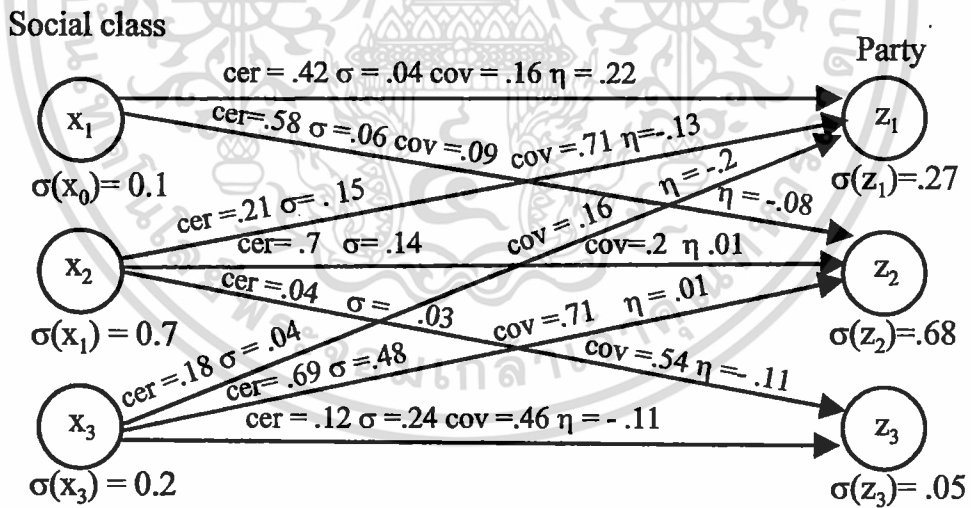


Fig. 3. Combined of flow graph.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 Experimental Results of the POSN Data Set

In this section, we describe experimental results from the scores of Olympiad mathematics' students' analysis.

We depict an initial data flow graph in Fig. 4. It contains 68 records collected from the Promotion of Academic Olympiads and Development of Science Education Foundation (POSN). It consists of four disjoint preference subjects:  $x_1$  (Geometry),  $x_2$  (Number theory),  $x_3$  (Algebra) and  $x_4$  (Combinatorics) which belong to four most practiced subjects:  $y_1$  (Geometry),  $y_2$  (Number theory),  $y_3$  (Algebra) and  $y_4$  (Combinatorics). The decision attribute is the students' total score,  $z$ , that we will compare 4 discretization cases presented below. The thresholds for association rule extraction are  $sup > 0.1$  and  $conf > 0.35$ .

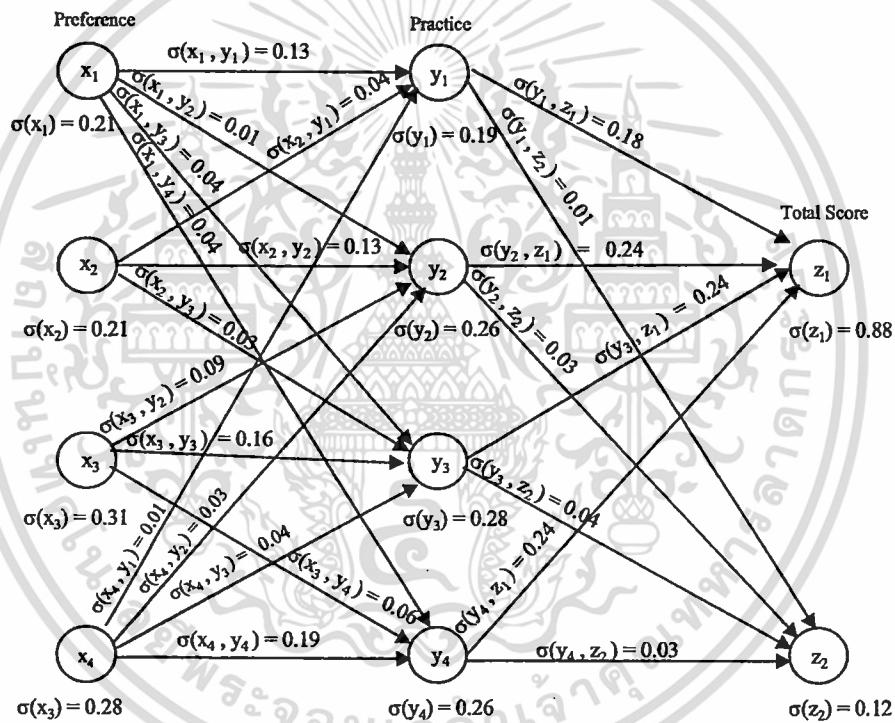


Fig. 4. Flow graph of initial Olympiad mathematics' students score from POSN.

Case 1. If  $z_1$  (score  $\geq 103$ ),  $z_2$  ( $80 \leq \text{score} \leq 102$ ),  $z_3$  ( $70 \leq \text{score} \leq 79$ ) and  $z_4$  (score  $\leq 69$ ), then the selected association rule is:

$$\boxed{\text{IF } \text{Prac\_Comb} \text{ THEN } \text{score} \geq 103, \text{ sup} = 0.10, \text{ conf} = 0.39^+,}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

with  $sup = 0.10$ ,  $conf = 0.39$  and  $accuracy = 97.6\%$ .

**Case 2.** If  $z_1$  (score  $\geq 147$ ),  $z_2$  ( $118 \leq \text{score} \leq 146$ ),  $z_3$  ( $88 \leq \text{score} \leq 117$ ) and  $z_4$  (score  $\leq 87$ ), then the selected association rules are:

```
IF Prefer_Alg and Prac_Alg THEN score  $\leq 87$ , sup = 0.12,
conf = 0.73,
IF Prefer_Alg THEN  $66 \leq \text{score} \leq 74$ , sup = 0.13, conf = 0.64,
IF Prefer_Comb THEN  $66 \leq \text{score} \leq 74$ , sup = 0.16, conf =
0.59,
IF Prefer_Number THEN  $66 \leq \text{score} \leq 74$ , sup = 0.17, conf =
0.57,
IF Prac_Alg THEN  $66 \leq \text{score} \leq 74$ , sup = 0.2, conf = 0.74,
```

with *average sup* = 0.16 *average conf* = 0.65 and *accuracy* = 86.76 %.

**Case 3.** If  $z_1$  (score  $\geq 94$ ),  $z_2$  ( $73 \leq \text{score} \leq 93$ ) and  $z_3$  ( $58 \leq \text{score} \leq 72$ ), then the selected association rules are:

```
IF Prefer_Comb THEN score  $\geq 180$ , sup = 0.10, conf = 0.37,
IF Prac_Alg THEN  $73 \leq \text{score} \leq 93$ , sup = 0.10, conf = 0.37,
IF Prac_Number THEN score  $\geq 180$ , sup = 0.10, conf = 0.379,
IF Prefer_Geo THEN score  $\geq 180$ , sup = 0.10, conf = 0.5,
IF Prac_Alg THEN  $58 \leq \text{score} \leq 72$ , sup = 0.12, conf = 0.44,
IF Prac_Comb THEN  $\geq 180$ , sup = 0.12, conf = 0.42,
```

with *average sup* = 0.11, *average conf* = 0.42 and *accuracy* = 80.88 %.

**Case 4.** If  $z_1$  (score  $\geq 61$ ) and  $z_2$  (score  $\leq 60$ ), then the selected association rules are:

```
IF Prefer_Geo and Prac_Geo THEN score  $\geq 61$ , sup = 0.13,
conf = 1,
IF Prefer_Number and Prac_Number THEN score  $\geq 61$ , sup =
0.13, conf = 0.78,
IF Prefer_Alg and Prac_Alg THEN score  $\geq 61$ , sup = 0.15,
conf = 0.90,
IF Prefer_Comb and Prac_Comb THEN score  $\geq 61$ , sup = 0.18,
conf = 0.92,
IF Prefer_Geo THEN score  $\geq 61$ , sup = 0.17, conf = 0.86,
IF Prefer_Alg THEN score  $\leq 60$ , sup = 0.17, conf = 0.86,
IF Prac_Geo THEN score  $\leq 60$ , sup = 0.18, conf = 0.92,
IF Prac_Alg THEN score  $\leq 60$ , sup = 0.23, conf = 0.84,
IF Prac_Num THEN score  $\leq 60$ , sup = 0.24, conf = 0.89,
IF Prac_Comb THEN score  $\leq 60$ , sup = 0.25, conf = 0.89,
IF Prefer_Comb THEN score  $\leq 60$ , sup = 0.28, conf = 0.90,
IF Prefer_Alg THEN score  $\leq 60$ , sup = 0.28, conf = 0.90,
```

with *average sup* = 0.19, *average conf* = 0.89 and *accuracy* = 51.47 %.

From these selected association rules, we can distil knowledge from this flow graph: if students practice their preferred subjects, then they will have high scores. In Case 4, we obtain maximum average support and confidence as shown in Fig. 5. The highest average accuracy to predict students' score is 97.6 % from Case 1 as shown in Fig. 5.

<sup>1</sup> Due to round off errors in computations, the equation may be not satisfied exactly in our examples.

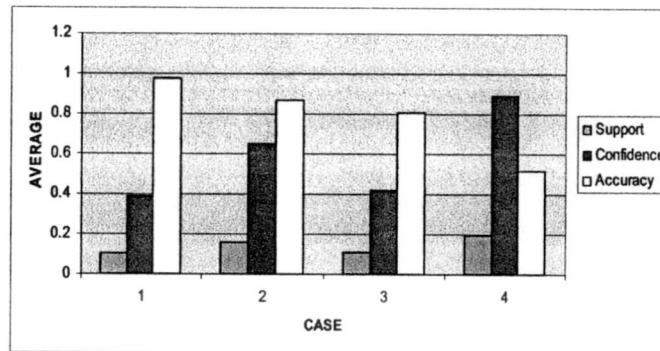


Fig. 5. Comparison of *sup*, *conf* and *accuracy/100* from four cases of POSN data set.

## 2 Rough Sets, Flow Graphs, Fuzzy Sets and their Extension to Decision Algorithm

Fuzzy set theory was introduced by Lofti Zadeh in 1965 [21]. A fuzzy set is a class of objects with a continuum of grader of membership. The characteristic of *membership function* of a fuzzy set is permitted to have value in the interval  $[0, 1]$ . If  $A$  is a classical set, its membership function can take on only two values  $0$  and  $1$ . It can be understood as a vague boundary set theory. Recently, fuzzy set theory and its applications were developed extensively and fascinated attention [21]. In real world data analysis, if collected data is in the form of fuzzy attributes, then it can be represented in the decision table with fuzzy attributes. We can analyze this decision table by applying concepts of fuzzy set. Alicja et al. introduced a concept of fuzzy flow graphs in [22]. They proposed to allow computation not only for descriptive attributes but also for fuzzy attributes in flow graphs.

In this research, we introduce an alternative approach to construct flow graphs, decision rules and perform data analysis for decision table with fuzzy attributes. Moreover, we relax mutual exclusion and find four new coefficients which are satisfied coefficient in flow graph.

### 2.1 Fuzzy Set Theory

The following fuzzy set theory concept is taken from [21]. Let  $X$  be a space of objects and  $x$  be a generic element of  $X$ . A *fuzzy set*  $A$  in  $X$  is characterized by a *membership function*,  $\mu_A(x) \rightarrow [0, 1]$ . The nearer the value of  $\mu_A(X)$  to  $1$ , the higher the grade of membership of  $x$  in  $A$ . A crucial issue for fuzzy sets applications is to express real-word problem by fuzzy sets attributes in decision table and perform data analysis. In further discussions, we consider decision table with fuzzy attributes in the form introduced in [22, 23].

Let  $(U, C \cup D)$  be a decision table with fuzzy attributes,  $U = \{x_1, x_2, \dots, x_N\}$  be a finite non-empty set of universe with  $N$  elements,  $C = \{c_1, c_2, \dots, c_n\}$  be a set of condition attributes and  $D = \{d_1, d_2, \dots, d_m\}$  be a set of decision

attributes. The domain of the family of linguistic values of condition attribute  $c_i$  denoted by  $V_i = \{V_{i_1}, V_{i_2}, \dots, V_{i_{n_i}}\}$ , where  $n_i$  is the number of the linguistic values of  $i$ -th conditions,  $i = 1, 2, \dots, n$ . The domain of the family of linguistic values of decision attribute  $d_i$  denoted by  $W_j = \{W_{j_1}, W_{j_2}, \dots, W_{j_{m_j}}\}$ , where  $m_j$  is the number of the linguistic values of  $j$ -th conditions,  $j = 1, 2, \dots, m$ .

In order to describe an element  $x \in U$  with fuzzy attribute, its membership degree should be determined. This process is called fuzzification state. In this state, for any element  $x$ , fuzzy value of condition attribute  $c_i$  and decision attribute  $d_j$  denoted by  $V_i(x)$  and  $W_j(x)$  are defined as

$$V_i(x) = \left\{ \frac{\mu_{v_{i_1}}(x)}{V_{i_1}}, \frac{\mu_{v_{i_2}}(x)}{V_{i_2}}, \dots, \frac{\mu_{v_{i_{n_i}}}(x)}{V_{i_{n_i}}} \right\} \text{ and}$$

$$W_j(x) = \left\{ \frac{\mu_{w_{j_1}}(x)}{W_{j_1}}, \frac{\mu_{w_{j_2}}(x)}{W_{j_2}}, \dots, \frac{\mu_{w_{j_{m_j}}}(x)}{W_{j_{m_j}}} \right\}. \quad 2$$

Furthermore, for any  $x \in U$  we assume that values of  $V_i(x)$  and  $W_j(x)$  satisfy  $power(V_i(x)) = \sum_{k=1}^{n_i} \mu_{V_{i_k}}(x) = 1$  and  $power(W_j(x)) = \sum_{k=1}^{m_j} \mu_{W_{j_k}}(x) = 1$ ,  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, m$ . An example of fuzzy values  $V_i(x)$  of  $c_i$  and fuzzy values  $W_j(x)$  of  $d_j$  is in Table 1.

Table 1. Decision table with fuzzy attributes

	$c_1$	$c_2$	...	$c_n$	$d_1$	$d_2$	...	$d_m$
$x_1$	$V_1(x_1)$	$V_2(x_1)$	...	$V_n(x_1)$	$W_1(x_1)$	$W_2(x_1)$	...	$W_m(x_1)$
$x_2$	$V_1(x_2)$	$V_2(x_2)$	...	$V_n(x_2)$	$W_1(x_2)$	$W_2(x_2)$	...	$W_m(x_2)$
...								
$x_N$	$V_1(x_N)$	$V_2(x_N)$	...	$V_n(x_N)$	$W_1(x_N)$	$W_2(x_N)$	...	$W_m(x_N)$

## 2.2 Decision Rules, Decision Algorithm and Flow Graph

In this section, basic concepts of decision rules and decision algorithms that can be used as classifiers in logical form [20] are provided.

Let  $S = (U, A)$  be an information system, for any subset  $B$  of  $A$ , there is an associated set of formulas  $For(B)$ .  $For(B)$  are built up from attribute value pairs  $(a, v)$  where  $a \in B$  and  $v \in V_a$ . For any  $\Phi \in For(B)$ , let  $\|\Phi\|$  be the set of all objects from universe  $U$  satisfy  $\Phi$  in  $S$ . By means of logical connectives AND ( $\wedge$ ), OR ( $\vee$ ) and NOT ( $\neg$ ), Pawlak [20] defined some inductively properties as follows:

$$\|\Phi \vee \Psi\| = \|\Phi\| \cup \|\Psi\|,$$

<sup>2</sup> If the values of all attributes are singletons or disjoint intervals on the original domain of attributes, then we obtain a classical crisp decision table.

$$\|\Phi \wedge \Psi\| = \|\Phi\| \cap \|\Psi\| \text{ and} \\ \|\neg\Phi\| = U - \|\Phi\|.$$

A decision rule in  $S$  is an expression in the form  $\Phi \rightarrow \Psi$ ,  $\Phi$  and  $\Psi$  are referred to as a *condition* and *decision* of the rule, respectively. For every decision rules, there are three associated coefficients:  $\sigma(\Phi, \Psi) = \frac{\text{card}(\|\Phi\| \cap \|\Psi\|)}{\text{card}(U)}$ ,  $\text{cer}(\Phi, \Psi) = \frac{\text{card}(\|\Phi\| \cap \|\Psi\|)}{\text{card}(\|\Phi\|)}$  and  $\text{cov}(\Phi, \Psi) = \frac{\text{card}(\|\Phi\| \cap \|\Psi\|)}{\text{card}(\|\Psi\|)}$  [4].

A set of all decision rules  $\text{Dec}(\Phi, \Psi) = \{\Phi_i \rightarrow \Psi_i\}_{i=1}^n, n \geq 2$ , will be called a *decision algorithm* if all decision rules are satisfied conditions:

- (1) *Mutually exclusive*: for every  $\Phi_i \rightarrow \Psi_i$  and  $\Phi_j \rightarrow \Psi_j$ ,  $\|\Phi_i\| \cap \|\Phi_j\| = \emptyset$  and  $\|\Psi_i\| \cap \|\Psi_j\| = \emptyset$  for  $i \neq j$ ,
- (2) *Admissible*:  $\|\Phi \wedge \Psi\| \neq \emptyset$ ,
- (3) *Covering*:  $\bigcup_{i=1}^n \|\Phi_i\| = U$  and  $\bigcup_{i=1}^n \|\Psi_i\| = U$ .

A flow graph can be interpreted as decision algorithms [4] and will be described as follows. Let us assume that the set of nodes of a flow graph is interpreted as a set of *formulas*, denoted by  $\Phi, \Psi$ <sup>3</sup>. Then every branch  $(\Phi, \Psi)$  can be understood as a *decision rule*  $\Phi \rightarrow \Psi$ .  $\Phi$  will be referred to as a *condition*, whereas  $\Psi$  is *decision*. and  $\sigma(\Phi, \Psi)$  denotes strength coefficient. Such a rule is characterized by three coefficients strength, certainty and coverage,  $\sigma(\Phi, \Psi) = \frac{\text{card}(\|\Phi\| \cap \|\Psi\|)}{\text{card}(U)}$ ,  $\text{cer}(\Phi, \Psi) = \frac{\text{card}(\|\Phi\| \cap \|\Psi\|)}{\text{card}(\|\Phi\|)}$  and  $\text{cov}(\Phi, \Psi) = \frac{\text{card}(\|\Phi\| \cap \|\Psi\|)}{\text{card}(\|\Psi\|)}$  [4].

Next, a *path*,  $[\Phi_1 \dots \Phi_n]$ , is a sequence of decision rules  $\Phi_1 \rightarrow \Phi_2, \Phi_2 \rightarrow \Phi_3, \dots, \Phi_{n-1} \rightarrow \Phi_n$ . As shown in [4], this sequence of decision rules can be interpreted as a single decision rule  $\Phi_1 \Phi_2 \dots \Phi_{n-1} \rightarrow \Phi_n$ , in short  $\Phi^* \rightarrow \Phi_n$ , where  $\Phi^* = \Phi_1 \wedge \Phi_2 \wedge \dots \wedge \Phi_{n-1}$ . It is characterized by three coefficients:  $\text{cer}(\Phi^*, \Phi_n) = \text{cer}[\Phi_1 \dots \Phi_n]$ ,  $\text{cov}(\Phi^*, \Phi_n) = \text{cov}[\Phi_1 \dots \Phi_n]$  and  $\sigma(\Phi^*, \Phi_n) = \sigma(\Phi_1) \text{cer}[\Phi_1 \dots \Phi_n] = \sigma(\Phi_n) \text{cov}[\Phi_1 \dots \Phi_n]$ , where  $\sigma(\Phi)$  can be considered as the truth value of the formula  $\Phi$ .

Let  $[\Phi_1 \dots \Phi_n]$  be a path such that  $\Phi_1$  is an input and  $\Phi_n$  is an output of flow graph  $G$ , respectively, then the  $[\Phi_1 \dots \Phi_n]$  will be referred as *complete*. The set of all paths from  $\Phi_1$  to  $\Phi_n$  denoted by  $\langle \Phi_1 \dots \Phi_n \rangle$  is called complete paths. The set of all decision rules associated with all complete paths will be called a *decision algorithm* induced by the flow graph.

### 2.3 New Certainty and Coverage Coefficients for Decision Rules from Relaxed Mutually Exclusive

In this section, we apply concept of mutually exclusive in decision algorithm introduced by Pawlak [4] to find new certainty and coverage coefficients. Since, we assume  $\|\Phi_i\|$  and  $\|\Psi_i\|, i = 1, 2, \dots, n$ , are set of all objects from universe  $U$  satisfy condition and decision attributes. If we want to connect any set of conditions (set of decisions) by logical connectives AND ( $\wedge$ ) or OR ( $\vee$ ) and find

<sup>3</sup> The formulas can be viewed as *propositional functions* in logic [4].

certainty and coverage of these decision rules. Then by Pawlak's definition of certainty and coverage, the following Remarks 1 and 2 are holded.

**Remark 1** Let  $\Phi$  and  $\Psi_i$ ,  $i = 1, 2, \dots, n$  be the set of condition and decision attributes, then

$$\begin{aligned} 1. \text{cer}(\Phi, \Psi_1 \vee \dots \vee \Psi_n) &= \frac{\text{card}(\|\Phi\| \cap (\|\Psi_1\| \cup \dots \cup \|\Psi_n\|))}{\text{card}(\|\Phi\|)}, \\ 2. \text{cer}(\Phi, \Psi_1 \wedge \dots \wedge \Psi_n) &= \frac{\text{card}(\|\Phi\| \cap (\|\Psi_1\| \cap \dots \cap \|\Psi_n\|))}{\text{card}(\|\Phi\|)}, \\ 3. \text{cov}(\Phi, \Psi_1 \vee \dots \vee \Psi_n) &= \frac{\text{card}(\|\Phi\| \cap (\|\Psi_1\| \cup \dots \cup \|\Psi_n\|))}{\text{card}(\|\Psi_1\| \cup \dots \cup \|\Psi_n\|)}, \\ 4. \text{cov}(\Phi, \Psi_1 \wedge \dots \wedge \Psi_n) &= \frac{\text{card}(\|\Phi\| \cap (\|\Psi_1\| \cap \dots \cap \|\Psi_n\|))}{\text{card}(\|\Psi_1\| \cap \dots \cap \|\Psi_n\|)}. \end{aligned}$$

**Remark 2** Let  $\Phi_i$  and  $\Psi$ ,  $i = 1, 2, \dots, n$  be the set of condition and decision attributes, then

$$\begin{aligned} 1. \text{cov}(\Phi_1 \vee \dots \vee \Phi_n, \Psi) &= \frac{\text{card}((\|\Phi_1\| \cup \dots \cup \|\Phi_n\|) \cap \|\Psi\|)}{\text{card}(\|\Psi\|)}, \\ 2. \text{cov}(\Phi_1 \wedge \dots \wedge \Phi_n, \Psi) &= \frac{\text{card}((\|\Phi_1\| \cap \dots \cap \|\Phi_n\|) \cap \|\Psi\|)}{\text{card}(\|\Psi\|)}, \\ 3. \text{cer}(\Phi_1 \vee \dots \vee \Phi_n, \Psi) &= \frac{\text{card}((\|\Phi_1\| \cup \dots \cup \|\Phi_n\|) \cap \|\Psi\|)}{\text{card}(\|\Phi_1\| \cup \dots \cup \|\Phi_n\|)}, \\ 4. \text{cer}(\Phi_1 \wedge \dots \wedge \Phi_n, \Psi) &= \frac{\text{card}((\|\Phi_1\| \cap \dots \cap \|\Phi_n\|) \cap \|\Psi\|)}{\text{card}(\|\Phi_1\| \cap \dots \cap \|\Phi_n\|)}. \end{aligned}$$

**Remark 3** is the property obtained from set theory.

**Remark 3** Let  $\Phi_i$ ,  $i = 1, 2, \dots, n$  be a set of attribute, then  $\text{card}(\|\Phi_1\| \cup \|\Phi_2\| \cup \dots \cup \|\Phi_n\|)$

$$\begin{aligned} &= \sum_{i=1}^n \text{card}(\|\Phi_i\|) - \sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^n \text{card}(\|\Phi_{i_1}\| \cap \|\Phi_{i_2}\|) \\ &+ \sum_{i_1=1}^{n-2} \sum_{i_2=i_1+1}^{n-1} \sum_{i_3=i_2+1}^n \text{card}(\|\Phi_{i_1}\| \cap \|\Phi_{i_2}\| \cap \|\Phi_{i_3}\|) \\ &- \dots + (-1)^{n-1} \text{card}(\|\Phi_{i_1}\| \cap \|\Phi_{i_2}\| \cap \dots \cap \|\Phi_n\|). \end{aligned}$$

From mutual exclusive of decision algorithm, there are assumptions that for every decision rules  $\Phi_i \rightarrow \Psi_i$  and  $\Phi_j \rightarrow \Psi_j$ ,

$$\|\Phi_i\| \cap \|\Phi_j\| = \emptyset \text{ and } \|\Psi_i\| \cap \|\Psi_j\| = \emptyset$$

for  $i \neq j$ .

In this research, we propose relaxation of the mutual exclusion property (mentioned above) as follows. There exist at least two decision rules  $\Phi_i \rightarrow \Psi$  and  $\Phi \rightarrow \Psi_i$  such that

$$\|\Phi_{i_m}\| \cap \|\Phi_{i_k}\| \neq \emptyset \text{ or } \|\Psi_{i_m}\| \cap \|\Psi_{j_k}\| \neq \emptyset$$

for  $i_m \neq j_k$ .

We can describe this of relax mutual exclusion as the following. For example, if we have decision rules,

IF *score\_of\_Number\_Theory* is high THEN *total\_score* is high  
and

IF *score\_of\_Algebra* is high THEN *total\_score* is high,

in a decision algorithm, then there may be some elements satisfy both conditions. Thus, our proposed certainty and coverage coefficients are more flexible with the relaxed mutually exclusive. If there exist at least two decisions  $\|\Psi_i\|$  and  $\|\Psi_j\|$  such that  $\|\Psi_i\| \cap \|\Psi_j\| \neq \emptyset$ , we can relax mutually exclusive for decisions (as in Propositions 1, 2). If there exist at least two conditions  $\|\Phi_i\|$  and  $\|\Phi_j\|$  such that  $\|\Phi_i\| \cap \|\Phi_j\| \neq \emptyset$ , we can relax mutually exclusive for condition (as shown Propositions 3, 4).

**Proposition 1** If we are given a set of decision rules,  $\{\Phi \rightarrow \Psi_i\}_n$ , such that  $\Psi_{i_m} \cap \Psi_{i_k} \neq \emptyset$ , for any  $i_m \neq i_k$ . Then  $cer(\Phi, \Psi_1 \vee \Psi_2 \vee \dots \vee \Psi_n)$

$$\begin{aligned} &= \sum_{i=1}^n cer(\Phi, \Psi_i) - \sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^n cer(\Phi, \Psi_{i_1} \wedge \Psi_{i_2}) \\ &+ \sum_{i_1=1}^{n-2} \sum_{i_2=i_1+1}^{n-1} \sum_{i_3=i_2+1}^n cer(\Phi, \Psi_{i_1} \wedge \Psi_{i_2} \wedge \Psi_{i_3}) - \\ &\dots + (-1)^{n-1} cer(\Phi, \Psi_1 \wedge \Psi_2 \wedge \dots \wedge \Psi_n). \end{aligned}$$

*Proof.* If we denote  $cer(\Phi, \Psi_1 \vee \Psi_2 \vee \dots \vee \Psi_n)$  by  $cer(\Phi, \Psi_i)$ , then by Remark 1(1) we have  $cer(\Phi, \Psi_i) = \frac{card(\|\Phi\| \cap (\|\Psi_1\| \cup \|\Psi_2\| \cup \dots \cup \|\Psi_n\|))}{card(\|\Phi\|)}$ . Next, by left-distributive over union, we obtain

$$\begin{aligned} cer(\Phi, \Psi_i) &= \frac{card((\|\Phi\| \cap \|\Psi_1\|) \cup (\|\Phi\| \cap \|\Psi_2\|) \cup \dots \cup (\|\Phi\| \cap \|\Psi_n\|))}{card(\|\Phi\|)} \\ &= \sum_{i=1}^n \frac{card(\|\Phi\| \cap \|\Psi_i\|)}{card(\|\Phi\|)} \\ &- \sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^n \frac{card(\|\Phi\| \cap \|\Psi_{i_1}\| \cap \|\Psi_{i_2}\|)}{card(\|\Phi\|)} \\ &+ \sum_{i_1=1}^{n-2} \sum_{i_2=i_1+1}^{n-1} \sum_{i_3=i_2+1}^n \frac{card(\|\Phi\| \cap \|\Psi_{i_1}\| \cap \|\Psi_{i_2}\| \cap \|\Psi_{i_3}\|)}{card(\|\Phi\|)} \\ &\dots + (-1)^{n-1} \frac{card(\|\Phi\| \cap \|\Psi_1\| \cap \|\Psi_2\| \cap \dots \cap \|\Psi_n\|)}{card(\|\Phi\|)} \text{ (by Remark 3)} \\ &= \sum_{i=1}^n cer(\Phi, \Psi_i) - \sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^n cer(\Phi, \Psi_{i_1} \wedge \Psi_{i_2}) \\ &+ \sum_{i_1=1}^{n-2} \sum_{i_2=i_1+1}^{n-1} \sum_{i_3=i_2+1}^n cer(\Phi, \Psi_{i_1} \wedge \Psi_{i_2} \wedge \Psi_{i_3}) \\ &\dots + (-1)^{n-1} cer(\Phi, \Psi_1 \wedge \Psi_2 \wedge \dots \wedge \Psi_n) \text{ (by Remark 1(2)).} \end{aligned}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**Proposition 2** If we are given a set of decision rules,  $\{\Phi \rightarrow \Psi_i\}_n$ , such that  $\Psi_{i_m} \cap \Psi_{i_k} \neq \emptyset$ , for any  $i_m \neq i_k$ . Then  $\text{cov}(\Phi, \Psi_1 \vee \Psi_2 \vee \dots \vee \Psi_n) < \sum_{i=1}^n \text{cov}(\Phi, \Psi_i)$ .

*Proof.* If we denote  $\text{cov}(\Phi, \Psi_1 \vee \Psi_2 \vee \dots \vee \Psi_n)$  by  $\text{cov}(\Phi, \Psi_i)$ , then by Remark 1(3) we have  $\text{cov}(\Phi, \Psi_i) = \frac{\text{card}(\|\Phi\| \cap (\|\Psi_1\| \cup \|\Psi_2\| \cup \dots \cup \|\Psi_n\|))}{\text{card}(\|\Psi_1\| \cup \|\Psi_2\| \cup \dots \cup \|\Psi_n\|)}$ . Next, by left-distributive over union and inequality property, we obtain

$$\begin{aligned} \text{cov}(\Phi, \Psi_i) &= \frac{\text{card}(\|\Phi\| \cap \|\Psi_1\| \cup (\|\Phi\| \cap \|\Psi_2\|) \cup \dots \cup (\|\Phi\| \cap \|\Psi_n\|))}{\text{card}(\|\Psi_1\| \cup \|\Psi_2\| \cup \dots \cup \|\Psi_n\|)} \\ &< \sum_{i=1}^n \frac{\text{card}(\|\Phi\| \cap \|\Psi_i\|)}{\text{card}(\|\Psi_1\| \cup \|\Psi_2\| \cup \dots \cup \|\Psi_n\|)} \\ &< \sum_{i=1}^n \frac{\text{card}(\|\Phi\| \cap \|\Psi_i\|)}{\text{card}(\|\Psi_i\|)}. \end{aligned}$$

Then  $\text{cov}(\Phi, \Psi_i) < \sum_{i=1}^n \text{cov}(\Phi, \Psi_i)$  (by Remark 1(3)).

**Proposition 3** If we are given a set of decision rules,  $\{\Phi_i \rightarrow \Psi\}_n$ , such that  $\Phi_{i_m} \cap \Phi_{i_k} \neq \emptyset$ , for any  $i_m \neq i_k$ . Then  $\text{cov}(\Phi_1 \vee \Phi_2 \vee \dots \vee \Phi_n, \Psi)$

$$\begin{aligned} &= \sum_{i=1}^n \text{cov}(\Phi_i, \Psi) - \sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^n \text{cov}(\Phi_{i_1} \wedge \Phi_{i_2}, \Psi) \\ &+ \sum_{i_1=1}^{n-2} \sum_{i_2=i_1+1}^{n-1} \sum_{i_3=i_2+1}^n \text{cov}(\Phi_{i_1} \wedge \Phi_{i_2} \wedge \Phi_{i_3}, \Psi) - \\ &\dots + (-1)^{n-1} \text{cov}(\Phi_1 \wedge \Phi_2 \wedge \dots \wedge \Phi_n, \Psi). \end{aligned}$$

*Proof.* If we denote  $\text{cov}(\Phi_1 \vee \Phi_2 \vee \dots \vee \Phi_n, \Psi)$  by  $\text{cov}(\Phi_i, \Psi)$ , then by Remark 2(1) we have  $\text{cov}(\Phi_i, \Psi) = \frac{\text{card}(\|\Phi_1\| \cup \|\Phi_2\| \cup \dots \cup \|\Phi_n\| \cap \|\Psi\|)}{\text{card}(\|\Psi\|)}$ . Next, by right-distributive over union, we obtain

$$\begin{aligned} \text{cov}(\Phi_i, \Psi) &= \frac{\text{card}(\|\Phi_1\| \cap \|\Psi\| \cup (\|\Phi_2\| \cap \|\Psi\|) \cup \dots \cup (\|\Phi_n\| \cap \|\Psi\|))}{\text{card}(\|\Psi\|)} \\ &= \sum_{i=1}^n \frac{\text{card}(\|\Phi_i\| \cap \|\Psi\|)}{\text{card}(\|\Psi\|)} \\ &- \sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^n \frac{\text{card}(\|\Phi_{i_1}\| \cap \|\Phi_{i_2}\| \cap \|\Psi\|)}{\text{card}(\|\Psi\|)} \\ &+ \sum_{i_1=1}^{n-2} \sum_{i_2=i_1+1}^{n-1} \sum_{i_3=i_2+1}^n \frac{\text{card}(\|\Phi_{i_1}\| \cap \|\Phi_{i_2}\| \cap \|\Phi_{i_3}\| \cap \|\Psi\|)}{\text{card}(\|\Psi\|)} \\ &- \dots + (-1)^{n-1} \frac{\text{card}(\|\Phi_1\| \cap \|\Phi_2\| \cap \dots \cap \|\Phi_n\| \cap \|\Psi\|)}{\text{card}(\|\Psi\|)} \text{ (by Remark 3)} \\ &= \sum_{i=1}^n \text{cov}(\Phi_i, \Psi) - \sum_{i_1=1}^{n-1} \sum_{i_2=i_1+1}^n \text{cov}(\Phi_{i_1} \wedge \Phi_{i_2}, \Psi) + \\ &\sum_{i_1=1}^{n-2} \sum_{i_2=i_1+1}^{n-1} \sum_{i_3=i_2+1}^n \text{cov}(\Phi_{i_1} \wedge \Phi_{i_2} \wedge \Phi_{i_3}, \Psi) - \dots \\ &+ (-1)^{n-1} \text{cov}(\Phi_1 \wedge \Phi_2 \wedge \dots \wedge \Phi_n, \Psi) \text{ (by Remark 2(1)).} \end{aligned}$$

**Proposition 4** If we are given a set of decision rules,  $\{\Phi_i \rightarrow \Psi\}_n$ , such that  $\Phi_{i_m} \cap \Phi_{i_k} \neq \emptyset$ , for any  $i_m \neq i_k$ . Then  $cer(\Phi_1 \vee \Phi_2 \vee \dots \vee \Phi_n, \Psi) < \sum_{i=1}^n cer(\Phi_i, \Psi)$ .

*Proof.* If we denote  $cer(\Phi_1 \vee \Phi_2 \vee \dots \vee \Phi_n, \Psi)$  by  $cer(\Phi_i, \Psi)$ , then by Remark 2(3) we have  $cer(\Phi_i, \Psi) = \frac{card((\Phi_1 \cup \Phi_2 \cup \dots \cup \Phi_n) \cap \Psi)}{card(\Phi_1 \cup \Phi_2 \cup \dots \cup \Phi_n)}$ . Next, by right-distributive over union and inequality property, we obtain

$$\begin{aligned} cer(\Phi_i, \Psi) &= \frac{card((\|\Phi_1\| \cap \|\Psi\|) \cup (\|\Phi_2\| \cap \|\Psi\|) \cup \dots \cup (\|\Phi_n\| \cap \|\Psi\|))}{card(\|\Phi_1\| \cup \|\Phi_2\| \cup \dots \cup \|\Phi_n\|)} \\ &< \sum_{i=1}^n \frac{card(\|\Phi_i\| \cap \|\Psi\|)}{card(\|\Phi_1\| \cup \|\Phi_2\| \cup \dots \cup \|\Phi_n\|)} \\ &< \sum_{i=1}^n \frac{card(\|\Phi_i\| \cap \|\Psi\|)}{card(\|\Phi_i\|)}. \end{aligned}$$

Then  $cer(\Phi_i, \Psi) < \sum_{i=1}^n cer(\Phi_i, \Psi)$  (by Remark 2(1)).

#### 2.4 Flow Graphs with Fuzzy Attributes

Flow graphs with fuzzy attributes have several advantages (e.g., represent data in form of fuzzy attributes) [22]. In further discussions, we apply flow graphs to represent and analyze decision table with fuzzy attributes in the form introduced in [22]. All possible decision rules with fuzzy attributes can be generated by using the *Cartesian product* of the set of attribute values with  $r = \prod_{i=1}^n n_i \prod_{j=1}^m m_j$  as a possible number of rules. Let  $R_k$  be the  $k$ -th decision rule. It can be expressed by

$R_k$ : IF  $c_1$  is  $V_1^k$ ,  $c_2$  is  $V_2^k$ , ... and  $c_n$  is  $V_n^k$ ,  
THEN  $d_1$  is  $W_1^k$ ,  $d_2$  is  $W_2^k$ , ... and  $d_m$  is  $W_m^k$ ,

where  $k = 1, 2, \dots, r$ , and  $V_i^k \in V_i$ ,  $i = 1, 2, \dots, n$ ,  $W_j^k \in W_j$ ,  $j = 1, 2, \dots, m$ .

Thus, Cartesian products  $C^k = V_1^k \times V_2^k \times \dots \times V_n^k$  and  $D^k = W_1^k \times W_2^k \times \dots \times W_m^k$  are the  $k$ -th decision rules expressed in the form of fuzzy implication,  $C^k \rightarrow D^k$ .

Next, we apply a  $T$ -norm operator to determine the *confirmation degree* of fuzzy decision rule. Let  $cda(x, k)$  and  $cdc(x, k)$  be the confirmation degrees of the antecedent and consequent of decision rules and  $cd(x, k)$  be the confirmation degrees of the  $k$ -th decision rules by the element  $x \in U$ . They are defined respectively as

$$cda(x, k) = T(\mu_{V_1^k}(x), \mu_{V_2^k}(x), \dots, \mu_{V_n^k}(x)) \quad (42)$$

$$cdc(x, k) = T(\mu_{W_1^k}(x), \mu_{W_2^k}(x), \dots, \mu_{W_m^k}(x)) \quad (43)$$

$$cd(x, k) = T(cda(x, k), cdc(x, k)). \quad (44)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

By using the introduced equations (42), (43) and (44), in fuzzy set domain, the *support* (*sup*) of decision rule's antecedent, consequent and itself are defined respectively as

$$\text{sup}(cda(x, k)) = \{cda(x_1, k)/x_1, \dots, cda(x_N, k)/x_N\},$$

$$\text{sup}(cdc(x, k)) = \{cdc(x_1, k)/x_1, \dots, cdc(x_N, k)/x_N\} \text{ and}$$

$$\text{sup}(R_k) = \{cd(x_1, k)/x_1, \dots, cd(x_N, k)/x_N\}.$$

Moreover, when consider fuzzy universe we can define the *throughflow*,  $\varphi(x, y)$  of the branch  $(x, y)$  in  $G$  to the cardinality of the product of fuzzy sets,  $\tilde{X}$  and  $\tilde{Y}$ . The *strength* of  $(x, y)$  is thus given by

$$\sigma(x, y) = \frac{\varphi(x, y)}{\varphi(G)}. \quad (45)$$

With every node  $x$  of a flow graph  $G$ , associated *normalized inflow* and *outflow* are defined respectively as

$$\varphi_-(x) = \text{power}\tilde{x} = \sum_{y \in O(x)} \varphi(x, y) = \sum_{y \in O(x)} \text{power}(\tilde{x} \cap \tilde{y}) \quad (46)$$

$$\varphi_+(x) = \text{power}\tilde{x} = \sum_{y \in I(x)} \varphi(x, y) = \sum_{y \in I(x)} \text{power}(\tilde{x} \cap \tilde{y}) \quad (47)$$

where  $\text{power}\tilde{x}$  denotes fuzzy cardinality. For any internal node  $x$ , it holds that  $\varphi_-(x) = \varphi_+(x) = \varphi(x)$ .

A (directed) *path* from  $x$  to  $y$  ( $x \neq y$ ) in  $G$ , denoted by  $[x \dots y]$ , is a sequence of nodes  $x_1, \dots, x_n$  such that  $x_1 = x$  and  $x_n = y$  and  $(x_i, x_{i+1}) \in B$ ,  $i, 1 \leq i \leq n - 1$ . We obtain the *certainty*, *coverage* and *strength* coefficients of the path  $[x_1 \dots x_n]$  [22, 23] as:  $\text{cer}[x_1 \dots x_n] = \prod_{i=1}^{n-1} \text{cer}(x_1 \dots x_i, x_{i+1})$ , where  $\text{cer}(x_1 \dots x_i, x_{i+1}) = \frac{\text{power}(\tilde{x}_1 \cap \tilde{x}_2 \cap \dots \cap \tilde{x}_{i+1})}{\text{power}(\tilde{x}_1 \cap \tilde{x}_2 \cap \dots \cap \tilde{x}_i)}$ ,  $\text{cov}[x_1 \dots x_n] = \prod_{i=1}^{n-1} \text{cov}(x_1 \dots x_i, x_{i+1})$ , where  $\text{cov}(x_1 \dots x_i, x_{i+1}) = \frac{\text{power}(\tilde{x}_1 \cap \tilde{x}_2 \cap \dots \cap \tilde{x}_{i+1})}{\text{power}(\tilde{x}_{i+1})}$ , and  $\sigma[x_1 \dots x_n] = \sigma(x_1) \text{cer}[x_1 \dots x_n] = \sigma(x_n) \text{cov}[x_1 \dots x_n]$ .

The decision rule  $R_k, \Phi^* \rightarrow \Phi_n$ , where  $\Phi^*$  is condition and  $\Phi_n$  is decision can be characterized by three coefficients:

$$\text{cer}(\Phi^*, \Phi_n) = \text{cer}(R_k) = \frac{\text{power}(\text{sup}(R_k))}{\text{power}(\text{sup}(cda(x, k)))} \quad (48)$$

$$\text{cov}(\Phi^*, \Phi_n) = \text{cov}(R_k) = \frac{\text{power}(\text{sup}(R_k))}{\text{power}(\text{sup}(cdc(x, k)))} \quad (49)$$

$$\sigma(\Phi^*, \Phi_n) = \text{strength}(R_k) = \frac{\text{power}(\text{sup}(R_k))}{\text{card}(U)}. \quad (50)$$

We will illustrate an example of introduced notions and experimental results in subsequence sections.

2.5 Illustrative Example

In this section, we present an example of how to compute coefficients: *certainty*, *coverage* and *strength* for flow graphs with fuzzy attributes. To compute these coefficients for flow graphs by the notion of rough set theory please refer to [3].

Table 2 shows a decision table  $S = (U, A)$ , where  $U = \{x_1, x_2, \dots, x_{10}\}$ , condition attributes are score of Number theory ( $N$ ) and score of Algebra ( $A$ ). The decision attribute is the total score ( $D$ ).

Table 2. An example decision table

	$N$	$A$	$D$
$x_1$	32	46	78
$x_2$	9	40	102
$x_3$	8	35	101
$x_4$	11	31	88
$x_5$	12	27	87
$x_6$	16	31	81
$x_7$	2	37	74
$x_8$	0	36	73
$x_9$	10	29	73
$x_{10}$	4	28	70

We next use triangular membership function below to the data in Table 2 fuzzification.

$\mu_{N_1} = \begin{cases} 1, & \text{if } 0 \leq x < 5 \\ (11-x)/6, & \text{if } 5 \leq x < 11 \\ 0, & \text{if } 11 \leq x \leq 45 \end{cases}$	$\mu_{N_2} = \begin{cases} 0, & \text{if } 0 \leq x < 5 \\ (x-5)/6, & \text{if } 5 \leq x < 11 \\ (40-x)/29, & \text{if } 11 \leq x < 40 \\ 0, & \text{if } 40 \leq x \leq 45 \end{cases}$
$\mu_{N_3} = \begin{cases} 0, & \text{if } 0 \leq x < 11 \\ (x-11)/29, & \text{if } 11 \leq x < 40 \\ 1, & \text{if } 40 \leq x \leq 45 \end{cases}$	$\mu_{A_1} = \begin{cases} 1, & \text{if } 0 \leq x < 20 \\ (30-x)/10, & \text{if } 20 \leq x < 30 \\ 0, & \text{if } 30 \leq x \leq 46 \end{cases}$
$\mu_{A_2} = \begin{cases} 0, & \text{if } 0 \leq x < 20 \\ (x-20)/10, & \text{if } 20 \leq x < 30 \\ (36-x)/6, & \text{if } 30 \leq x < 36 \\ 0, & \text{if } 36 \leq x \leq 46 \end{cases}$	$\mu_{A_3} = \begin{cases} 0, & \text{if } 0 \leq x < 30 \\ (x-30)/6, & \text{if } 30 \leq x < 36 \\ 1, & \text{if } 36 \leq x \leq 46 \end{cases}$
$\mu_{D_1} = \begin{cases} 1, & \text{if } 0 \leq x < 69 \\ (85-x)/16, & \text{if } 69 \leq x < 85 \\ 0, & \text{if } 85 \leq x \leq 195 \end{cases}$	$\mu_{D_2} = \begin{cases} 0, & \text{if } 0 \leq x < 69 \\ (x-69)/16, & \text{if } 69 \leq x < 85 \\ (110-x)/25, & \text{if } 85 \leq x < 110 \\ 0, & \text{if } 110 \leq x \leq 195 \end{cases}$
$\mu_{D_3} = \begin{cases} 0, & \text{if } 0 \leq x < 85 \\ (x-85)/25, & \text{if } 85 \leq x < 110 \\ 1, & \text{if } 110 \leq x \leq 195 \end{cases}$	

The result is a decision table with fuzzy attributes as shown in Table 3. Afterwards, we can compute strength  $\sigma(N_i, A_j)$  of condition attributes in Table 3 by equation (45).

Table 4 shows strength coefficient between condition attributes, last row depicts inflow of each  $A_j$ , and last column depicts outflow of each  $N_i$ . For example, since throughflow  $\varphi(N_1, A_1)$  is sum of the cardinality of product of fuzzy set  $\tilde{N}_1$  and  $\tilde{A}_1$ , we get  $N_1 A_1 = \tilde{N}_1 \cap \tilde{A}_1 = (0 \times 0) + (0.33 \times 0) + (0.5 \times 0) + (0 \times 0)$

**Table 3.** Decision table with fuzzy attributes

	$N_i$			$A_j$			$D_k$		
	$N_1$	$N_2$	$N_3$	$A_1$	$A_2$	$A_3$	$D_1$	$D_2$	$D_3$
$x_1$	0	0.276	0.724	0	0	1	0	0	1
$x_2$	0.333	0.667	0	0	0	1	0	0.320	0.680
$x_3$	0.500	0.500	0	0	0.167	0.833	0	0.360	0.640
$x_4$	0	1	0	0	0.833	0.167	0	0.880	0.120
$x_5$	0	0.965	0.035	0.300	0.700	0	0	0.920	0.080
$x_6$	0	0.828	0.172	0	0.833	0.167	0.250	0.750	0
$x_7$	1	0	0	0	0	1	0.688	0.312	0
$x_8$	1	0	0	0	0	1	0.750	0.250	0
$x_9$	0.167	0.833	0	0.100	0.900	0	0.75	0.250	0
$x_{10}$	1	0	0	0.200	0.800	0	0.938	0.062	0

**Table 4.** Strength  $\sigma(N_i, A_j)$  between condition and decision attribute

	$A_1$	$A_2$	$A_3$	$\sum$
$N_1$	0.02167	0.10333	0.27500	0.4000
$N_2$	0.03730	0.30322	0.16638	0.5069
$N_3$	0.00103	0.01678	0.07529	0.0931
$\sum$	0.06000	0.42330	0.51667	1

$(0 \times 0.3) + (0 \times 0) + (1 \times 0) + (1 \times 0) + (0.167 \times 0.1) + (1 \times 0.2) = 0.2167$ .  
 Then by (45),  $\sigma(N_1, A_1) = \frac{\varphi(N_1, A_1)}{10} = 0.02167$ .

**Table 5.** Strength  $\sigma(A_j, D_k)$  between condition and decision attribute

	$D_1$	$D_2$	$D_3$	$\sum$
$A_1$	0.02625	0.03135	0.00240	0.06
$A_2$	0.16333	0.23373	0.02627	0.42333
$A_3$	0.14792	0.14542	0.22333	0.51667
$\sum$	0.33750	0.41050	0.2520	1

Next, we can use (45) and (46) to compute the strength of inflow of  $A_1$  and obtain  $\sigma_-(A_1) = \frac{power(\tilde{A}_1)}{N} = \sum_{i=1}^3 \sigma(N_i, A_1) = \sigma(N_1, A_1) + \sigma(N_2, A_1) + \sigma(N_3, A_1) = 0.02167 + 0.0373 + 0.00103 = 0.06$ .

Table 5 shows strength  $\sigma(A_j, D_k)$  between condition and decision attributes last row depicts inflow of each  $D_k$  and last column depicts outflow of each  $A_j$ . If we use (45) and (47) to compute the strength of outflow of  $A_1$ , we obtain  $\sigma_+(A_1) = \frac{power(\tilde{A}_1)}{N} = \sum_{j=1}^3 \sigma(A_j, D_1) = \sigma(A_1, D_1) + \sigma(A_2, D_1) + \sigma(A_3, D_1) = 0.02625 + 0.03135 + 0.0024 = 0.06$  (first row in last column) (equal to  $\sigma_-(A_1)$ ). Moreover, last row from Table 4 and last column from Table 5 show that  $\sigma_-(A_j) = \sigma_+(A_j)$ , where  $j = 1, 2, 3$ .

**Table 6.** Certainty and Coverage coefficients of  $N_i$  and  $A_j$ 

	$A_1$		$A_2$		$A_3$	
	<i>cer</i>	<i>cov</i>	<i>cer</i>	<i>cov</i>	<i>cer</i>	<i>cov</i>
$N_1$	0.0542	0.3612	0.2583	0.2440	0.6875	0.5322
$N_2$	0.0736	0.6217	0.5982	0.7163	0.3282	0.3220
$N_3$	0.0111	0.0171	0.1802	0.0397	0.8087	0.1458

Table 6 shows the certainty and coverage coefficients of  $N_i$  and  $A_j$ , where  $i, j = 1, 2, 3$ . They are computed by (48), as  $cer(N_i, A_j) = \frac{power(\tilde{N}_i \cap \tilde{A}_j)}{power(\tilde{N}_i)} = \frac{\sigma(N_i, A_j)}{\sigma(N_i)}$  and by (49) as  $cov(N_i, A_j) = \frac{power(\tilde{N}_i \cap \tilde{A}_j)}{power(A_j)} = \frac{\sigma(N_i, A_j)}{\sigma(A_j)}$ . For example,  $cer(N_1, A_1) = \frac{\sigma(N_1, A_1)}{\sigma(N_1)} = \frac{0.02167}{0.4} = 0.0542$  (in top left cell) and  $cov(N_1, A_1) = \frac{\sigma(N_1, A_1)}{\sigma(A_1)} = \frac{0.02167}{0.06} = 0.3612$  (next cell).

**Table 7.** Certainty and Coverage coefficients of  $A_i$  and  $D_k$ 

	$D_1$		$D_2$		$D_3$	
	<i>cer</i>	<i>cov</i>	<i>cer</i>	<i>cov</i>	<i>cer</i>	<i>cov</i>
$A_1$	0.4375	0.0778	0.5225	0.0764	0.0400	0.0095
$A_2$	0.3858	0.4840	0.5521	0.5694	0.0620	0.1042
$A_3$	0.2863	0.4383	0.2815	0.3542	0.4323	0.8862

Table 7 shows certainty and coverage coefficients of  $A_j$  and  $D_k$ , where  $j, k = 1, 2, 3$ . They are computed by the same approach as in Table 6.

**Table 8.** Value of normalized strength between condition and decision attributes

	$D_1$	$D_2$	$D_3$	$\sum$
$N_1 A_1$	0.0200	0.0017	0	0.02167
$N_1 A_2$	0.0863	0.0118	0.0053	0.10333
$\vdots$				$\vdots$
$N_3 A_3$	0.0007	0.0022	0.0724	0.07529
$\sum$	0.3375	0.4105	0.2520	1

Table 8 shows the value of normalized strength between condition and decision attributes,  $\sigma(N_i, A_j, D_k)$  where  $i, j, k = 1, 2, 3$  by (50). Last row depicts inflow of  $D_k$  and last column depicts outflow of  $N_i A_j$ .

Next, from (45), (48) and (49) we can find certainty and coverage coefficients of  $N_i, A_j$  and  $D_k$ . The results are shown in Table 9, e.g.,  $cer(N_1 A_1, D_1) = \frac{0.02}{0.02167} = 0.9229$  (top left cell) and  $cov(N_1 A_1, D_1) = \frac{0.02}{0.3375} = 0.0593$  (next cell).

Table 10 shows certainty and coverage coefficients of  $N_i$  and  $D_k$ , where  $i, k = 1, 2, 3$ . They are computed by the same approach as in Table 6.

**Table 9.** Certainty and Coverage coefficients of  $N_i, A_j$  and  $D_k$

	$D_1$		$D_2$		$D_3$	
	<i>cer</i>	<i>cov</i>	<i>cer</i>	<i>cov</i>	<i>cer</i>	<i>cov</i>
$N_1 A_1$	0.9229	0.0593	0.0769	0.0041	0	0
$N_1 A_2$	0.8347	0.2556	0.1137	0.0286	0.0516	0.0212
$\vdots$						$\vdots$
$N_3 A_3$	0.0095	0.0021	0.0286	0.0053	0.9618	0.2874

**Table 10.** Certainty and Coverage coefficients of  $N_i$  and  $D_k$

	$D_1$		$D_2$		$D_3$	
	<i>cer</i>	<i>cov</i>	<i>cer</i>	<i>cov</i>	<i>cer</i>	<i>cov</i>
$N_1$	0.6250	0.7407	0.2383	0.2322	0.1367	0.2169
$N_2$	0.1641	0.2465	0.5900	0.7285	0.2459	0.4946
$N_3$	0.0463	0.0128	0.173	0.0392	0.7807	0.2885

**Table 11.** Certainty and Coverage coefficients of  $N_i \vee A_j$  and  $D_k$

	$D_1$		$D_2$		$D_3$	
	<i>cer</i>	<i>cov</i>	<i>cer</i>	<i>cov</i>	<i>cer</i>	<i>cov</i>
$N_1 \vee A_1$	$\leq 1.0000$	0.7407	$\leq 0.7608$	0.3045	$\leq 0.1767$	0.2264
$N_1 \vee A_2$	$\leq 1.0000$	0.9691	$\leq 0.7904$	0.7730	$\leq 0.1987$	0.2995
$\vdots$						$\vdots$
$N_3 \vee A_3$	$\leq 0.3326$	0.4490	$\leq 0.4545$	0.4545	$\leq 1.0000$	0.8873

Finally, if we want to relax mutual exclusion of condition, then from Propositions 3 and 4 we can find  $cer(N_i \vee A_j, D_k)$  and  $cov(N_i \vee A_j, D_k)$ , as shown in Table 11. For example,  $cer(N_1 \vee A_2, D_2) < cer(N_1, D_2) + cer(A_2, D_2) = 0.2383 + 0.5521 = 0.7904$  and  $cov(N_1 \vee A_2, D_2) = cov(N_1, D_2) + cov(A_2, D_2) - cov(N_1 \wedge A_2, D_2) = 0.2322 + 0.5694 - 0.0286 = 0.7730$ .

From Table 11, we can use Propositions 3 and 4 to find certainty and coverage coefficients for relax mutual exclusion of condition.

## 2.6 Inverse Flow Graphs

A flow graph is an alternative representation of knowledge in the given data in terms of certainty, coverage and strength between ordered pairs of inputs and outputs.

Alternatively, if we focus on the relations between ordered pairs of outputs and inputs, then we consider an *inverse flow graph*. Thus, in this section, we present inverse flow graphs and three coefficients between flow graphs and their inverse flow graphs.

**Definition 1** [9] Let  $G = (N, B, \sigma)$  be a flow graph. If we invert direction of all branches in  $G$ , then the resulting graph  $G^{-1} = (N, B', \sigma')$  will be called an inverted graph of  $G$ , where  $N$  is a set of *nodes*,  $B'$  is a set of *inverse directed branches* and  $\sigma'$  is a normalized flow function.

From this definition, we see that an inverse flow graph also forms a flow graph. Inputs and outputs of flow graph  $G$  are outputs and inputs of flow graph  $G^{-1}$ , respectively. For this reason, an inverted graph of  $G$  is also referred to as an *inverse flow graph*.

Next, we prove the subsequent properties of this inverse flow graph.

**Proposition 5** Let  $G$  be a flow graph and  $G^{-1}$  be its inverse flow graph. For any inverse flow graph  $G^{-1}$ , certainty, coverage and strength are holded as the following.

1.  $\sigma_{G^{-1}}(y, x) = \sigma_G(x, y)$ ,
2.  $cov_{G^{-1}}(y, x) = cer_G(x, y)$ ,
3.  $cer_{G^{-1}}(y, x) = cov_G(x, y)$ .

*Proof.* Let  $(x, y)$  be branch of flow graph  $G$ , then  $\sigma_G(x, y) = \frac{\varphi(x, y)}{\varphi(G)}$ ,  $cer_G(x, y) = \frac{\sigma(x, y)}{\sigma(x)}$  and  $cov_G(x, y) = \frac{\sigma(x, y)}{\sigma(y)}$ .

From Definition 1, if  $(y, x)$  is a branch in an inverted graph of  $G$ , then  $\sigma_{G^{-1}}(y, x) = \frac{\varphi(y, x)}{\varphi(G)}$ ,  $cov_{G^{-1}}(y, x) = \frac{\sigma(y, x)}{\sigma(x)}$  and  $cer_{G^{-1}}(y, x) = \frac{\sigma(y, x)}{\sigma(y)}$ .

Thus,  $\sigma_G(x, y) = \sigma_{G^{-1}}(y, x)$ ,  $cer_G(x, y) = cov_{G^{-1}}(y, x)$  and  $cov_G(x, y) = cer_{G^{-1}}(y, x)$ .

**Proposition 6** Let  $G$  be a flow graph and  $G^{-1}$  be its inverse flow graph. For any inverse flow graph  $G^{-1}$ , certainty, coverage and strength are the following.

1.  $cov_{G^{-1}}[x_n \dots x_1] = cer_G[x_1 \dots x_n]$ ,
2.  $cer_{G^{-1}}[x_n \dots x_1] = cov_G[x_1 \dots x_n]$ ,
3.  $\sigma_{G^{-1}}[x_n \dots x_1] = \sigma_G[x_1 \dots x_n]$ .

*Proof.* Let  $[x_1, \dots, x_n]$  be a path from  $x_1$  to  $x_n$  in a flow graph  $G$ , then

$$cer_G[x_1 \dots x_n] = \prod_{i=1}^{n-1} cer_G(x_i, x_{i+1}),$$

$$cov_G[x_1 \dots x_n] = \prod_{i=1}^{n-1} cov_G(x_i, x_{i+1}) \text{ and}$$

$$\sigma_G[x \dots y] = \sigma(x)cer_G[x \dots y] = \sigma(y)cov_G[x \dots y].$$

From Definition 1, if  $[x_n, \dots, x_1]$  is path in the corresponding inverse flow graph  $G^{-1}$ , then

$$cer_{G^{-1}}[x_n \dots x_1] = \prod_{i=1}^{n-1} cer_{G^{-1}}(x_{n+1-i}, x_{n-i}),$$

$$cov_{G^{-1}}[x_n \dots x_1] = \prod_{i=1}^{n-1} cov_{G^{-1}}(x_{n+1-i}, x_{n-i}) \text{ and}$$

$$\sigma[x_n \dots x_1] = \sigma(x_n)cer_{G^{-1}}[x_n \dots x_1] = \sigma(x_1)cov_{G^{-1}}[x_n \dots x_1].$$

From Proposition 5, we obtain

$$cer_G[x_1, \dots, x_n] = cov_{G^{-1}}[x_n, \dots, x_1],$$

$$cov_G[x_1, \dots, x_n] = cer_{G^{-1}}[x_n, \dots, x_1] \text{ and}$$

$$\sigma_G[x_1, \dots, x_n] = \sigma_{G^{-1}}[x_n, \dots, x_1].$$

**Proposition 7** Let  $G$  be a flow graph and  $G^{-1}$  be its inverse flow graph. For any inverse flow graph  $G^{-1}$ , certainty, coverage and strength are holded as the following.

1.  $cov_{G^{-1}}(x_n, x_1) = cer_G(x_1, x_n)$ ,
2.  $cer_{G^{-1}}(x_n, x_1) = cov_G(x_1, x_n)$ ,
3.  $\sigma_{G^{-1}}(x_n, x_1) = \sigma_G(x_1, x_n)$ .

*Proof.* Let  $\langle x_1, x_n \rangle$  be a connection from  $x_1$  to  $x_n$  in a flow graph  $G$ , then

$$cer\langle x_1, x_n \rangle = \sum_{[x_1 \dots x_n] \in \langle x_1, x_n \rangle} cer[x_1 \dots x_n],$$

$$cov\langle x_1, x_n \rangle = \sum_{[x_1 \dots x_n] \in \langle x_1, x_n \rangle} cov[x_1 \dots x_n] \text{ and}$$

$$\begin{aligned} \sigma\langle x_1, x_n \rangle &= \sum_{[x_1 \dots x_n] \in \langle x_1, x_n \rangle} \sigma[x_1 \dots x_n] \\ &= \sigma(x_1)cer\langle x_1, x_n \rangle = \sigma(x_n)cov\langle x_1, x_n \rangle. \end{aligned}$$

Since a connection  $\langle x_1, x_n \rangle$  contains all of path from  $x_1$  to  $x_n$  and from Definition 1, if  $[x_1 \dots x_n]$  is a path in flow graph  $G$ , then  $[x_n \dots x_1]$  is a path in its corresponding inverse flow graph  $G^{-1}$ . Then  $\langle x_n, x_1 \rangle$  is an inverse of a connection  $\langle x_1, x_n \rangle$  in  $G$ , and hence

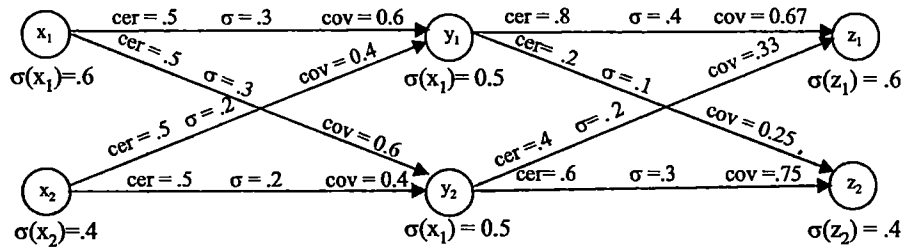
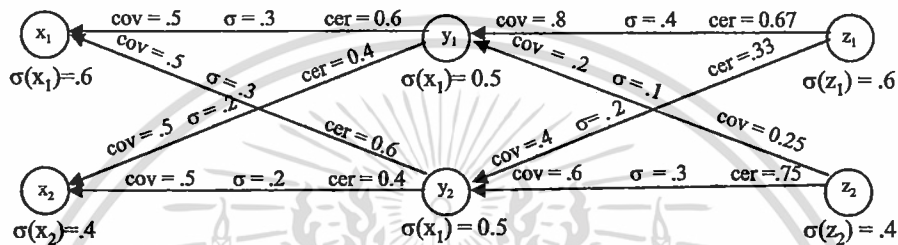
$$cer\langle x_n, x_1 \rangle = \sum_{[x_n \dots x_1] \in \langle x_n, x_1 \rangle} cer[x_n \dots x_1],$$

$$cov\langle x_n, x_1 \rangle = \sum_{[x_n \dots x_1] \in \langle x_n, x_1 \rangle} cov[x_n \dots x_1] \text{ and}$$

$$\begin{aligned} \sigma\langle x_n, x_1 \rangle &= \sum_{[x_n \dots x_1] \in \langle x_n, x_1 \rangle} \sigma[x_n \dots x_1] \\ &= \sigma(x_n)cer\langle x_n, x_1 \rangle = \sigma(x_1)cov\langle x_n, x_1 \rangle. \end{aligned}$$

From Proposition 6, we obtain  $cer_G\langle x_1, x_n \rangle = cov_{G^{-1}}\langle x_n, x_1 \rangle$ ,  $cov_G\langle x_1, x_n \rangle = cer_{G^{-1}}\langle x_n, x_1 \rangle$  and  $\sigma_G\langle x_1, x_n \rangle = \sigma_{G^{-1}}\langle x_n, x_1 \rangle$ .

**Example 1.** Suppose we are given the flow graph  $G$  in Fig. 6, the inputs are  $x_1$  and  $x_2$ , the outputs are  $z_1$  and  $z_2$  and the internal nodes are  $y_1$  and  $y_2$ . Every branch in the flow graph  $G$  is associated with three coefficients as shown in Fig. 6. If we invert direction of all branches in  $G$ , then by Definition 1 we obtain the inverse flow graph  $G^{-1}$  and its certainty, coverage and strength as

Fig. 6. Flow graph  $G$ .Fig. 7. Inverse flow graph  $G^{-1}$ .

depicted in Fig. 7. From Fig. 6 and Fig. 7 and their branches' directions, we see that Proposition 5 holds.

Next, if we consider every path in the flow graph  $G$  and every path in the inverse flow graph  $G^{-1}$ , then Proposition 6 is also satisfied. For example, consider path  $x_1 y_1 z_1$  in Fig. 6, we can calculate  $cer_G[x_1, y_1 z_1] = 0.40$ ,  $cov_G[x_1, y_1 z_1] = 0.402$  and  $\sigma_G[x_1, y_1 z_1] = 0.24$ . Then for the path  $z_1 y_1 x_1$  in the inverse flow graph in Fig. 7, we can calculate  $cov_{G^{-1}}[z_1, y_1 x_1] = 0.40$ ,  $cer_{G^{-1}}[z_1, y_1 x_1] = 0.402$  and  $\sigma_{G^{-1}}[z_1, y_1 x_1] = 0.24$ .

Similarly, if we consider all connections in the flow graph  $G$  and its inverse flow graph  $G^{-1}$ , then we see that Proposition 7 is also satisfied. For instance, consider the connection  $\langle x_1, z_1 \rangle$  in the flow graph  $G$  we can calculate  $cer_G\langle x_1, z_1 \rangle = 0.6$ ,  $cov_G\langle x_1, z_1 \rangle = 0.6$  and  $\sigma_G\langle x_1, z_1 \rangle = 0.36$ . Then for the connection  $\langle z_1, x_1 \rangle$  in the inverse flow graph  $G^{-1}$  we can calculate  $cov_{G^{-1}}\langle z_1, x_1 \rangle = 0.6$ ,  $cer_{G^{-1}}\langle z_1, x_1 \rangle = 0.6$  and  $\sigma_{G^{-1}}\langle z_1, x_1 \rangle = 0.36$ .

## 2.7 Experimental Results

In this section, we describe experimental results from the scores of Olympiad mathematics' students' analysis. Data contains 75 records collected from the Promotion of Academic Olympiads and Development of Science Education Foundation (POSN) in 2006–2007. In this paper, we study decision rules between conditions and decision as shown below:

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Condition	Symbol	Condition	Symbol
score of Geometry	<i>score_Geo</i>	preference of Algebra	<i>Pref_Alg</i>
score of Number theory	<i>score_Num</i>	preference of Combinatorics	<i>Pref_Com</i>
score of Algebra	<i>score_Alg</i>	preference of Inequality	<i>Pref_I.E.</i>
score of Combinatorics	<i>score_Com</i>	practiced of Geometry	<i>Prac_Geo</i>
score of Inequality	<i>score_I.E.</i>	practiced of Number theory	<i>Prac_Num</i>
preference of Geometry	<i>Pref_Geo</i>	practiced of Algebra	<i>Prac_Alg</i>
preference of Number theory	<i>Pref_Num</i>	practiced of Combinatorics	<i>Prac_Com</i>
practiced of Inequality	<i>Prac_I.E.</i>		
Decision	Symbol		
student total score	<i>total_score</i>		

Next, in Cases 1–2, we consider attributes in POSN data set as descriptive attributes. Possible values of first five condition attributes and decision attribute are low score and high score. Several case studies are then presented according to the combination of these attributes. The threshold for decision rule extraction is  $cer > 0.6$ .

**Case 1.** Decision rules between score of any subject and total score. The selected decision rules are:

```

IF Score_Num is low THEN total_score is low, cer = 0.867,
cov = 0.812,4
IF Score_Geo is low THEN total_score is low, cer = 0.792,
cov = 0.775,
.
.
IF Score_I.E. is high THEN total_score is high, cer = 0.956,
cov = 0.814,

```

with *average cer* = 0.791 and *average cov* = 0.731.

**Case 2.** Decision rules between score of any two subjects and total score. Then the selected decision rules are:

```

IF score_Geo and I.E. are low
THEN total_score is low, cer = 0.633, cov =
0.581,
IF score_Geo and Alg are low
THEN total_score is low, cer = 0.622, cov =
0.571,
.
.
IF score_Alg and I.E. are low
THEN total_score is high, cer = 0.625, cov =
0.523,

```

with *average cer* = 0.620 and *average cov* = 0.556.

From Cases 1–2, we can summarize that, *if score of one or two subjects are low, then total score tends to be low too and if score of one or two subjects are high, then total score tends to be high too*. We see that decision rules obtain from Cases 1–2 depend on only two possible values of attributes (high and low). Thus, we next consider cases that scores are high and low with their degrees which are fuzzy attributes.

<sup>4</sup> Due to round off errors in computations, the equation may be not satisfied exactly in our examples.

Next, in Cases 3–8, we transform POSN data set to be in fuzzy attributes form by triangular membership function. Several case studies are then presented according to the combination of these attributes. The threshold for decision rule extraction is  $cer > 0.6$ .

**Case 3.** Decision rules between score of any subject and total score. Then the selected decision rule is:

<p>IF <i>Score_Num</i> is high THEN <i>total_score</i> is high, <math>cer = 0.627</math>, <math>cov = 0.504</math></p>
--

with  $cer = 0.627$  and  $cov = 0.504$ .

**Case 4.** Decision rules between score of any two subjects and total score. Then the selected decision rules are:

<p>IF <i>score_Geo</i> and <i>I.E.</i> are low THEN <i>total_score</i> is low, <math>cer = 0.651</math>, <math>cov = 0.409</math>, IF <i>score_Geo</i> and <i>I.E.</i> are high THEN <i>total_score</i> is high, <math>cer = 0.68</math>, <math>cov = 0.301</math>, . . IF <i>score_Alg</i> and <i>I.E.</i> are high THEN <i>total_score</i> is high, <math>cer = 0.618</math>, <math>cov = 0.226</math>,</p>
---

with average  $cer = 0.656$  and average  $cov = 0.343$ .

**Case 5.** Decision rules between score of any five subjects and total score. Then the selected decision rules are:

<p>IF <i>score_Geo</i>, <i>Num</i>, <i>Alg</i>, <i>Com</i> and <i>score_I.E.</i> are low THEN <i>total_score</i> is low, <math>cer = 0.782</math>, <math>cov = 0.139</math>, IF <i>score_Geo</i>, <i>Num</i>, <i>Alg</i>, and <i>Com</i> are low, <i>I.E.</i> is high THEN <i>total_score</i> is low, <math>cer = 0.645</math>, <math>cov = 0.09</math>, . . IF <i>score_Geo</i>, <i>Num</i>, <i>Alg</i>, <i>Comb</i> and <i>I.E.</i> are high THEN <i>total_score</i> is high, <math>cer = 0.81</math>, <math>cov = 0.024</math>,</p>
--

with average  $cer = 0.691$  and average  $cov = 0.047$ .

**Case 6.** Decision rules between practiced and preferred subjects, score of practiced and preferred subject and total score. The selected decision rules are:

<p>IF <i>Pref</i> and <i>Prac_Geo</i> and <i>score_Geo</i> is low THEN <i>total_score</i> is low, <math>cer = 0.638</math>, <math>cov = 0.719</math>, IF <i>Pref</i> and <i>Prac_Num</i> and <i>score_Num</i> is high THEN <i>total_score</i> is high, <math>cer = 0.834</math>, <math>cov = 0.6</math>,</p>
--

with *average cer* = 0.736 and *average cov* = 0.660.

**Case 7.** Decision rules between practiced subjects, score of practiced subjects and total score. The selected decision rules are:

<p>IF <i>Prac_Geo</i> and <i>score_Geo</i> is low          THEN <i>total_score</i> is low, <i>cer</i> = 0.602, <i>cov</i> = 0.731,          IF <i>Prac_I.E.</i> and <i>score_I.E.</i> is low          THEN <i>total_score</i> is low, <i>cer</i> = 0.665, <i>cov</i> = 0.623,</p>
---

with *average cer* = 0.634 and *average cov* = 0.677.

**Case 8.** Decision rules between preferred subjects, score of preferred subjects and total score. The selected decision rules are:

<p>IF <i>Pref_Geo</i> and <i>score_Geo</i> is high  <i>total_score</i> is high, <i>cer</i> = 0.671, <i>cov</i> = 0.541,          IF <i>Pref_Num</i> and <i>score_Num</i> is high          THEN <i>total_score</i> is high, <i>cer</i> = 0.633, <i>cov</i> = 0.602,          IF <i>Pref_I.E.</i> and <i>score_I.E.</i> is low          THEN <i>total_score</i> is low, <i>cer</i> = 0.829, <i>cov</i> = 0.72,          IF <i>Pref_I.E.</i> and <i>score_I.E.</i> is high          THEN <i>total_score</i> is low, <i>cer</i> = 0.831, <i>cov</i> = 0.28,</p>
---

with *average cer* = 0.741 and *average cov* = 0.536.

From these selected decision rules, we can distil useful knowledge from this flow graph as follows. Number theory affects total score substantially. Because in Case 3, 62.7% of students with high score of Number theory also have high total score. Moreover, 50.4% of students which have high total score are from the group in which Number theory score is high. Obviously, students who get high score from any two (or more) subjects will successively get high total score as shown in Cases 4 and 5.

Interestingly, the combinations of either preferred subject or practiced subject and high score for such subject resulted in high total scores from Cases 6, 7 and 8. In Case 8, we obtain maximum certainty of selected decision rules as one can see in Fig. 1. The results are desirable values of *cer* and *cov*.

Finally, in Case 9 we relax mutual exclusion of condition and used Proposition 3 to find coverage coefficients of any one or two subjects and total score. The threshold for decision rule extraction is  $cov > 0.6$ .

**Case 9.** Decision rules between one or two subjects and total score. The selected decision rules are:

<p>IF <i>score_Geo</i> or <i>I.E.</i> are low THEN <i>total_score</i> is low, <i>cov</i> = 0.895,          IF <i>score_Geo</i> or <i>Alg</i> are low THEN <i>total_score</i> is low, <i>cov</i> = 0.888,          .          .          IF <i>score_Com</i> or <i>I.E.</i> are high THEN <i>total_score</i> is high, <i>cov</i> = 0.758,</p>
--

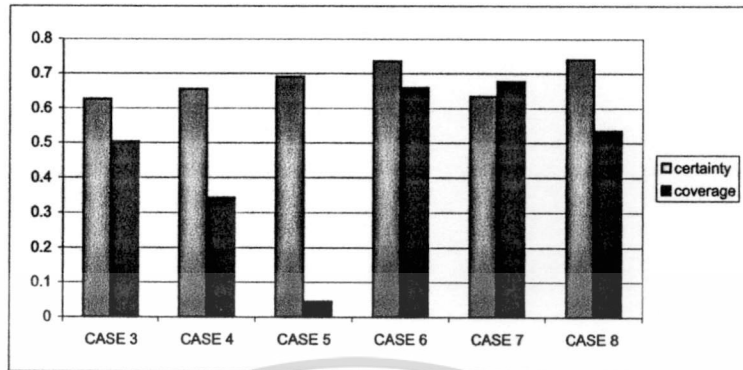


Fig. 8. Comparison of *cer*, *cov* of POSN data.

with average *cov* = 0.754.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3 Suggestions and Extensions of Flow Graphs' Research

As we mentioned above, a flow graph shows the information flow which can be useful for illustration and knowledge-based decision support. Nevertheless, the following are some possible extensions to enhance the utilities of flow graphs.

**Flow graphs representation** New representations of nodes, branches and their associate coefficients (e.g., *cer*, *cov*) are required to be able to interpret and understand flow graphs easier. Some attempts were in [11, 12].

**Dynamic flow graphs** Like rough set theory, flow graphs are generated from a finite (fixed) data set. Thus, if the data set is changes, we have to create an entirely new flow graph. Dynamic approaches to update, delete or add some new information are open issues.

**Hybridization** Several combinations and strong relationships of flow graphs and other pertinent techniques were studied by Pawlak et al. [2, 4, 5, 8–10, 15, 16] and others [6, 7, 11–14, 24]. Further interesting hybridization of flow graphs is likely still possible, e.g., to optimization.

**Flow graphs software** In order to generate flow graphs and provide interactive results to user, algorithms and software implementation that consider scalability and complexity are necessary.

## 4 Conclusion

In this research, an extension of rough set theory, namely flow graphs, is considered. It is a knowledge representation in form of information flow. In addition surveying the research in this area, we also performed an experiment on realword data to demonstrate the utility of flow graphs. Further, we presented several mathematical properties to reveal the relationship between flow graphs and association rules. New certainty and coverage coefficients are introduced. They are successively applied to POSN data set efficiency.

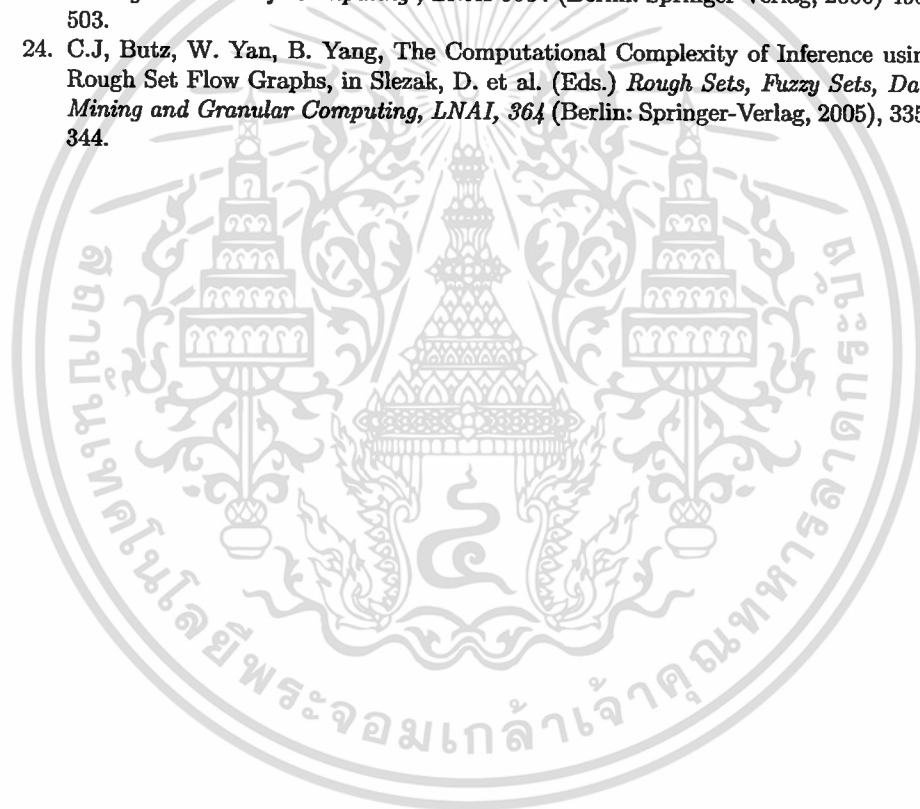


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## References

1. Pawlak, Z.: *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht (1991)
2. Pawlak, Z.: *Rough Set and Flow Graphs*. In: Slezak, D. (et al.) (Eds): *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. Lecture Notes in Artificial Intelligence*, Vol. 3641. Springer-Verlag, Berlin Heidelberg (2005) 1–11
3. Ford, L.R., Fulkerson, D.R.: *Flows in Networks*. Princeton University Press, Princeton. New Jersey (1962)
4. Pawlak, Z.: *Decision Algorithms and Flow Graphs: A Rough Set Approach*. *Journal of Telecommunications and Information Technology*. Vol. 3. (2003) 98–101
5. Pawlak, Z.: *Some Remarks on Conflict Analysis*. *European Journal of Operational Research*. Vol. 166. (2005) 649–654
6. Pattaraintakorn, P., Cercone, N., Naruedomkul, K.: *Rule Learning: Ordinal Prediction based on Rough Set and Soft-computing*. *Applied Mathematics Letters*. Vol. 19. (12) (2006) 1300–1307
7. Greco, S., Pawlak, Z., Slowinski, R.: *Generalized Decision Algorithms, Rough Inference Rules, and Flow Graphs*. In: Alpigini, J.J. (et al.) (Eds): *Rough Sets and Current Trends in Computing. Lecture Notes in Artificial Intelligence*, Vol. 2475. Springer-Verlag, Berlin Heidelberg (2002) 93–104
8. Pawlak, Z.: *Probability, Truth and Flow Graphs*. In: *Proceedings of Rough Sets in Knowledge Discovery and Soft Computing*, Warsaw, Poland (2003) 1–9
9. Pawlak, Z.: *Flow Graphs and Data Mining*. In: Peters, J.F., Skowron, A. (Eds.): *Transactions on Rough Sets III. Lecture Notes in Computer Science*, Vol. 3400. Springer-Verlag, Berlin Heidelberg (2005) 1–36
10. Pawlak, Z.: *Decision Trees and Flow Graphs*. In: Greco, S. (Eds.): *Rough Sets and Current Trends in Computing. Lecture Notes in Artificial Intelligence*, Vol. 4259. Springer-Verlag, Berlin Heidelberg (2006) 1–11
11. Sun, J., Liu, H., Qi, C., Zhang H.: *An Interpretation of Flow Graphs by Granular Computing*. In: Greco, S. (Eds.): *Rough Sets and Current Trends in Computing. Lecture Notes in Computer Science*, Vol. 4259. Springer-Verlag, Berlin Heidelberg (2006) 448–457
12. Sun, J., Liu, H., Zhang H.: *An Extension of Pawlak's Flow Graphs*. *Information Science*. Vol. 177. (2007) 28–40
13. Czdrzej, A., Kostek, B.: *Musical Metadata Retrieval with Flow Graphs*. In: Tsumoto, S. (et al.) (Eds.): *Rough Sets and Current Trends in Computing. Lecture Notes in Artificial Intelligence*, Vol. 3066. Springer-Verlag, Berlin Heidelberg (2004) 691–698
14. Kostek, B., Czdrzej, A.: *Processing of Musical Metadata Employing Pawlak's Flow Graphs*. In: Peters, J.F., Skowron, A. (Eds.): *Transactions on Rough Sets I. Lecture Notes in Artificial Intelligence*, Vol. 3100. Springer-Verlag, Berlin Heidelberg (2004) 279–298
15. Pawlak, Z.: *Flow Graphs and Decision Algorithms*. In: Wang, G. (et al.) (Eds.): *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. Lecture Notes in Artificial Intelligence*, Vol. 2639. Springer-Verlag, Berlin Heidelberg (2003) 1–10
16. Pawlak, Z.: *Flow Graphs and Intelligent Data Analysis*. *Fundamenta Informaticae*. Vol. 64. (1–4) (2005) 369–377
17. Agrawal, R., Imielinski, T., Swami, A.: *Mining Association Rule between Sets of Items in Large Database*. In *Proceeding of the ACM SIGMOD Conference*, Wasington DC, USA (1993)

18. Srikant, R., Agrawal, R.: Mining Generalized Association Rules. In: Proceeding of the 21st International Conference on Very Large Data Bases, Zurich, Switzerland (1995) 407–419
19. Butz, C.J., Yan, W., Yang, B.: The Computational Complexity of Inference using Rough Set Flow Graphs. In: Slezak, D. (et al.) (Eds.): Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. Lecture Notes in Artificial Intelligence, Vol. 364. Springer-Verlag, Berlin Heidelberg (2005) 335–344
20. Z. Pawlak, “Rough Sets, Decision Algorithms and Bayes’ Theorem.” *Eur. J. of Oper. Res.* Vol. 136, 2002, pp 181–189.
21. L.A. Zadeh, Fuzzy set. *Internation and Control* 8, 1965, 338–353
22. M. Alicja, M. Leszek, Flow Graph and Decision Tables with Fuzzy Attribdtes, *International Conference Artificial Intelligence and Soft Computing , LNAI 4029* (Berlin: Springer-Verlag, 2006) 268–277.
23. M. Alicja, M. Leszek, Fuzzy Implication Operators in Variable Precision Fuzzy Rough Sets Model, in L. Rutkowski et al. (eds.) *International Conference Artificial Intelligence and Soft Computing , LNAI 3037* (Berlin: Springer-Verlag, 2006) 498–503.
24. C.J, Butz, W. Yan, B. Yang, The Computational Complexity of Inference using Rough Set Flow Graphs, in Slezak, D. et al. (Eds.) *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, LNAI, 364* (Berlin: Springer-Verlag, 2005), 335–344.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ภาคผนวก

### ประโยชน์ทางการประยุกต์และการเผยแพร่ผลงานวิจัย

ในช่วง 12 เดือนที่ผ่านมาได้มีบทความวิชาการที่อยู่ระหว่างการพิจารณาให้ตีพิมพ์ในวารสารวิชาการระดับนานาชาติ คือ

Doungnat Chitcharone, Puntip Pattaraintakorn: "Novel Matrix Forms of Flow Graphs to Multi-Datasets Applications". (submitted).

นอกจากนี้ ในช่วง 12 เดือนที่ผ่านมาได้มีการนำเสนอผลงานวิจัยแบบ Oral presentation ในงานประชุมวิชาการในระดับนานาชาติ ดังนี้

Doungnat Chitcharone, Puntip Pattaraintakorn: "Knowledge Discovery by Rough Sets Mathematical Flow Graphs and its Extension". *The IASTED International Conference on Artificial Intelligence and Applications, (AIA'2008)*, Innsbruck, Austria (2008) 340-345.

Doungnat Chitcharone, Puntip Pattaraintakorn: "*Towards Theories of Fuzzy Set and Rough Set to Flow Graphs*". The 2008 IEEE World Congress on Computational Intelligence (WCCI 2008), Hong Kong (2008) 1675-1682.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้