

```

addchoice(X,Res):- length(X,Len),
                  numlist(1,Len,NumL),
                  map(NumL,X,Res).
map([], [], [99-exitShell]).
map([H|T], [X|TT], [H-X|T1]) :- map(T, TT, T1).

```

ตัวอย่างการใช้งานระบบผู้เชี่ยวชาญเพื่อขอคำแนะนำเกี่ยวกับใช้คอนแท็กเลนส์ แสดงได้ดังรูปที่ 5.6 เมื่อรัน โปรแกรม expertshell1.pl ผู้ใช้จะเริ่มต้นใช้งาน expert system shell ด้วยการพิมพ์คำสั่ง expertshell (ทุกคำสั่งในภาษาปรัลล็อกจะต้องจบคำสั่งด้วยเครื่องหมาย '.') เมื่อ expert system shell เริ่มทำงาน ที่ต้นบรรทัดจะปรากฏข้อความ expert-shell> เพื่อเตรียมรับคำสั่งจากผู้ใช้ คำสั่งแรกของการใช้งานคือคำสั่ง load เพื่อเรียกใช้ไฟล์ฐานความรู้ที่เกี่ยวข้อง ซึ่งในตัวอย่างนี้ใช้ไฟล์ '1.knb' จากนั้นใช้คำสั่ง solve โปรแกรมจะเริ่มถามข้อมูลต่างๆจากผู้ใช้ เมื่อได้ข้อมูลที่ต้องการเพียงพอแล้ว จะแสดงคำแนะนำให้ผู้ใช่ทราบ พร้อมทั้งค่าความน่าจะเป็นเพื่อให้ผู้ใช้ทราบว่าเชื่อมั่นในคำแนะนำนั้นได้มากน้อยเพียงใด และถ้าผู้ใช้ต้องการคำอธิบายประกอบคำแนะนำ สามารถเรียกดูคำอธิบายได้โดยการพิมพ์คำสั่ง why

ในกรณีที่ข้อมูลที่ใช้ระบุไม่ปรากฏในกฎใดๆ ของฐานความรู้ ระบบผู้เชี่ยวชาญจะโต้ตอบด้วยข้อความดังรูปที่ 5.7

```

SWI-Prolog -- c:/Documents and Settings/Nittaya/Desktop/expertshell1.pl
File Edit Settings Run Debug Help

1 ?- expertshell.
This is the Easy Expert System shell.
Type help. load. solve. why. quit. or 99.
at the prompt.
expert-shell> load.
Enter file name in single quotes (ex. '1.knb'.): '1.knb'.
% 1.knb compiled 0.01 sec, 2,336 bytes
expert-shell> solve.

What is the value for tear?
[1-reduced, 2-normal, 99-exitShell]
Enter the choice> 2.

What is the value for age?
[1-young, 2-pre_presbyopic, 3-presbyopic, 99-exitShell]
Enter the choice> 1.
The answer is __yes__ with probability 0.166667
expert-shell> why.

The answer is ...yes... with probability = 0.166667.
The known storage are
[age(young), tear(normal)]
expert-shell>

```

รูปที่ 5.6 ระบบผู้เชี่ยวชาญที่ให้คำแนะนำเกี่ยวกับคอนแทกเลนส์

```

SWI-Prolog -- c:/Documents and Settings/Nittaya/Desktop/expertshell1.pl
File Edit Settings Run Debug Help

1 ?- expertshell.
This is the Easy Expert System shell.
Type help. load. solve. why. quit. or 99.
at the prompt.
expert-shell> load.
Enter file name in single quotes (ex. '1.knb'.): '1.knb'.
% 1.knb compiled 0.00 sec, 2,336 bytes
expert-shell> solve.

What is the value for tear?
[1-reduced, 2-normal, 99-exitShell]
Enter the choice> 2.

What is the value for age?
[1-young, 2-pre_presbyopic, 3-presbyopic, 99-exitShell]
Enter the choice> 3.
No answer found.
expert-shell>

```

รูปที่ 5.7 การโต้ตอบของระบบผู้เชี่ยวชาญกรณีไม่มีข้อมูลปรากฏในฐานความรู้

5.2 การทดสอบโปรแกรมเพื่อการประมวลผลหลังการทำเหมืองข้อมูล

5.2.1 วิธีการทดสอบโปรแกรมเพื่อการประมวลผลหลังการทำเหมืองข้อมูล

การทำเหมืองข้อมูลในงานวิจัยนี้เน้นที่การทำเหมืองข้อมูลเพื่อการจำแนก วิธีการสร้างโมเดลข้อมูลใช้หลักการของการสร้างต้นไม้ตัดสินใจตามอัลกอริทึม ID3 (Quinlan 1993) ที่ใช้การสร้างโหนดของต้นไม้เป็นเครื่องมือในการแยกข้อมูลที่มีหลายคลาสปะปนกัน ให้เหลือเป็นกลุ่มข้อมูลย่อยที่เป็นคลาสเดียวกัน จากนั้นอ่านลักษณะของข้อมูลในแต่ละกลุ่มย่อยจากโครงสร้างโหนดและกิ่งของต้นไม้ตัดสินใจ วิธีการจำแนกข้อมูลเพื่อให้ได้โมเดลในลักษณะนี้ได้รับการยอมรับว่ามีความถูกต้องของโมเดลสูง และข้อเด่นของวิธีการนี้คือ โมเดลเข้าใจได้ง่าย

โมเดลที่ได้จากการทำเหมืองข้อมูลนี้จะถูกประมวลผลต่อไป โดยโมเดลจะถูกเปลี่ยนเป็นกฎการตัดสินใจที่มีค่าความน่าจะเป็นกำกับ ค่าความน่าจะเป็นนี้ถูกใช้เป็นเกณฑ์เพื่อวัดอัตราส่วนของจำนวนข้อมูลที่สามารถอธิบายได้ด้วยกฎการตัดสินใจ (หรือเรียกว่าค่า coverage) ซึ่งค่านี้จะแปลความหมายได้ถึง โอกาสที่กฎนั้นจะประยุกต์ใช้ได้กับข้อมูลใหม่ที่จะเกิดขึ้นในอนาคต

ดังนั้นวิธีการทดสอบความสามารถของโปรแกรมเพื่อการประมวลผลหลังการทำเหมืองข้อมูลจะใช้วิธีแยกข้อมูลส่วนใหญ่เป็นข้อมูลฝึก เพื่อสร้างโมเดลข้อมูลและสร้างกฎการตัดสินใจที่จะถูกแปลงเป็นกฎในระบบผู้เชี่ยวชาญ จากนั้นจะทดสอบความถูกต้องในการให้คำแนะนำของระบบผู้เชี่ยวชาญ โดยข้อมูลที่ใช้ทดสอบเป็นข้อมูลขนาดเล็กมีจำนวน 16 รายการ ชุดข้อมูลที่ใช้ในการทดสอบโปรแกรมนี้เป็นข้อมูลมาตรฐานของ UCI Machine Learning Repository: Data Sets (<http://archive.ics.uci.edu/ml/datasets.html>) โดยเลือกใช้ข้อมูล post-operative patients (จำนวนข้อมูลฝึกมี 70 เรคคอร์ด แต่ละเรคคอร์ดประกอบด้วย 8 แอททริบิวต์) และข้อมูล breast-cancer recurrences (จำนวนข้อมูลฝึกมี 175 เรคคอร์ด แต่ละเรคคอร์ดประกอบด้วย 9 แอททริบิวต์) ข้อมูลฝึกและข้อมูลทดสอบของชุดข้อมูลทั้งสองในรูปแบบของโปรล็อก ปรากฏในภาคผนวก ข

ตัวอย่างวิธีการทดสอบความถูกต้องของระบบผู้เชี่ยวชาญกับข้อมูลคนไข้รายที่ 71 ในชุดข้อมูล post-operative patients แสดงได้ดังรูปที่ 5.8 ขั้นตอนการใช้งาน expert system shell ที่แสดงในรูปแบบเป็นการสอบถามคำแนะนำกรณีคนไข้หลังผ่าตัด มีค่าอุณหภูมิภายในร่างกาย ค่าความดัน และค่าอื่นๆ เป็น [internalTemp= mid, surfaceTemp=mid, oxygenSaturation=excellent, bloodPressure=high, tempStability=stable, coreTempStability=stable, bpStability=stable, comfort=10] ผลการวินิจฉัยของแพทย์ระบุให้ส่งคนไข้รายนี้ไปที่หอผู้ป่วยหรือ ward และผลจากคำแนะนำของระบบผู้เชี่ยวชาญเป็น ward (คำแนะนำอนุมานจากกฎที่ 12 ในฐานะความรู้ของระบบผู้เชี่ยวชาญ) แสดงว่าคำแนะนำที่ระบบผู้เชี่ยวชาญให้แก่คนไข้รายนี้ถูกต้อง กฎที่ใช้ในระบบผู้เชี่ยวชาญตามตัวอย่างนี้ สร้างจากโปรแกรมคัดเลือกกฎการตัดสินใจที่กำหนดค่าความน่าจะเป็น

ขั้นต่ำ 0.001 กฎทั้งหมดในระบบผู้เชี่ยวชาญแสดงได้ดังรูปที่ 5.9 ในกรณีของข้อมูล breast-cancer recurrences ฐานความรู้ของระบบผู้เชี่ยวชาญแสดงดังรูปที่ 5.10

```

/* *** Test Data
%
=====
instance(71, class=ward, [internalTemp=mid, surfaceTemp=mid, oxygenSaturation=excellent,
bloodPressure=high, tempStability=stable, coreTempStability=stable,
bpStability=stable, ccmfort=10]).

% 1.knb
% for expert shell. --- written by Postprocess
% top_goal where the inference starts.

top_goal(X,V) :- type(X,V).

type(ward,0.1):-comfort(10),bloodPressure(high),surfaceTemp(low). % generated rule
type(ward,0.0714286):-comfort(10),bloodPressure(mid),surfaceTemp(low),bpStability(stable). % generated rule
type(ward,0.0714286):-comfort(10),bloodPressure(mid),surfaceTemp(high),bpStability(mod_stable). % generated rule
type(ward,0.0571429):-comfort(10),bloodPressure(high),surfaceTemp(mid),bpStability(mod_stable). % generated rule
type(ward,0.0428571):-comfort(15),bpStability(unstable),surfaceTemp(mid). % generated rule
type(ward,0.0428571):-comfort(10),bloodPressure(mid),surfaceTemp(mid),bpStability(stable),internalTemp(mid),tempSt
type(ward,0.0428571):-comfort(10),bloodPressure(low). % generated rule
type(ward,0.0285714):-comfort(15),bpStability(unstable),surfaceTemp(high). % generated rule
type(ward,0.0285714):-comfort(15),bpStability(stable),internalTemp(mid),surfaceTemp(mid). % generated rule
type(ward,0.0285714):-comfort(15),bpStability(mod_stable). % generated rule
type(home,0.0285714):-comfort(10),bloodPressure(mid),surfaceTemp(mid),bpStability(mod_stable). % generated rule
type(ward,0.0285714):-comfort(10),bloodPressure(mid),surfaceTemp(mid),bpStability(mod_stable),oxygenSaturation(good)
type(home,0.0142857):-comfort(15),bpStability(unstable),surfaceTemp(low). % generated rule
type(home,0.0142857):-comfort(15),bpStability(stable),internalTemp(mid),surfaceTemp(low),oxygenSaturation(good).

```

```

C:\Program Files\Microsoft Windows\Winlogon\Shell\Shell.nls
File Edit Format Run Debug Help
expert-shell> solve.

What is the value for comfort?
[1-5, 2-7, 3-10, 4-15, 99-exitShell]
Enter the choice> 3.

What is the value for bloodPressure?
[1-high, 2-mid, 3-low, 99-exitShell]
Enter the choice> 1.

What is the value for surfaceTemp?
[1-mid, 2-high, 3-low, 99-exitShell]
Enter the choice> 1.

What is the value for bpStability?
[1-stable, 2-mod_stable, 3-unstable, 99-exitShell]
Enter the choice> 1.

What is the value for oxygenSaturation?
[1-excellent, 2-good, 3-fair, 4-poor, 99-exitShell]
Enter the choice> 1.
The answer is __ward__ with probability 0.0285714
expert-shell> why.

The answer is ...ward... with probability = 0.0285714.
The known storage are
[oxygenSaturation(excellent), bpStability(stable), surfaceTemp(mid), bloodPressure(high), comfort(
expert-shell>

```

รูปที่ 5.8 ตัวอย่างการทดสอบความถูกต้องของระบบผู้เชี่ยวชาญ


```

1 WordPad
File Edit View Insert Format Help
top_goal(X,V) :- type(X,V).

type(ward,0.1):-comfort(10),bloodPressure(high),surfaceTemp(low). % generated rule
type(ward,0.0714286):-comfort(10),bloodPressure(mid),surfaceTemp(low),bpStability(stable). % gener
type(ward,0.0714286):-comfort(10),bloodPressure(mid),surfaceTemp(high),bpStability(mod_stable). % g
type(ward,0.0571429):-comfort(10),bloodPressure(high),surfaceTemp(mid),bpStability(mod_stable). % g
type(ward,0.0428571):-comfort(15),bpStability(unstable),surfaceTemp(mid). % generated rule
type(ward,0.0428571):-comfort(10),bloodPressure(mid),surfaceTemp(mid),bpStability(stable),internalT
type(ward,0.0428571):-comfort(10),bloodPressure(low). % generated rule
type(ward,0.0285714):-comfort(15),bpStability(unstable),surfaceTemp(high). % generated rule
type(ward,0.0285714):-comfort(15),bpStability(stable),internalTemp(mid),surfaceTemp(mid). % genera
type(ward,0.0285714):-comfort(15),bpStability(mod_stable). % generated rule
type(home,0.0285714):-comfort(10),bloodPressure(mid),surfaceTemp(mid),bpStability(mod_stable). % g
type(home,0.0285714):-comfort(10),bloodPressure(high),surfaceTemp(mid),bpStability(stable),oxygensa
type(home,0.0142857):-comfort(15),bpStability(unstable),surfaceTemp(low). % generated rule
type(home,0.0142857):-comfort(15),bpStability(stable),internalTemp(mid),surfaceTemp(low),oxygenSatur
type(ward,0.0142857):-comfort(15),bpStability(stable),internalTemp(mid),surfaceTemp(low),oxygenSatur
type(home,0.0142857):-comfort(15),bpStability(stable),internalTemp(low),surfaceTemp(mid). % genera
type(ward,0.0142857):-comfort(15),bpStability(stable),internalTemp(low),surfaceTemp(high). % gener
type(home,0.0142857):-comfort(15),bpStability(stable),internalTemp(high). % generated rule
type(home,0.0142857):-comfort(10),bloodPressure(mid),surfaceTemp(mid),bpStability(unstable),tempStab
type(ward,0.0142857):-comfort(10),bloodPressure(mid),surfaceTemp(mid),bpStability(unstable),tempStab
type(ward,0.0142857):-comfort(10),bloodPressure(mid),surfaceTemp(mid),bpStability(unstable),tempStab
type(home,0.0142857):-comfort(10),bloodPressure(mid),surfaceTemp(mid),bpStability(stable),internalT
type(home,0.0142857):-comfort(10),bloodPressure(mid),surfaceTemp(mid),bpStability(stable),internalT
type(home,0.0142857):-comfort(10),bloodPressure(mid),surfaceTemp(low),bpStability(unstable),interna
type(home,0.0142857):-comfort(10),bloodPressure(mid),surfaceTemp(low),bpStability(mod_stable). % g
type(ward,0.0142857):-comfort(10),bloodPressure(mid),surfaceTemp(high),bpStability(unstable). % ge
type(home,0.0142857):-comfort(10),bloodPressure(mid),surfaceTemp(high),bpStability(stable),internal
type(ward,0.0142857):-comfort(10),bloodPressure(mid),surfaceTemp(high),bpStability(unstable).
type(home,0.0142857):-comfort(10),bloodPressure(high),surfaceTemp(mid),bpStability(unstable).
type(ward,0.0142857):-comfort(10),bloodPressure(high),surfaceTemp(mid),bpStability(stable),oxyg
type(home,0.0142857):-comfort(10),bloodPressure(high),surfaceTemp(mid),bpStability(stable),oxyg
type(home,0.0142857):-comfort(10),bloodPressure(high),surfaceTemp(high),internalTemp(mid). % g
type(ward,0.0142857):-comfort(10),bloodPressure(high),surfaceTemp(high),internalTemp(high). % g
type(home,0.0142857):-comfort(7). % generated rule
type(home,0.0142857):-comfort(5). % generated rule

internalTemp(X):-menuask(internalTemp,X,[mid,high,low]). %generated menu
surfaceTemp(X):-menuask(surfaceTemp,X,[mid,high,low]). %generated menu
oxygenSaturation(X):-menuask(oxygenSaturation,X,[excellent,good,fair,poor]). %generated menu
bloodPressure(X):-menuask(bloodPressure,X,[high,mid,low]). %generated menu
tempStability(X):-menuask(tempStability,X,[stable,mod_stable,unstable]). %generated menu
coreTempStability(X):-menuask(coreTempStability,X,[stable,mod_stable,unstable]). %generated m
bpStability(X):-menuask(bpStability,X,[stable,mod_stable,unstable]). %generated menu
comfort(X):-menuask(comfort,X,[5,7,10,15]). %generated menu
class(X):-menuask(class,X,[home,ward]). %generated menu

```

รูปที่ 5.9 กฎทั้งหมดของระบบผู้เชี่ยวชาญที่สร้างด้วยเกณฑ์ความน่าจะเป็นขั้นต่ำ 0.001

```

WordPad
File Edit View Insert Format Help
[Icons]

% 1.knb
% for expert shell. --- written by Postprocess
% top_goal where the inference starts.

top_goal(X,V) :- type(X,V).

type(no,0.0514286) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(1),age(range5
type(no,0.04) :- invNodes(range0_2),menopause(premeno),breastQuad(right_low). % generated ru
type(no,0.04) :- invNodes(range0_2),menopause(ge40),breast(left),degMalign(1). % generated ru
type(no,0.0342857) :- invNodes(range0_2),menopause(ge40),breast(left),degMalign(3). % generat
type(no,0.0342857) :- invNodes(range0_2),menopause(ge40),breast(left),degMalign(2),irradiat(no
type(no,0.0285714) :- invNodes(range0_2),menopause(premeno),breastQuad(left_low),irradiat(no)
type(no,0.0228571) :- invNodes(range0_2),menopause(premeno),breastQuad(central),breast(left)
type(no,0.0228571) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(2),irradiat(r
type(yes,0.0171429) :- invNodes(range6_8),breastQuad(right_low). % generated rule
type(yes,0.0171429) :- invNodes(range3_5),degMalign(2),breast(left),irradiat(yes). % generate
type(no,0.0171429) :- invNodes(range3_5),degMalign(1). % generated rule
type(yes,0.0171429) :- invNodes(range0_2),menopause(premeno),breastQuad(right_up),irradiat(no
type(no,0.0171429) :- invNodes(range0_2),menopause(premeno),breastQuad(left_up),degMalign(2),k
type(no,0.0171429) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(3),irradiat(r
type(no,0.0171429) :- invNodes(range0_2),menopause(ge40),breast(left),degMalign(2),irradiat(ye
type(no,0.0114286) :- invNodes(range9_11),irradiat(yes),nodeCaps(yes). % generated rule
type(yes,0.0114286) :- invNodes(range9_11),irradiat(no). % generated rule
type(yes,0.0114286) :- invNodes(range6_8),breastQuad(left_low). % generated rule
type(yes,0.0114286) :- invNodes(range3_5),degMalign(3),irradiat(yes),breast(left). % generate
type(no,0.0114286) :- invNodes(range3_5),degMalign(2),breast(right),irradiat(yes). % generate

type(no,0.00571429) :- invNodes(range0_2),menopause(premeno),breastQuad(left_low),irradiat(no)
type(no,0.00571429) :- invNodes(range0_2),menopause(premeno),breastQuad(central),breast(right
type(no,0.00571429) :- invNodes(range0_2),menopause(lt40),degMalign(3). % generated rule
type(no,0.00571429) :- invNodes(range0_2),menopause(lt40),degMalign(2). % generated rule
type(no,0.00571429) :- invNodes(range0_2),menopause(lt40),degMalign(1),age(range60_69). % ger
type(yes,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(3),irradiat
type(yes,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(3),irradiat
type(no,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(3),irradiat
type(yes,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(2),irradiat
type(no,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(2),irradiat
type(yes,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(2),irradiat
type(no,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(2),irradiat
type(yes,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(1),age(range
type(no,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(1),age(range
type(yes,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(1),age(range
type(no,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(right),degMalign(1),age(range
type(yes,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(left),degMalign(2),irradiat
type(no,0.00571429) :- invNodes(range0_2),menopause(ge40),breast(left),degMalign(2),irradiat(r

age(X) :- menuask(age,X,[range20_29, range30_39, range40_49, range50_59, range60_69]). % gener
menopause(X) :- menuask(menopause,X,[lt40, ge40, premeno]). %generated menu
tumorSize(X) :- menuask(tumorSize,X,[rang0_4, range5_9, range10_14, range15_19, range20_24, r
invNodes(X) :- menuask(invNodes,X,[range0_2, range3_5, range6_8, range9_11, range15_17, range
nodeCaps(X) :- menuask(nodeCaps,X,[missing, yes, no]). %generated menu
degMalign(X) :- menuask(degMalign,X,[1, 2, 3]). %generated menu
breast(X) :- menuask(breast,X,[left, right]). %generated menu
breastQuad(X) :- menuask(breastQuad,X,[left_up, left_low, right_up, right_low, central]). %ge
irradiat(X) :- menuask(irradiat,X,[yes, no]). %generated menu
class(X) :- menuask(class,X,[no, yes]). %generated menu

%end of automatic post process

```

รูปที่ 5.10 ฐานความรู้ของระบบผู้เชี่ยวชาญในการแนะนำเกี่ยวกับ breast-cancer recurrences

การทดสอบความถูกต้องในการให้คำแนะนำของระบบผู้เชี่ยวชาญที่พัฒนาขึ้นในงานวิจัยนี้ จะเปรียบเทียบกับผลการทดสอบในการจำแนกข้อมูลด้วยอัลกอริทึม ID3 (decision-tree induction algorithm), PRISM (rule induction algorithm) และ neural network (multi-layer perceptron algorithm)

5.2.2 ผลการทดสอบโปรแกรม

ผลการทดสอบความถูกต้องในการให้คำแนะนำของระบบผู้เชี่ยวชาญ ในกรณีของข้อมูล post-operative patients แสดงดังตารางที่ 5.1 และผลการทดสอบความถูกต้องในการให้คำแนะนำของระบบผู้เชี่ยวชาญ ในกรณีของข้อมูล breast-cancer recurrences แสดงดังตารางที่ 5.2

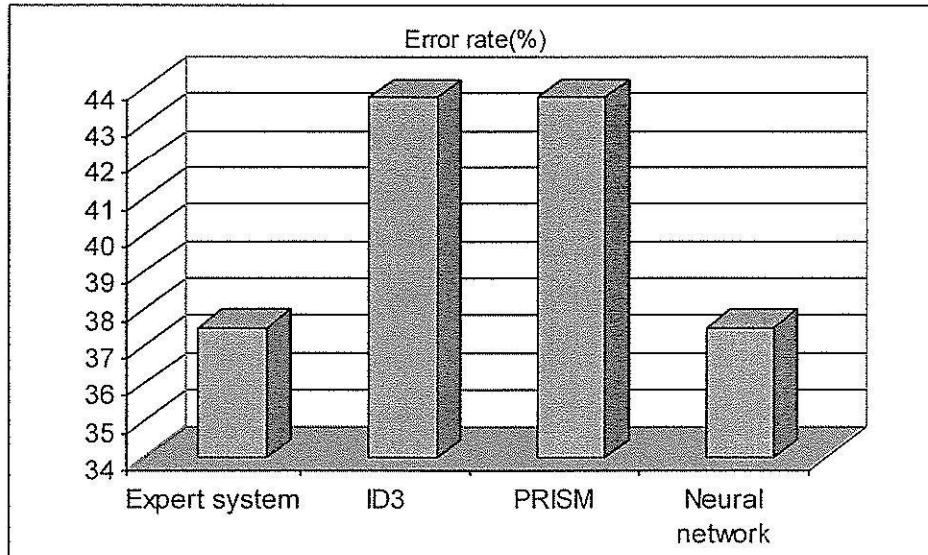
ตารางที่ 5.1 ผลการทดสอบความถูกต้องของระบบผู้เชี่ยวชาญกับข้อมูล post-operative patients

ข้อมูลที่	การวินิจฉัยของแพทย์	คำแนะนำของระบบผู้เชี่ยวชาญ	ผลการทำนายของ ID3	ผลการทำนายของ PRISM	ผลการทำนายของ Neural network
1	Ward	Ward	Ward	Ward	Ward
2	Ward	Ward	Ward	Ward	Ward
3	Ward	Ward	Unclassified	Home	Ward
4	Ward	Home	Home	Home	Home
5	Ward	Ward	Ward	Unclassified	Ward
6	Home	Ward	Ward	Home	Home
7	Ward	Home	Home	Home	Home
8	Ward	Ward	Ward	Ward	Ward
9	Home	No answer found	Ward	Ward	Ward
10	Ward	Home	Home	Home	Home
11	Ward	Ward	Ward	Unclassified	Ward
12	Ward	Ward	Ward	Ward	Ward
13	Home	Ward	Ward	Ward	Ward
14	Ward	Ward	Ward	Ward	Ward
15	Ward	Ward	Ward	Ward	Ward
16	Home	Ward	Ward	Ward	Ward

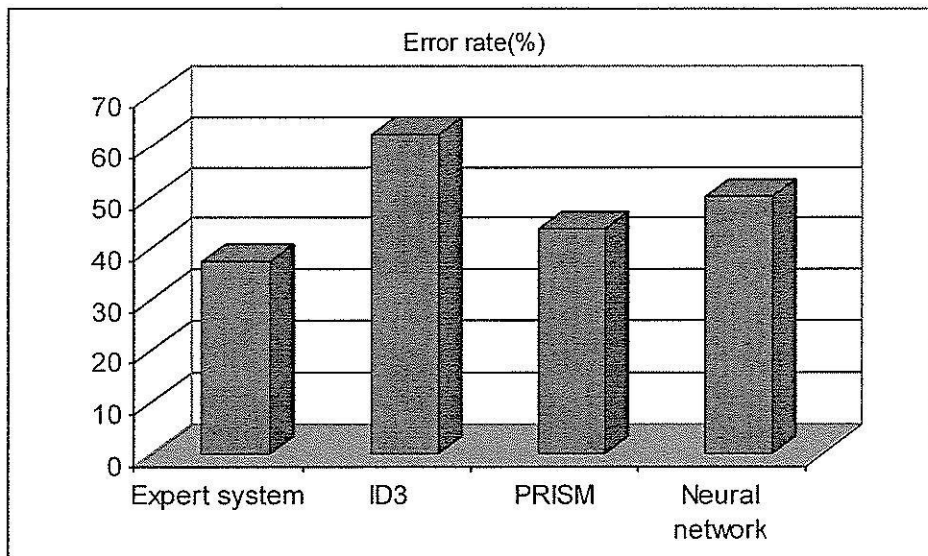
ตารางที่ 5.2 ผลการทดสอบความถูกต้องของระบบผู้เชี่ยวชาญกับข้อมูล breast-cancer recurrences

ข้อมูลที่	การวินิจฉัยของแพทย์	คำแนะนำของระบบผู้เชี่ยวชาญ	ผลการทำนายของ ID3	ผลการทำนายของ PRISM	ผลการทำนายของ Neural network
1	No	No	Yes	No	Yes
2	No	Yes	Yes	Yes	Yes
3	Yes	Yes	No	No	No
4	Yes	No answer found	No	No	Yes
5	No	Yes	Unclassified	No	No
6	No	No answer found	Yes	No	No
7	Yes	No	No	No	No
8	No	No	Yes	Yes	No
9	Yes	No	No	Yes	No
10	No	Yes	No	No	Yes
11	No	No	No	No	No
12	No	Yes	Yes	Yes	Yes
13	Yes	Yes	No	No	Yes
14	No	No answer found	No	No	No
15	No	No	Unclassified	No	Yes
16	Yes	Yes	Yes	Yes	Yes

ผลการเปรียบเทียบความผิดพลาดในการให้คำแนะนำ หรือการทำนายคลาสของข้อมูลทดสอบของโปรแกรม expert system, ID3, PRISM, Neural network กับข้อมูล post-operative patients และ breast-cancer recurrences แสดงดังกราฟในรูปที่ 5.11 และ 5.12 ตามลำดับ



รูปที่ 5.11 กราฟเปรียบเทียบ error rate เมื่อทดสอบกับข้อมูล post-operative patients



รูปที่ 5.12 กราฟเปรียบเทียบ error rate เมื่อทดสอบกับข้อมูล breast-cancer recurrences

อภิปรายผล

จากผลการทดสอบความถูกต้องของโปรแกรมประมวลผลหลังการทำเหมืองข้อมูล หรือโปรแกรม expert system ในทั้งสองชุดข้อมูลเมื่อเทียบกับโปรแกรม ID3 พบว่าโปรแกรม expert system ให้ผลลัพธ์ที่ถูกต้องมากกว่าทั้งๆที่โปรแกรม expert system ใช้โมเดลข้อมูลเริ่มต้นเป็นต้นไม้ตัดสินใจเหมือนกับโปรแกรม ID3 แต่ข้อแตกต่างอยู่ที่ โปรแกรม expert system มีการจัดลำดับกฎการตัดสินใจตามค่าความน่าจะเป็นและคัดเลือกใช้เฉพาะกฎที่มีค่าความน่าจะเป็นสูง

เมื่อเปรียบเทียบความผิดพลาดในการให้คำแนะนำ หรือความผิดพลาดในการทำนาย (error rate) ของโปรแกรมทั้งสี่พบว่าโปรแกรม expert system ให้ค่าความผิดพลาดที่ต่ำที่สุด (คือ 37.50% ในข้อมูลทั้งสองชุด) และเมื่อวิเคราะห์ความผิดพลาดในลักษณะของ false negative (ดังตารางที่ 5.3) พบว่าโปรแกรม expert system ให้ความผิดพลาดชนิดนี้ต่ำที่สุด ซึ่งในการวินิจฉัยทางการแพทย์ความผิดพลาดประเภท false negative (เช่นคนไข้เป็นมะเร็ง แต่ได้รับการวินิจฉัยว่าสุขภาพแข็งแรงเป็นปกติ) ถือว่ามีความร้ายแรงมากกว่าความผิดพลาดประเภท false positive (เช่นคนไข้ที่ไม่เป็นมะเร็ง แต่ได้รับการวินิจฉัยว่าเป็นมะเร็งและจะต้องถูกส่งไปตรวจทางคลินิกเพิ่มเติม)

ตารางที่ 5.3 ผลการวิเคราะห์ความผิดพลาดในลักษณะของ false negative

False negative error	Expert system	ID3	PRISM	Neural network
ข้อมูล post-operative patients ลักษณะการทำนายผิด จากส่งกลับหอผู้ป่วย (ward) เป็นส่งกลับบ้าน (home)	Error = 3/16 = 18.75%	Error = 3/16 = 18.75%	Error = 4/16 = 25%	Error = 3/16 = 18.75%
ข้อมูล breast-cancer recurrences ลักษณะการทำนายผิด จากเป็นมะเร็งซ้ำ (yes) เป็นไม่เป็นมะเร็ง (no)	Error = 2/16 = 12.50%	Error = 5/16 = 31.25%	Error = 4/16 = 25%	Error = 3/16 = 18.75%

โปรแกรมที่ใช้โมเดลข้อมูลในลักษณะของต้นไม้ตัดสินใจ (ID3) หรือในลักษณะของกฎ (expert system และ PRISM) จะมีข้อบกพร่องที่เห็นได้ชัดเจนจากผลการทดลองคือ โมเดลที่ได้ไม่สมบูรณ์ เนื่องจากจะมีบางกรณีที่ไม่สามารถให้คำแนะนำ หรือไม่สามารถทำนายคลาสของข้อมูลได้ ซึ่งจะต่างจากโปรแกรม Neural network ที่มีพื้นฐานการสร้างโมเดลจากการใช้ฟังก์ชันทางคณิตศาสตร์ (ในการทดลองนี้ใช้ sigmoid function) ทำให้สามารถทำนายคลาสของข้อมูลได้ครอบคลุมหมดทุกกรณี

5.3 สรุปผลโครงการวิจัยที่ 4

โปรแกรมเพื่อการทำเหมืองข้อมูลในปัจจุบัน มักจะได้รับการพัฒนาโปรแกรมจนถึงขั้นให้สามารถแสดงผลลัพธ์เป็นโมเดลข้อมูล หรือแพทเทิร์นของกลุ่มข้อมูล ขึ้นตอนต่อจากนั้นที่เป็นการนำโมเดลหรือแพทเทิร์นไปใช้ประโยชน์มักจะให้อยู่ในดุลพินิจของนักวิเคราะห์ข้อมูล ซึ่งนักวิเคราะห์ข้อมูลที่ไม่คุ้นเคยกับกระบวนการทำเหมืองข้อมูล มักจะประสบกับอุปสรรคสำคัญของการคัดเลือกความรู้ใน โมเดลไปใช้งานต่อไป เนื่องจากโปรแกรมทำเหมืองข้อมูลมักจะแสดง

ผลลัพธ์เป็นโมเดลที่มีความซับซ้อนและมีขนาดใหญ่ทำให้แปลผลได้ยาก ตัวอย่างเช่นจากข้อมูล post-operative patients เมื่อสร้างโมเดลข้อมูลด้วยโปรแกรม ID3 จะได้ผลลัพธ์ที่เป็นโครงสร้างต้นไม้ขนาดใหญ่ (รูปที่ 5.13) และข้อมูลเดียวกันนี้เมื่อสร้างโมเดลด้วยโปรแกรม Multi-layer perceptron ซึ่งเป็นโปรแกรมในกลุ่ม Neural network จะได้ผลลัพธ์ (รูปที่ 5.14) เป็นโมเดลข้อมูลในลักษณะฟังก์ชันที่แปลผลได้ค่อนข้างยากสำหรับผู้ที่ไม่คุ้นเคย

comfort = 05: home comfort = 07: home comfort = 10 bloodPressure = high surfaceTemp = mid bpStability = stable oxygenSaturation = excellent: ward oxygenSaturation = good surfaceTempStability = stable: home surfaceTempStability = mod-stable: null surfaceTempStability = unstable: ward oxygenSaturation = fair: null oxygenSaturation = poor: null bpStability = mod-stable: ward bpStability = unstable: home surfaceTemp = high internalTemp = mid: home internalTemp = high: ward internalTemp = low: null surfaceTemp = low: ward bloodPressure = mid surfaceTemp = mid bpStability = stable internalTemp = mid surfaceTempStability = stable: ward surfaceTempStability = mod-stable: null surfaceTempStability = unstable: ward internalTemp = high: home internalTemp = low oxygenSaturation = excellent: home oxygenSaturation = good: home oxygenSaturation = fair: null oxygenSaturation = poor: null bpStability = mod-stable: home bpStability = unstable internalTemp = mid	surfaceTempStability = stable: ward surfaceTempStability = mod-stable: null surfaceTempStability = unstable: home internalTemp = high: ward internalTemp = low: null surfaceTemp = high bpStability = stable internalTemp = mid: home internalTemp = high: null internalTemp = low: ward bpStability = mod-stable: ward bpStability = unstable: ward surfaceTemp = low bpStability = stable: ward bpStability = mod-stable: home bpStability = unstable internalTemp = mid: home internalTemp = high: null internalTemp = low: home bloodPressure = low: ward comfort = 15 bpStability = stable bloodPressure = high: home bloodPressure = mid internalTemp = mid: ward internalTemp = high: null internalTemp = low surfaceTemp = mid: home surfaceTemp = high: ward surfaceTemp = low: null bloodPressure = low: null bpStability = mod-stable: ward bpStability = unstable surfaceTemp = mid: ward surfaceTemp = high: ward surfaceTemp = low: home
--	---

รูปที่ 5.13 โมเดลของข้อมูล breast-cancer recurrences สร้างจากโปรแกรม ID3

<p>Sigmoid Node 0</p> <p>Inputs Weights</p> <p>Threshold 4.838013774039095</p> <p>Node 2 -3.908582636087111</p> <p>Node 3 -3.162267392436414</p> <p>Node 4 1.0981619846468047</p> <p>Node 5 0.18908775947239695</p> <p>Node 6 -5.74938601906933</p> <p>Node 7 4.422399319810025</p> <p>Node 8 0.1999742473645021</p> <p>Node 9 -3.5324578914467377</p> <p>Node 10 0.11046630190222735</p> <p>Node 11 -1.7079259280625576</p> <p>Node 12 2.566868144965138</p> <p>Node 13 -1.1374211603297253</p> <p>Node 14 -2.7865026549167866</p> <p>Node 15 0.5042240856020124</p> <p>Sigmoid Node 1</p> <p>Inputs Weights</p> <p>Threshold -4.837381256263012</p> <p>Node 2 3.9522624615165793</p> <p>Node 3 3.1682972818602746</p> <p>Node 4 -1.094499095299558</p> <p>Node 5 -0.13283814684003517</p> <p>Node 6 5.746872610366041</p> <p>Node 7 -4.424392875653778</p> <p>Node 8 -0.18722208735744159</p> <p>Node 9 3.527387810727785</p> <p>Node 10 -0.10472586398059963</p> <p>Node 11 1.6962159718097416</p> <p>Node 12 -2.560804638307743</p> <p>Node 13 1.0737048825607642</p> <p>Node 14 2.7877184996555053</p> <p>Node 15 -0.5273986540734854</p> <p>Sigmoid Node 2</p> <p>Inputs Weights</p> <p>Threshold -0.27814804385505276</p> <p>Attrib internalTemp=mid 0.5810283641710741</p> <p>Attrib internalTemp=high 0.21179494161259477</p> <p>Attrib internalTemp=low -0.5449481721523186</p> <p>Attrib surfaceTemp=mid 1.1141644074041142</p> <p>Attrib surfaceTemp=high 1.5150741028274102</p> <p>...</p>	<p>...</p> <p>Sigmoid Node 15</p> <p>Inputs Weights</p> <p>Threshold -0.2930395536777638</p> <p>Attrib internalTemp=mid 0.062120538342274045</p> <p>Attrib internalTemp=high 0.3128132567811115</p> <p>Attrib internalTemp=low 0.018576988447827396</p> <p>Attrib surfaceTemp=mid 0.04411226345984964</p> <p>Attrib surfaceTemp=high 0.5845640457361527</p> <p>Attrib surfaceTemp=low -0.3416470192893855</p> <p>Attrib oxygenSaturation=excellent -0.022407433837484465</p> <p>Attrib oxygenSaturation=good -0.0017430323732858114</p> <p>Attrib oxygenSaturation=fair -0.01474001719084124</p> <p>Attrib oxygenSaturation=poor -0.02503529326219822</p> <p>Attrib bloodPressure=high 0.3330125093594764</p> <p>Attrib bloodPressure=mid -0.3816134197463625</p> <p>Attrib bloodPressure=low 0.32842425053023594</p> <p>Attrib surfaceTempStability=stable 0.11418920446981239</p> <p>Attrib surfaceTempStability=mod-stable -0.0132397526586</p> <p>Attrib surfaceTempStability=unstable -0.07986194908507</p> <p>Attrib coreTempStability=stable -0.14532864396285983</p> <p>Attrib coreTempStability=mod-stable 0.31474360142549024</p> <p>Attrib coreTempStability=unstable 0.12862619667684774</p> <p>Attrib bpStability=stable 0.21950090896499233</p> <p>Attrib bpStability=mod-stable -0.10037656428787754</p> <p>Attrib bpStability=unstable 0.1820629305281999</p> <p>Attrib comfort=05 0.04690470481373389</p> <p>Attrib comfort=07 0.09299279029838585</p> <p>Attrib comfort=10 -0.024417921014600892</p> <p>Attrib comfort=15 0.4216833375087676</p> <p>Class home</p> <p>Input</p> <p>Node 0</p> <p>Class ward</p> <p>Input</p> <p>Node 1</p>
---	--

รูปที่ 5.14 โมเดลของข้อมูล breast-cancer recurrences สร้างจากโปรแกรม Multi-layer perceptron

โครงการวิจัยเรื่องการประมวลผลหลังกระบวนการทำเหมืองข้อมูลนี้ จึงมีวัตถุประสงค์หลักเพื่อจะพัฒนาโปรแกรมเพื่อรับอินพุตเป็นโมเดลข้อมูล จากนั้นแปลงโมเดลข้อมูลให้อยู่ในรูปแบบที่นักวิเคราะห์ข้อมูลโดยทั่วไปเข้าใจได้ง่ายกว่าในรูปแบบของ decision tree ในรูปที่ 5.13 หรือรูปแบบ sigmoid functions ในรูปที่ 5.14 นอกจากนี้แปลงรูปแบบโมเดลข้อมูลให้เข้าใจได้ง่ายแล้ว งานวิจัยนี้ยังได้เพิ่มเทคนิคของการถ่วงน้ำหนักของโมเดลให้มีขนาดเล็กลง ทั้งนี้เพื่อความสะดวกในการแปลผลและนำไปใช้งาน งานวิจัยนี้ได้อำนวยความสะดวกด้านการใช้งานโมเดลที่ได้รับการถ่วงน้ำหนักแล้วด้วยการจัดสร้างฐานความรู้โดยอัตโนมัติจากโมเดลข้อมูล และสร้าง expert system shell ที่ผู้ใช้สามารถสอบถามโมเดลข้อมูลที่อยู่ในฐานความรู้ได้อย่างสะดวก ตัวอย่างของ

ฐานความรู้ที่สร้างจากข้อมูล post-operative patients และกั้นกรองไว้เฉพาะโมเดลที่มีค่าความน่าจะเป็นสูงกว่า 0.05 แสดงได้ดังรูปที่ 5.15 และแสดงการใช้ experts system shell สอบถามโมเดล ข้อมูลนี้ในรูปที่ 5.16

```

type(ward,0.1) :- comfort(10), bloodPressure(high), surfaceTemp(low).
type(ward,0.07142) :- comfort(10), bloodPressure(mid), surfaceTemp(low), bpStability(stable).
type(ward,0.07142) :- comfort(10), bloodPressure(mid), surfaceTemp(high), bpStability(mod_stable).
type(ward,0.05714) :- comfort(10), bloodPressure(high), surfaceTemp(mid), bpStability(mod_stable).
type(ward,0.04285) :- comfort(15), bpStability(unstable), surfaceTemp(mid).
type(ward,0.04285) :- comfort(10), bloodPressure(mid), surfaceTemp(mid), bpStability(stable),
                    internalTemp(mid), tempStability(unstable).
type(ward,0.04285) :- comfort(10), bloodPressure(low).
type(ward,0.02857) :- comfort(15), bpStability(unstable), surfaceTemp(high).
type(ward,0.02857) :- comfort(15), bpStability(stable), internalTemp(mid), surfaceTemp(mid).
type(ward,0.02857) :- comfort(15), bpStability(mod_stable).
type(home,0.02857) :- comfort(10), bloodPressure(mid), surfaceTemp(mid), bpStability(mod_stable).
type(ward,0.02857) :- comfort(10), bloodPressure(high), surfaceTemp(mid),
                    bpStability(stable), oxygenSaturation(excellent).

```

รูปที่ 5.15 โมเดลของข้อมูล post-operative patents ที่มีค่าความน่าจะเป็นสูงกว่า 0.02

```

SWI-Prolog  c:/Documents and Settings/Nittaya/Desktop/expertshell.pl
File Edit Settings Run Debug Help
% c:/Documents and Settings/Nittaya/Desktop/expertshell.pl compiled 0.00 sec. 6,620 bytes
Welcome to SWI-Prolog (Multi-threaded, 32 bits, Version 5.7.11)
Copyright (c) 1990-2009 University of Amsterdam.
SWI-Prolog comes with ABSOLUTELY NO WARRANTY. This is free software,
and you are welcome to redistribute it under certain conditions.
Please visit http://www.swi-prolog.org for details.

For help, use ?- help(Topic). or ?- apropos(Word).

1 ?- expertshell.
This is the Easy Expert System shell.
Type help, load, solve, why, quit, or 99.
at the prompt.
expert-shell> load.
Enter file name in single quotes (ex. '1.knb'): '1.knb'.
% 1.knb compiled 0.00 sec, 4,844 bytes
expert-shell> solve.

What is the value for comfort?
[1-5, 2-7, 3-10, 4-15, 99-exitShell]
Enter the choice> 3.

What is the value for bloodPressure?
[1-high, 2-mid, 3-low, 99-exitShell]
Enter the choice> 3.
The answer is __ward__ with probability 0.0428571
expert-shell> why.

The answer is ...ward... with probability = 0.0428571.
The known storage are
[bloodPressure(low), comfort(10)]
expert-shell>

```

รูปที่ 5.16 การใช้ expert system shell สอบถามโมเดลข้อมูล

ข้อเสนอแนะ

(1) โปรแกรมสร้างโมเดลข้อมูลที่พัฒนาขึ้นนี้มีข้อกำหนดเกี่ยวกับชนิดข้อมูล โดยข้อมูลที่ใช้จะต้องเป็นชนิด nominal หรือ categorical ถ้าข้อมูลใดเป็นจำนวนเลข (numeric) จะต้องแปลงตัวเลขให้เป็นข้อความ และการจำแนกคลาสของข้อมูลเป็นชนิด binary classification ดังนั้นแนวทางการปรับปรุงงานวิจัยนี้ต่อไปคือพัฒนาวิธีการจำแนกข้อมูลให้เป็นแบบหลายคลาส หรือ multi-class classification และเพิ่มฟังก์ชันการแปลงชนิดข้อมูลจากจำนวนเลขเป็นช่วงของค่า (interval)

(2) การสร้างกฎการตัดสินใจในงานวิจัยนี้สร้างจาก classification rules ทั้งนี้โมเดลที่ได้การทำเหมืองข้อมูลประเภทอื่น เช่น clustering model, association model สามารถแปลงให้เป็นกฎการตัดสินใจได้เช่นเดียวกัน

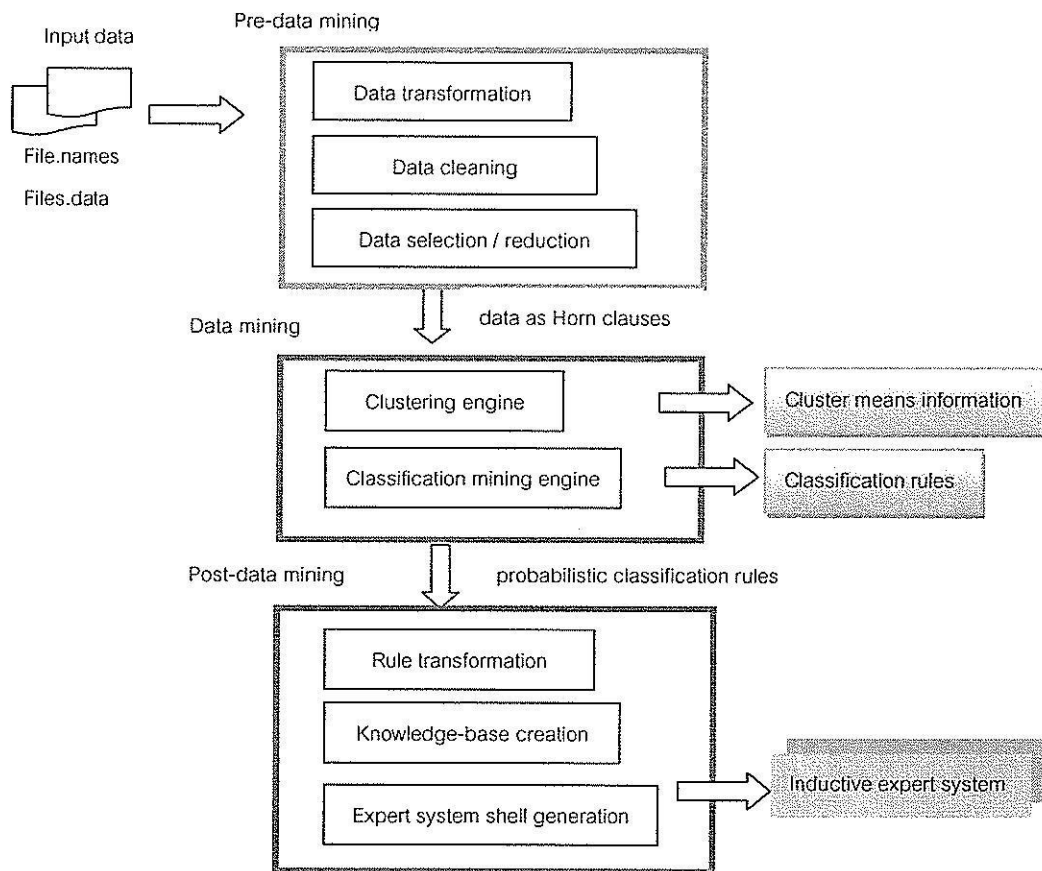
(3) จากแนวทางการลดขนาดของโมเดล ด้วยการใช้เฉพาะกฎการตัดสินใจที่มีค่าความน่าจะเป็นสูง จะทำให้ได้โมเดลที่มีจำนวนกฎน้อย แต่ผลที่อาจจะตามมาคือกฎที่ถูกคัดเลือกไว้ อาจจะมีเงื่อนไขการพิจารณาไม่ครบถ้วนทำให้โมเดลไม่สมบูรณ์ หรือกฎบางส่วนอาจขัดแย้งกัน แนวทางแก้ไขปัญหากรณีเงื่อนไขไม่ครบถ้วนคือต้องมี default rule เตรียมไว้ ส่วนในกรณีที่กฎขัดแย้งอาจใช้วิธีการโหวตจากทุกกฎที่เกี่ยวข้อง จากนั้นใช้ค่าทำนายเป็นค่าส่วนใหญ่จากผลโหวต

บทที่ 6

บทสรุป

6.1 ลักษณะเด่นของระบบ SUT Miner

ระบบเหมืองข้อมูล SUT Miner ประกอบขึ้นจากโมดูลต่างๆ เพื่อทำหน้าที่ในส่วน pre-data mining, data mining และ post-data mining โมดูลหลักในระบบเหมืองข้อมูลแต่ละส่วนแสดงไว้ดังรูปที่ 6.1



รูปที่ 6.1 ระบบเหมืองข้อมูล SUT Miner และ โมดูลหลักในแต่ละส่วน

ข้อมูลที่เป็นอินพุตเข้ามาในระบบเหมืองข้อมูล จะเป็นชุดข้อมูลที่ประกอบด้วยไฟล์สองชนิดคือ names file และ data file ซึ่งลักษณะไฟล์แบบนี้เป็นรูปแบบปกติของข้อมูลใน UCI repository ระบบเหมืองข้อมูล SUT Miner จะเริ่มดำเนินงานด้วยส่วน Pre-data mining (หรือ preprocessing) ที่ทำหน้าที่รวมทั้งสองไฟล์แล้วแปลงรูปแบบไฟล์ให้อยู่ในลักษณะคลอสของภาษาโปรล็อก ในส่วน Pre-data mining ยังมีโมดูล Data cleaning และ Data selection/reduction ทำหน้าที่

จัดการกับ missing values และลดขนาดข้อมูลด้วยการคัดเลือกแอททริบิวต์ และสุ่มเลือกข้อมูลตัวแทน เทคนิคการสุ่มจะมีทั้งการสุ่มแบบปกติ (simple random sampling with/without replacement) และการสุ่มตามความหนาแน่น density-biased sampling

ส่วนที่ทำหน้าที่ค้นหาโมเดลจากข้อมูล คือส่วน Data mining ที่ประกอบด้วย Clustering engine (โมเดลทำเหมืองข้อมูลแบบจัดกลุ่ม) และ Classification mining engine (โมเดลทำเหมืองข้อมูลแบบจำแนก)

โมเดลทำเหมืองข้อมูลแบบจัดกลุ่มพัฒนาจากอัลกอริทึม incremental k-means clustering ซึ่งเป็นการจัดกลุ่มข้อมูลแบบเพิ่มพูน เพื่อให้สามารถรองรับข้อมูลขนาดใหญ่ได้ ผลลัพธ์ที่ได้จากโมเดลนี้คือ ค่าตัวแทนกลุ่ม (cluster means) และกลุ่มข้อมูล (clusters)

ในส่วนของการทำเหมืองข้อมูลแบบจำแนก จะแสดงโมเดลที่เป็นผลลัพธ์ของโมเดลนี้ได้ ในสองรูปแบบคือ โมเดลในลักษณะของต้นไม้ตัดสินใจ และโมเดลในลักษณะของกฎการตัดสินใจ โดยส่วนสร้างกฎการตัดสินใจจะมีฟังก์ชันให้ผู้ใช้สามารถระบุการสร้างกฎ ให้สร้างเฉพาะกฎการตัดสินใจที่มีค่าความน่าจะเป็นสูง (probabilistic decision rules) ที่จะสามารถประยุกต์ใช้โมเดลกับข้อมูลในอนาคต

Probabilistic decision rules ที่ได้จากโมเดลทำเหมืองข้อมูลแบบจำแนก จะถูกส่งต่อไปยังส่วน Post-data mining (หรือ postprocessing) เพื่อเปลี่ยนกฎการตัดสินใจให้เป็นความรู้ในฐานความรู้ของระบบผู้เชี่ยวชาญ ความรู้นี้จะถูกสร้างด้วยกระบวนการอัตโนมัติในโมเดล Expert system shell generation และถูกผนวกเข้าเป็นส่วนหนึ่งของ expert system shell ของระบบผู้เชี่ยวชาญ กระบวนการนี้จะช่วยให้ผู้ใช้สามารถเรียกดูผลลัพธ์ของการทำเหมืองข้อมูลกับข้อมูลอินพุตผ่านระบบผู้เชี่ยวชาญ ซึ่งจะทำความเข้าใจได้ง่ายกว่าการพิจารณาโดยตรงจากโมเดลในลักษณะต้นไม้ตัดสินใจที่เป็นผลลัพธ์จากโมเดลทำเหมืองข้อมูล

เมื่อเปรียบเทียบกับระบบเหมืองข้อมูลในปัจจุบันเช่น WEKA (Hall et al, 2009) ระบบ SUT Miner จะมีความสามารถที่สูงกว่าในด้านการผนวกผลลัพธ์ของการทำเหมืองข้อมูลแบบจำแนกให้เข้าสู่ฐานความรู้ของระบบผู้เชี่ยวชาญได้โดยอัตโนมัติ ความสามารถในด้านนี้เกิดขึ้นได้จากการที่ระบบเหมืองข้อมูล SUT Miner พัฒนาขึ้นด้วยภาษาโปรแกรมที่เหมาะสมกับงาน rule processing และการค้นหาคำตอบทั้งหมดด้วยเทคนิค backtracking

6.2 ข้อจำกัดของระบบ SUT Miner

ระบบเหมืองข้อมูล SUT Miner ได้รับการพัฒนาขึ้นในลักษณะของ rapid prototyping เพื่อให้สามารถทดสอบความถูกต้องของอัลกอริทึมใหม่ที่พัฒนาขึ้น (ได้แก่อัลกอริทึม density-

biased sampling, probabilistic rule induction, incremental k-means clustering และ expert system shell generation) ภายในเวลาที่รวดเร็ว เนื่องจากระบบเหมืองข้อมูล SUT Miner ยังเป็น โปรแกรม ต้นแบบ จึงยังมีข้อจำกัดหลายประการ ได้แก่ ส่วนติดต่อกับผู้ใช้ยังขาดลักษณะของ graphical user interface, ข้อมูลที่จะนำเข้ายังระบบต้องอยู่ในหน่วยความจำหลัก ดังนั้นขนาดของข้อมูลที่ระบบ รองรับได้จะขึ้นอยู่กับขนาดหน่วยความจำของเครื่องคอมพิวเตอร์, รูปแบบของข้อมูลเข้ายังจำกัดอยู่ เฉพาะรูปแบบที่ใช้โดย UCI repository, ยังขาดฟังก์ชันในการวิเคราะห์แอททริบิวต์ เช่น principal component analysis

6.3 ข้อเสนอแนะและแนวทางการพัฒนาในอนาคต

ระบบเหมืองข้อมูล SUT Miner ที่พัฒนาขึ้นนี้ยังสามารถปรับปรุงให้มีความสามารถ สูงขึ้นในด้านต่างๆ ดังนี้

- (1) การคัดเลือกแอททริบิวต์ข้อมูลในงานวิจัยนี้ ใช้วิธีแสดงการกระจายค่าของข้อมูลใน ลักษณะของฮิสโตแกรม จากนั้นให้ผู้ใช้กำหนดว่าจะคัดเลือกแอททริบิวต์ใดบ้าง ใน ส่วนนี้สามารถพัฒนาให้ดีขึ้นได้โดยการเพิ่มเทคนิคการพิจารณาความเหมาะสมของ แอททริบิวต์ (เช่น คำนวณ gain value) จากนั้นแสดงคำแนะนำให้ผู้ใช้ตัดสินใจว่า ควรจะเลือกแอททริบิวต์ใด
- (2) โมเดลทำเหมืองข้อมูลแบบจัดกลุ่ม เน้นการจัดกลุ่มข้อมูลที่มีชนิดข้อมูลเป็นข้อความ (nominal, categorical) การจัดกลุ่มข้อมูลที่แอททริบิวต์มีค่าเป็นจำนวนเลข สามารถ ทำได้โดยปรับปรุงซอฟต์แวร์นี้เพียงเล็กน้อย แต่ถ้าค่าตัวเลขในแอททริบิวต์มีช่วง ของค่าที่เป็นไปได้กว้างมาก เช่นในกรณีแอททริบิวต์อายุ ที่มีค่าอยู่ระหว่าง 0-120 ควรเพิ่มฟังก์ชัน normalization เช่นใช้เทคนิค max-min normalization หรือ z-score standardization เพื่อปรับลดช่วงกว้างของค่าสูงสุด-ต่ำสุด ทั้งนี้เพื่อลด bias ของการ คำนวณค่า mean ของกลุ่มข้อมูล
- (3) โมเดลทำเหมืองข้อมูลแบบจำแนกมีแนวทางการลดขนาดของโมเดล ด้วยการ ใช้เฉพาะกฎการตัดสินใจที่มีค่าความน่าจะเป็นสูง จะทำให้ได้โมเดลที่มีจำนวนกฎน้อย แต่ผลที่อาจตามมาคือกฎที่ถูกคัดเลือกไว้อาจจะมีเงื่อนไขการพิจารณาไม่ครบถ้วน ทำให้โมเดลไม่สมบูรณ์ หรือกฎบางส่วนอาจขัดแย้งกัน แนวทางแก้ไขปัญหาคือ เงื่อนไขไม่ครบถ้วนคือจะต้องมี default rule เตรียมไว้ ส่วนในกรณีที่กฎขัดแย้งอาจ ใช้วิธีการ โหวตจากทุกกฎที่เกี่ยวข้อง จากนั้นใช้ค่าทำนายเป็นค่าส่วนใหญ่จากผล โหวต

- (4) การสร้างกฎการตัดสินใจในงานวิจัยนี้สร้างจาก classification rules ทั้งนี้โมเดลที่ได้การทำเหมืองข้อมูลประเภทอื่น เช่น clustering model, association model ควรสามารถแปลงให้เป็นกฎที่มีลักษณะคล้ายกฎการตัดสินใจได้เช่นเดียวกัน
- (5) การทำเหมืองข้อมูลเป็นกระบวนการที่ต้องมีการทำซ้ำหลายครั้ง ทำให้ใช้เวลาในการประมวลผลค่อนข้างสูง โดยเฉพาะเมื่อข้อมูลมีขนาดใหญ่ การปรับปรุงในส่วนนี้อาจใช้แนวทางของการประมวลผลแบบขนาน และใช้วิธีการประมวลผลแบบหลายหน่วยประมวลผล (multi-core programming) ซึ่งแนวทางนี้จะเป็แนวทางที่ผู้วิจัยเลือกใช้เพื่อพัฒนาระบบเหมืองข้อมูล SUT Miner ให้มีประสิทธิภาพสูงขึ้นในเวอร์ชันต่อไป

บรรณานุกรม

- G. Adomavicius & A. Tuzhilin (2001). Expert-driven validation of rule-based user models in personalization applications. *Journal of Data Mining and Knowledge Discovery*, 5(1/2): 33-58.
- D.W. Aha (1992). Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies*, 36(1), 267-287.
- D.P. Ballou & G.K. Tayi (1999). Enhancing data quality in data warehouse environments. *Communications of ACM*, 42, 73-78.
- D. Barbara, W. DuMouchel, C. Faloutsos, P.J. Haas, J.H. Hellerstein, Y. Ioannidis, H.V. Jagadish, T. Johnson, R. Ng., V. Poosala, K.A. Ross, & K.C. Servcik (1997). The New Jersey data reduction report. *Bulletin of the Technical Committee on Data Engineering*, 20, 3-45.
- R.J. Bayardo, R. Agrawal, & D. Gunopulos (1999). Constraint-based rule mining in large, dense databases. In *Proceedings of the 15th International Conference on Data Engineering*.
- C. Blake, E. Keogh, & C.J. Merz (1998). UCI Repository of machine learning databases, Department of Information and Computer Science, University of California, Irvine. [<http://www.ics.uci.edu/~mlearn/MLRepository.html>]
- L. Breiman, J. Friedman, R. Olshen, & C. Stone (1984). *Classification and Regression Trees*. Monterey, CA: Wadsworth International Group.
- S. Brin, R. Motwani, J.D. Ullman, & S. Tsur (1997). Dynamic itemset counting and implication rules for market basket data. In *Proceedings of the ACM SIGMOD Conference*.
- P. Clark and S. Matwin (1993). Using qualitative models to guide induction learning. In *Proceedings of the International Conference on Machine Learning*.
- M. Dash & H. Liu (1997). Feature selection methods for classification. *Intelligent Data Analysis: An International Journal*, 1.
- M. Dash, H. Liu, & J. Yao (1997). Dimensionality reduction of unsupervised data. In *Proceedings of 9th IEEE International Conference on Tools with AI (ICTAI)*, 532-539.

- J. Devore & R. Peck (1997). *Statistics: The Exploration and Analysis of Data*. New York: Duxbury Press.
- W.J. Frawley, G. Piatetsky-Shapiro, & C.J. Matheus (1991). Knowledge discovery in databases: an overview. In G.Piatetsky-Shapiro and W.J. Frawley (Eds.), *Knowledge Discovery in Databases*, AAAI/MIT Press.
- J.H. Friedman (1977). A recursive partitioning decision rule for nonparametric classifiers. *IEEE Transactions on Computing*, 26, 404-408.
- J. Han & M. Kamber (2001). *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann.
- M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, & I. Witten (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, 11(1), 10-18.
- G.H. John & P. Langley (1996). Static versus dynamic sampling for data mining. In *Proceedings 1996 International Conference on Knowledge Discovery and Data Mining (KDD'96)*, 367-370.
- K. Josien, G. Wang, T.W. Liao, E. Triantaphyllou, & M.C. Liu (2001). An evaluation of sampling methods for data mining with fuzzy c-means. In *Data Mining for Design and Manufacturing*, Chapter 15, 351-365. Kluwer Academic Publishers.
- R.L. Kennedy, Y. Lee, B. Van Roy, C.D. Reed, & R.P. Lippman (1998). *Solving Data Mining Problems Through Pattern Recognition*. Upper Saddle River, NJ:Prentice Hall.
- J. Kivinen & H. Mannila (1994). The power of sampling in knowledge discovery. In *Proceedings 13th ACM Symposium on Principles of Database Systems*, 77-85.
- M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A.I. Verkamo (1994). Finding interesting rules from large sets of discovered association rules. In *Proceedings of the Third International Conference on Information and Knowledge Management*.
- R. Kohavi & G.H. John (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97, 273-324.
- P. Langley, W. Iba, & K. Thompson (1992). An analysis of bayesian classifiers. *Proceedings of the 10th National Conference on Artificial Intelligence*, 223-228.