

การคัดเลือกตัวแปรโดยการหาค่าเอยู่ซีที่เหมาะสมที่สุด

วรัญญู วงษ์เสรี, ปวีศ ธาริชาญ

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

บทคัดย่อ

เอยู่ซีเป็นเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของตัวจำแนก การวิเคราะห์ทางสถิติหาความสัมพันธ์ระหว่างเอยู่ซีและอัตราผิดพลาด (Error rate) พบว่ากรณีที่คลาสไม่สมดุลที่มีอัตราผิดพลาดสูง ตัวจำแนกที่มีความถูกต้อง (Accuracy) สูงอาจจะไม่ได้มีค่าเอยู่ซีสูง เนื่องจากความถูกต้องจะแปรผันตามจำนวนตัวอย่างที่จำแนกผิดพลาด ในขณะที่เอยู่ซีนั้นจะแปรผันตามจำนวนตัวอย่างที่จำแนกผิดพลาดแล้วยังแปรผันตามลำดับ (Rank) ของตัวอย่างที่จำแนกผิดพลาดด้วย งานวิจัยนี้นำเสนอการทดลองเปรียบเทียบความสอดคล้องและความสามารถในการจำแนกระหว่างดัชนีค่าความถูกต้องและค่า auc พบว่า auc มีความสอดคล้อง สูงกว่าความถูกต้อง ทั้งในกรณีที่คลาสสมดุลและไม่สมดุล นอกจากนี้ตัวจำแนกที่มีค่าเอยู่ซีสูงมีแนวโน้มที่จะมีความถูกต้องสูงด้วย การออกแบบตัวจำแนกที่มีค่าเอยู่ซีสูง จึงมีความเหมาะสมมากกว่าตัวจำแนกที่มีความถูกต้องสูง ดังนั้นจึงมีความจำเป็นในการหาฟังก์ชันความสูญเสีย (Loss function) ที่เหมาะสมสำหรับการหาค่าเอยู่ซีที่เหมาะสมที่สุด

1 บทนำ

เป้าหมายของขั้นตอนวิธีการเรียนรู้สำหรับปัญหาการจำแนกคือการสร้างตัวจำแนกจากชุดข้อมูลที่มีป้ายกำกับเพื่อให้แบบจำลองสามารถใช้ ในการพยากรณ์ชุดข้อมูลทดสอบ โดยทั่วไปความสามารถในการทำนายของขั้นตอนวิธีการเรียนรู้สำหรับปัญหาการจำแนกสามารถวัดได้จากค่าความแม่นยำ (หรือ อัตราผิดพลาดซึ่งเท่ากับ 1 ลบด้วยค่าความแม่นยำ) ของชุดข้อมูลทดสอบ และโดยส่วนใหญ่ของแบบจำลองการจำแนกนั้นสามารถประมาณค่าความน่าจะเป็นของการเกิดคลาส ได้แต่ยังไม่ค่อยนำมาประเมินประสิทธิภาพของแบบจำลองเท่าไร ทำให้ความแม่นยำถูกพิจารณาเพียงถูกต้องหรือผิดพลาดเพียงเท่านั้น ค่าความแม่นยำนั้นอาจไม่เพียงพอในการประเมินประสิทธิภาพของแบบจำลองการจำแนก เช่น ในทางการตลาด ที่ต้องการกระตุ้นยอดขายสูงสุดให้เพิ่มขึ้นจากลูกค้า จึงทำให้ต้องการดำเนินการกลยุทธ์ทางการค้าต่อลูกค้าที่ส่งผลมากที่สุดต่อการขายในแต่ละบุคคล มิใช่เพียงแค่การดำเนินการกลยุทธ์ต่อลูกค้าที่เพียงสนใจแค่ว่าจะทำให้ลูกค้าซื้อหรือไม่เท่านั้น เพราะต้องการที่จะเพิ่มโอกาสการซื้อของลูกค้าให้เกิดผลสูงสุด ดังนั้นในกรณีนี้เพียงแค่เพิ่มโอกาสการซื้อของลูกค้านั้นไม่เพียงพอ แต่ต้องเป็นวิธีที่เพิ่มโอกาสการซื้อของลูกค้าได้มากที่สุดด้วย

ดังนั้นการจัดอันดับจึงเป็นที่ต้องการมากกว่าแค่การจัดประเภท และสามารถคำนวณได้ง่ายเนื่องจากแบบจำลองการจำแนกส่วนใหญ่จะสร้างการประมาณความน่าจะเป็นที่สามารถใช้ในการจัดอันดับได้

เส้นโค้ง ROC (Receiver Operating Characteristics) ได้ถูกนำมา

ใช้ในการประเมินประสิทธิภาพการจัดอันดับของขั้นตอนวิธีการเรียนรู้สำหรับปัญหาการจำแนกโดยพบว่า AUC มีคุณสมบัติที่พึงประสงค์หลายประการเมื่อเทียบกับความแม่นยำ ในบทความนี้จะแสดงให้เห็นในเชิงประจักษ์ว่า AUC เป็นการประเมินประสิทธิภาพแบบจำลองที่ดีกว่าความแม่นยำ

2 เกณฑ์ สำหรับการ เปรียบ เทียบ มาตรการ การ ประเมิน

เริ่มต้นด้วยการเปรียบเทียบ AUC และความแม่นยำจากนั้นอธิบายคำจำกัดความที่เป็นทางการในการเปรียบเทียบตัวประเมินประสิทธิภาพแบบจำลองการจำแนกทั้งสองประเภท

2.1 ค่า AUC เทียบกับ ค่าความแม่นยำ

การคำนวณ AUC สามารถคำนวณได้จาก[1]

$$AUC = \frac{\sum r_i - n_p \frac{n_p + 1}{2}}{n_p n_n} \quad (1)$$

ตารางที่ 1 ตัวอย่างข้อมูลการคำนวณ AUC

| | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|----|
| | - | - | - | - | + | - | + | + | + | + |
| i | | | | | 1 | | 2 | 3 | 4 | 5 |
| r_i | | | | | 5 | | 7 | 8 | 9 | 10 |

ตารางที่ 2 ตัวอย่างแบบจำลองการจำแนกทั้งสองที่มีค่าความแม่นยำเท่ากัน แต่มี AUC ต่างกัน

| | | | | | | | | | | |
|---------------|---|---|---|---|---|---|---|---|---|---|
| ตัวจำแนกที่ 1 | - | - | - | - | + | - | + | + | + | + |
| ตัวจำแนกที่ 2 | + | - | - | - | - | + | + | + | + | - |

เมื่อ n_p และ n_n คือจำนวนตัวอย่างทั้งหมดของคลาสบวก และคลาสลบตามลำดับ และ r_i คือหมายเลขอันดับของคลาสบวกที่ i จากตัวอย่างในตารางที่ 1 พบว่ามีคลาสบวกและ คลาสลบอยู่อย่างละ 5 ตัว และเมื่อคำนวณหาค่า AUC จะได้ดังนี้ $\frac{(5+7+8+9+10)-5 \times \frac{6}{2}}{5 \times 5}$ ซึ่งเท่ากับ $\frac{24}{25}$ โดยค่าสูงสุดของ AUC จะมีค่าเท่ากับ 1 ในตัวอย่างถัดไปจะเห็นว่าเหตุใด AUC จึงเป็นหน่วยวัดที่ดีกว่าความแม่นยำ

พิจารณาแบบจำลองการจำแนก 2 แบบจำลองที่มีการประมาณความน่าจะเป็นสำหรับชุดตัวอย่างการทดสอบ 10 ชุด โดยเป็นคลาสบวกและ คลาสลบอย่างละ 5 ตัว ซึ่งเห็นได้ชัดเจนว่าการจำแนกทั้ง 2 ตัว มีค่าความแม่นยำ

เท่ากับ 80% (จำแนกคลาสบวกถูกต้อง(จริงบวก) 4 ตัว จำแนกคลาสลบถูกต้อง(จริงลบ) 4 ตัวและ จำแนกคลาสบวกผิด(เท็จลบ) 1 ตัว จำแนกคลาสลบผิด(เท็จบวก) 1 ตัว รวมถูกต้องทั้งหมด 8 ตัวจาก 10 ตัว) แต่ค่า AUC ของตัวจำแนกที่ 1 และ 2 นั้นเท่ากับ $\frac{24}{25}$ และ $\frac{16}{25}$ ตามลำดับ พบว่าความแม่นยำไม่สามารถแยกความแตกต่างของทั้งสองแบบจำลองได้ ในขณะที่ค่า AUC สามารถแยกความแตกต่างของทั้งสองแบบจำลองได้

2.2 ความสอดคล้อง (Consistency) และความสามารถในการจำแนก (Discriminancy)

เมื่อต้องเปรียบเทียบประสิทธิภาพของแบบจำลองที่แตกต่างกันดัชนีหนึ่งที่เราควรคำนึงถึงคือความสอดคล้อง เพื่อระบุว่าแบบจำลองที่เปรียบเทียบกันนั้นจะมีการทำงานหรือเปลี่ยนแปลงไปในทิศทางเดียวกัน หรือกลับกัน หรือไม่

อีกดัชนีหนึ่งความสามารถในการจำแนก (Discriminancy) ความสามารถในการแยกรูปแบบที่แตกต่างกัน ดัชนีในการประเมินประสิทธิภาพแบบใดสามารถจำแนกสูงกว่าอีกแบบนั้นจะต้องมีเหตุการณ์ที่การประเมินประสิทธิภาพหนึ่งไม่สามารถแยกรูปแบบสองชุดข้อมูลที่มีความต่างกันได้ แต่อีกการประเมินประสิทธิภาพสามารถทำได้

3 ทดลองเปรียบเทียบ

การทดลองเพื่อเปรียบเทียบประสิทธิภาพของดัชนีความถูกต้องและ auc จะทดลองด้วยข้อมูลที่สามารถเกิดขึ้นในทุกกรณีโดยวิธีการเรียงสับเปลี่ยนทางคณิตศาสตร์ โดยข้อมูลจะมีทั้งหมดสองคลาสกำหนดให้เป็นคลาสบวกและ คลาสลบการทดลองจะแบ่งเป็นสองกรณีคือ ข้อมูลที่สมดุลกันและ ไม่สมดุลกัน

3.1 ข้อมูลสองคลาสที่สมดุล

ดังนั้นชุดข้อมูลในการทดลองนี้จะประกอบด้วยตัวอย่างบวกและลบจำนวนเท่ากัน (binary class) โดยจะทดลองข้อมูลที่มีขนาด 4, 6, 8, 10, 12, 14, 16, 18 และ 20 ตัวอย่าง

โดยเมื่อข้อมูลมีขนาด $2n$ จะมีรูปแบบที่เป็นไปได้ทั้งหมด $\binom{2n}{n}$ และรูปแบบที่สนใจทั้งหมดนี้ให้ a, b เป็นรูปแบบชุดข้อมูลที่แตกต่างกัน ถ้า $AUC(a) > AUC(b)$ และ $acc(a) > acc(b)$ ด้วยจะนับว่าค่า AUC มีความสอดคล้องกับค่าความแม่นยำ แต่ถ้า $AUC(a) > AUC(b)$ แต่ $acc(a) < acc(b)$ จะถูกนับว่าค่า AUC ไม่มีความสอดคล้องกับค่าความแม่นยำ โดยให้ค่าความสอดคล้อง C คิดได้จากเหตุการณ์ที่ค่า AUC และค่าความแม่นยำ เป็นไปในทิศทางเดียวกัน หากด้วยผลรวมของทั้งสองเหตุการณ์ ผลลัพธ์จากการทดลองแสดงดังตารางที่ 3

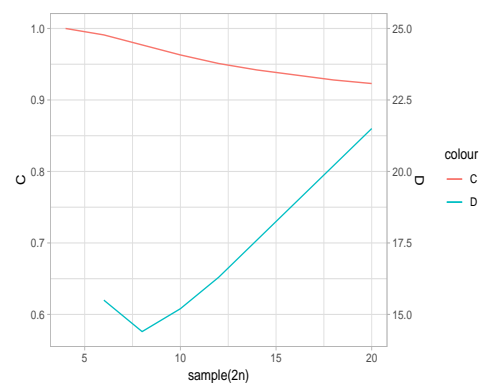
ถ้า $acc(a) = acc(b)$ แต่ $AUC(a) \neq AUC(b)$ ด้วยจะนับว่าค่า AUC นั้นมีความสามารถในการจำแนกสูงกว่าค่าความแม่นยำ แต่เมื่อ $AUC(a) = AUC(b)$ และ $acc(a) \neq acc(b)$ จะถูกนับว่าค่าความแม่นยำ นั้นมีความสามารถในการจำแนกสูงกว่าค่า AUC ความสามารถในการจำแนก D คิดได้จากเหตุการณ์ที่ AUC มีความสามารถในการจำแนกสูงกว่าค่าความแม่นยำ หากด้วย เหตุการณ์ที่ค่าความแม่นยำ มีความสามารถในการจำแนกสูงกว่า AUC ผลลัพธ์จากการทดลองแสดงดังตารางที่ 4

ตารางที่ 3 ความสอดคล้องกันของ AUC และ ความแม่นยำ

| $2n$ | ค่า AUC และ ค่าความแม่นยำ สอดคล้องกัน | ค่า AUC และ ค่าความแม่นยำ ไม่สอดคล้องกัน | C |
|------|---|--|-------|
| 4 | 9 | 0 | 1 |
| 6 | 113 | 1 | 0.991 |
| 8 | 1,459 | 34 | 0.977 |
| 10 | 19,742 | 766 | 0.963 |
| 12 | 273,600 | 13,997 | 0.951 |
| 14 | 3,864,673 | 237,303 | 0.942 |
| 16 | 55,370,122 | 3,868,959 | 0.935 |
| 18 | 802,343,521 | 61,797,523 | 0.928 |
| 20 | 11,733,729,456 | 975,464,160 | 0.923 |

ตารางที่ 4 ความสามารถในการจำแนกของค่า AUC และ ความแม่นยำ

| $2n$ | ค่า AUC จำแนกได้ดีกว่า ค่าความแม่นยำ | ค่าความแม่นยำ จำแนกได้ดีกว่า ค่า AUC | D |
|------|--|--|----------|
| 4 | 5 | 0 | ∞ |
| 6 | 62 | 4 | 15.5 |
| 8 | 762 | 52 | 14.4 |
| 10 | 9,416 | 618 | 15.2 |
| 12 | 120,374 | 7,369 | 16.3 |
| 14 | 1,578,566 | 89,828 | 17.6 |
| 16 | 21,161,143 | 1,121,120 | 18.9 |
| 18 | 288,745,778 | 14,290,466 | 20.2 |
| 20 | 3,998,425,154 | 185,536,518 | 21.5 |



รูปที่ 1 ค่าความสอดคล้องและค่าความสามารถในการจำแนก เทียบจำนวนตัวอย่าง

ในการทดลองที่ข้อมูลเป็นมีจำนวนคลาสสองคลาสและเป็นข้อมูลที่สมดุลพบว่า AUC นั้นมีความสอดคล้องกับ ค่าความแม่นยำและ AUC นั้นสามารถจำแนกเหตุการณ์ที่แตกต่างกันที่ค่าความแม่นยำไม่สามารถจำแนกได้มากกว่า และเมื่อพิจารณาความสามารถในการจำแนกนั้นพบว่ายิ่งจำนวนข้อมูลเยอะมากขึ้นนั้น ความสามารถในการจำแนกของ AUC จะสูงขึ้นด้วย แสดงดังรูปที่

3.2 ข้อมูลสองคลาสที่สมดุล

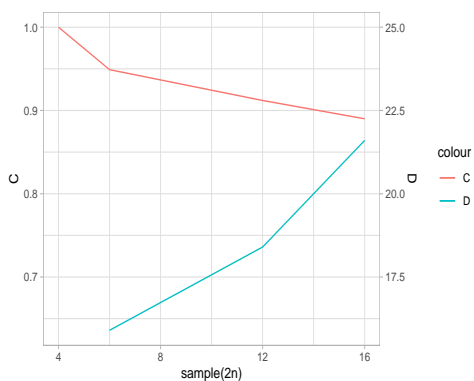
ข้อมูลแบบอาร์คาสที่สมดุลโดยจะกำหนดให้มีตัวอย่างคลาสบวก 25% และตัวอย่างคลาสลบ 75% โดยข้อมูลที่ให้จะมีจำนวน 4, 8, 12 และ 16 ตัวอย่าง และยังคงใช้สูตรการคำนวณค่า AUC เหมือนเดิมและเนื่องจากจำนวนตัวอย่างนั้นไม่สมดุลทำให้การคำนวณค่าความแม่นยำจะเปลี่ยนจากเดิมที่ 5 ตัวอย่างแรกเป็นคลาสลบและ 5 ตัวอย่างถัดไปเป็นคลาสบวกหรืออีกนัยหนึ่งคือแบ่งตรงกลางอย่างละครึ่ง แต่เมื่อข้อมูลนั้นมีขนาดไม่เท่ากันทำให้การแยกคลาสบวกและคลาสลบเป็น 75% แรกเป็นคลาสลบ และ 25% เป็นต่อมาเป็นคลาสบวกตามอัตราส่วนของข้อมูลเข้าที่เปลี่ยนไป

ตารางที่ 5 ความสอดคล้องกันของ AUC และ ความแม่นยำ(ไม่สมดุล)

| $2n$ | ค่า AUC และ ค่าความแม่นยำ สอดคล้องกัน | ค่า AUC และ ค่าความแม่นยำ ไม่สอดคล้องกัน | C |
|------|---|--|-------|
| 4 | 3 | 0 | 1 |
| 6 | 187 | 10 | 0.949 |
| 12 | 12,716 | 1,225 | 0.912 |
| 16 | 926,884 | 114,074 | 0.890 |

ตารางที่ 6 ความสามารถในการจำแนกของค่า AUC และ ความแม่นยำ(ไม่สมดุล)

| $2n$ | ค่า AUC จำแนกได้ดีกว่า ค่าความแม่นยำ | ค่าความแม่นยำ จำแนกได้ดีกว่า ค่า AUC | D |
|------|--|--|----------|
| 4 | 3 | 0 | ∞ |
| 8 | 159 | 10 | 15.9 |
| 12 | 8,986 | 489 | 18.4 |
| 16 | 559,751 | 25,969 | 21.6 |



รูปที่ 2 ค่าความสอดคล้องและ ค่าความสามารถในการจำแนก เทียบจำนวนตัวอย่าง กรณีข้อมูลไม่สมดุล

ตารางที่ 7 ความสอดคล้องและความสามารถในการจำแนกของค่า AUC และ ความแม่นยำ(ไม่สมดุล ขนาดข้อมูล 10 ตัวอย่าง)

| คลาสบวก | คลาสลบ | C | D |
|---------|--------|-------|----------|
| 1 | 9 | 1.0 | ∞ |
| 2 | 8 | 0.926 | 22.3 |
| 3 | 7 | 0.939 | 15.5 |
| 4 | 6 | 0.956 | 14.9 |
| 5 | 5 | 0.963 | 15.2 |

และสุดท้ายเป็นการทดลองในหลายๆ อัตราส่วนของคลาสบวกและคลาสลบ โดยกำหนดให้มีข้อมูลทั้งหมด 10 ตัวอย่าง โดยเริ่มจากสมดุลคือมีทั้งหมดอย่างละ 5 ตัวอย่างจากนั้นเพิ่มและ ลดคลาสใดคลาสหนึ่งไปเรื่อยๆ จนไม่สามารถลดได้ ในกรณีนี้คือเหลือตัวเดียว

จากการทดลองทั้งสองไม่ว่าเป็นข้อมูลทั้งแบบที่สมดุลและ ไม่สมดุลก็ตาม ผลการทดลองยังคงเป็นไปในทิศทางเดียวกัน ทั้งในมุมมองสอดคล้องที่ยังคงสอดคล้องกันสูง และในมุมมองความสามารถในการจำแนกที่ AUC มีความสามารถในการจำแนกสูง ขึ้นเรื่อยๆ ตามขนาดของ ข้อมูล และ ยังมีความสามารถในการจำแนกสูงมากขึ้นเมื่อข้อมูลเกิดความไม่สมดุลของทั้งสองคลาส

4 การประยุกต์ใช้

จากการทดลองที่ผ่านมาได้เปรียบเทียบตัวประเมินประสิทธิภาพทั้งสองคือค่า AUC และ ค่าความแม่นยำ โดยค่า AUC มีประสิทธิภาพดีกว่าค่าความแม่นยำ แต่อย่างไรก็ตามในการใช้งานจริงทั้ง AUC และ ความแม่นยำไม่ใช่เป้าหมายสุดท้าย เช่น ธนาคาร หรือ บริษัทประกันภัย อาจจะมีข้อมูลของลูกค้าอยู่มหาศาลโดยสิ่งที่ต้องการสุดท้ายคือการคาดการณ์การทำให้กับ บริษัท

สมมติว่าลูกค้าที่ถูกเก็บในฐานะข้อมูลมีการเก็บด้วยแอตทริบิวต์จำนวนหนึ่ง และ ลูกค้าแต่ละ ราย อาจ เป็นผู้ซื้อ หรือ ไม่ใช่ ผู้ซื้อ ผลลัพธ์ บาง อย่างเนื่องจากปัญหานี้เป็นปัญหาการจำแนกแบบไบนารี ลูกค้าจะได้รับการติดต่อจากแคมเปญการส่งเสริมการขายสำหรับลูกค้าแต่ละรายโดย บริษัทต้องคาดการณ์ว่าในสินค้าชนิดๆ หนึ่งนั้นลูกค้าแต่ละรายมีความต้องการสินค้านั้นมากเพียงใด และ ต้องเพิ่มโอกาสการซื้อ มากน้อยเพียงใด

อย่างไรก็ตามการประยุกต์ใช้ บริษัท อาจต้องการโปรโมตเพียงเล็กน้อยให้กับลูกค้าที่มีแนวโน้มสูงที่สุดที่คาดการณ์ไว้ และต้องโปรโมตมากขึ้นสำหรับลูกค้าที่มีแนวโน้มลดลง ซึ่งทำให้กำไรที่ได้ต่อลูกค้าแต่ละคนนั้นต่างกันไปด้วยซึ่งในความเป็นจริง ก่อให้เกิดผลดีต่อรายได้ของบริษัทเพราะสามารถลดการโปรโมตเกินจำเป็นสำหรับลูกค้าที่มีแนวโน้มจะซื้อสินค้าสูงๆ อยู่แล้ว เช่น ลูกค้าที่มีแนวโน้มจะซื้อสินค้าสูงสุด 10% แรกนั้นอาจจะเป็นลูกค้าที่มีการซื้อสินค้าเป็นประจำในการโปรโมตสินค้าที่ลูกค้ากลุ่มนี้ซื้อเป็นประจำอยู่แล้วอาจไม่จำเป็น และเพิ่มโอกาสให้ลูกค้าที่มีแนวโน้มลดลงมีโอกาสซื้อสินค้ามากขึ้นด้วย

5 สรุป

ใน บทความ นี้ ได้ ให้ คำ จำกั ด ความ อย่าง เป็น ทาง การ เกี่ยว กับ ความ สอดคล้องและ ความสามารถในการจำแนก เพื่อใช้ประเมินผลสำหรับขั้นตอนวิธีการเรียนรู้ในปัญหาการจำแนก กำหนดรูปแบบและเกณฑ์ที่ใช้สำหรับการเปรียบเทียบตัวประเมินประสิทธิภาพทั้งสอง และแสดงให้เห็นอย่างชัดเจนว่า AUC นั้นเป็นตัวประเมินประสิทธิภาพที่ดีกว่าค่าความแม่นยำ และได้นำไปเปรียบเทียบกับเหตุการณ์จริงในธุรกิจเพื่อแสดงผลลัพธ์ที่น่าสนใจว่า AUC เกี่ยข้องโดยตรงกับกำไรสุทธิมากกว่าค่าความแม่นยำในการตลาดทางตรง

การเพิ่ม ประสิทธิภาพ AUC นั้น เป็น ที่ ต้องการ มากกว่า การเพิ่ม ความแม่นยำในการนำขั้นตอนวิธีการเรียนรู้สำหรับปัญหาการจำแนก รวมถึงการทำเหมืองข้อมูลที่จะนำไปประยุกต์ใช้ในโลกความเป็นจริง

ข้อมูลอ้างอิง

- [1] D.J. Hand and R.J. Till. A simple generalisation of the area under the ROC curve for multipleclass classification problems.

Machine Learning, 45:171–186, 2001.

- [2] Ling, Charles X and Huang, Jin and Zhang, Harry. *AUC: a Statistically Consistent and more Discriminating Measure than Accuracy*. Fredericton, NB, Canada
- [3] Jin Huang and Ling, C.X. *Using AUC and accuracy in evaluating learning algorithms*. London, Ontario, Canada