

การคัดเลือกตัวแปรโดยการหาค่าเอยู่ซีที่เหมาะสมที่สุด

ผศ.ดร. วรณัฐ วงษ์เสรี, นาย ปวีศ ธารีชาญ

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

บทคัดย่อ

ในปัจจุบันเอยู่ซีเป็นเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของตัวจำแนก มีการวิเคราะห์ทางสถิติหาความสัมพันธ์ระหว่างเอยู่ซีและอัตราผิดพลาด (Error rate) พบว่ากรณีทีคลาสไม่สมดุลที่มีอัตราผิดพลาดสูง ตัวจำแนกที่มีความถูกต้อง (Accuracy) สูงอาจจะไม่ได้มีค่าเอยู่ซีสูง เนื่องจากความถูกต้องจะแปรผันตามจำนวนตัวอย่างที่จำแนกผิดพลาด ในขณะที่เอยู่ซีนอกจากจะแปรผันตามจำนวนตัวอย่างที่จำแนกผิดพลาดแล้วยังแปรผันตามลำดับ (Rank) ของตัวอย่างที่จำแนกผิดพลาดด้วย และมีการพิสูจน์เชิงทฤษฎีและทดลองพบว่า เอยู่ซีมีความสอดคล้อง (Consistency) และความสามารถในการจำแนก (Discriminancy) สูงกว่าความถูกต้องทั้งในกรณีทีคลาสสมดุลและไม่สมดุล นอกจากนี้ตัวจำแนกที่มีค่าเอยู่ซีสูงมีแนวโน้มที่จะมีความถูกต้องสูงด้วย การออกแบบตัวจำแนกที่มีค่าเอยู่ซีสูงจึงมีความเหมาะสมมากกว่าตัวจำแนกที่มีความถูกต้องสูง ดังนั้นจึงมีความจำเป็นในการหาฟังก์ชันความสูญเสีย (Loss function) ที่เหมาะสมสำหรับการหาค่าเอยู่ซีที่เหมาะสมที่สุด

1 บทนำ

เป้าหมายของขั้นตอนวิธีการเรียนรู้สำหรับปัญหาการจำแนกคือการสร้างตัวจำแนกจากชุดข้อมูลที่มีป้ายกำกับเพื่อให้แบบจำลองสามารถใช้ลักษณะนั้นๆ ในการคาดเดาชุดข้อมูลที่ใช้ทดสอบที่ไม่เคยพบมาก่อน โดยทั่วไปความสามารถในการทำนายของขั้นตอนวิธีการเรียนรู้สำหรับปัญหาการจำแนกสามารถวัดได้จากค่าความแม่นยำ (หรือ อัตราผิดพลาดซึ่งเท่ากับ 1 ลบด้วยค่าความแม่นยำ) ของชุดข้อมูลที่ใช้ทดสอบ และโดยส่วนใหญ่ของแบบจำลองการจำแนกนั้นสามารถประมาณค่าความน่าจะเป็นของการเกิดคลาสนั้นๆ ได้แต่น่าเสียดายเนื่องจากมักไม่ค่อยเป็นที่สนใจเท่าไรทำให้

ความแม่นยำถูกพิจารณาเพียงถูกต้องหรือผิดพลาดเพียงเท่านั้น โดยไม่นำความน่าจะเป็น (เช่น 0.51 หรือ 0.7 หรือ 0.99) ของการทำนายมาพิจารณาด้วย

จึงทำให้ค่าความแม่นยำนั้นอาจไม่เพียงพอในการประเมินประสิทธิภาพของแบบจำลองการจำแนก เช่นตัวอย่างในทางการตลาด เราล้วนต้องการกระตุ้นยอดขายสูงสุดให้เพิ่มขึ้นมากที่สุด $X\%$ จากลูกค้า จึงทำให้ต้องการดำเนินกลยุทธ์ทางการค้าต่อลูกค้าที่ส่งผลมากที่สุดต่อการขายในแต่ละบุคคล มิใช่เพียงแค่การดำเนินกลยุทธ์ต่อลูกค้าที่เพียงสนใจแค่จะทำให้ลูกค้าซื้อหรือไม่เท่านั้น เพราะเราต้องการที่จะเพิ่มโอกาสการซื้อของลูกค้าให้เกิดผลสูงสุด ดังนั้นในกรณีนี้เพียงแค่เพิ่มโอกาสการซื้อของลูกค้านั้นไม่เพียงพอ แต่ต้องเป็นวิธีที่เพิ่มโอกาสการซื้อของลูกค้าได้มากที่สุดด้วย

ดังนั้นการจัดอันดับจึงเป็นที่ต้องการมากกว่าแค่การจัดประเภท และสามารถทำได้ง่ายเนื่องจากแบบจำลองการจำแนกส่วนใหญ่จะสร้างการประมาณความน่าจะเป็นที่สามารถใช้ในการจัดอันดับได้

เส้นโค้ง ROC (Receiver Operating Characteristics) ปัจจุบันได้ถูกนำมาใช้ในการประเมินประสิทธิภาพการจัดอันดับของขั้นตอนวิธีการเรียนรู้สำหรับปัญหาการจำแนกโดยพบว่า AUC มีคุณสมบัติที่พึงประสงค์หลายประการเมื่อเทียบกับความแม่นยำ แต่ยังไม่ชัดเจนว่า AUC เป็นตัวประเมินประสิทธิภาพแบบจำลองที่ดีกว่าความแม่นยำ โดยทั่วไปเราจะเปรียบเทียบการวัดผลของการประเมินประสิทธิภาพแบบจำลองได้อย่างไร และจะสามารถบอกได้อย่างไรว่าการประเมินประสิทธิภาพแบบจำลองแบบใดดีกว่าแบบอื่นๆ

โดยในบทความนี้จะให้คำจำกัดความอย่างเป็นทางการเกี่ยวกับความสอดคล้อง และ ความสามารถในการจำแนก โดยจะแสดงให้เห็นทั้งในเชิงประจักษ์และเป็นทางการว่า AUC เป็นการประเมินประสิทธิภาพแบบจำลองที่ดีกว่าความแม่นยำ

2 เกณฑ์สำหรับการเปรียบเทียบมาตรการการประเมิน

เริ่มต้นด้วยการเปรียบเทียบ AUC และความแม่นยำ จากนั้นเราจึงนำเสนอ คำจำกัดความ ที่เป็นทางการ ในการ เปรียบ เทียบ ตัว ประเมิน ประสิทธิภาพแบบจำลองการจำแนกทั้งสองประเภท

2.1 ค่า AUC เทียบกับ ค่าความแม่นยำ

การคำนวณ AUC สามารถคำนวณได้จาก

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

สมการที่ 1 สูตรการคำนวณ AUC

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

ตารางที่ 1 ตัวอย่างการคำนวณ AUC

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

ตารางที่ 2 ตัวอย่างแบบจำลองการจำแนกทั้งสองที่มีค่าความแม่นยำเท่ากัน แต่มี AUC ต่างกัน

เมื่อ n_p และ n_n คือจำนวนทั้งหมดของคลาสบวก และคลาสลบตามลำดับ และ r_i และ i คือหมายเลขอันดับของคลาสบวกเทียบทั้งหมดและหมายเลขลำดับคลาสบวก ตามลำดับ จากตัวอย่างในตารางที่ 1 จะเห็นได้ว่ามีคลาสบวกและ คลาสลบอยู่อย่างละ 5 ตัว และเมื่อคำนวณหาค่า AUC จะได้ดังนี้ $\frac{(5+7+8+9+10)-5 \times \frac{5}{2}}{5 \times 5}$ ซึ่งเท่ากับ $\frac{24}{25}$ โดยค่าสูงสุดของ AUC จะมีค่าเท่ากับ 1 ในตัวอย่างถัดไปจะเห็นว่าเหตุใด AUC จึงเป็นหน่วยวัดที่ดีกว่าความแม่นยำจากตัวอย่างต่อไปนี้

พิจารณาการจำแนกทั้ง 2 ตัวซึ่งทั้งสองมีการประมาณความน่าจะเป็นสำหรับชุดตัวอย่างการทดสอบ 10 ชุด โดยเป็นคลาสบวกและ คลาสลบอย่างละ 5 ตัว ซึ่งเห็นได้ชัดแจ้งว่าการจำแนกทั้ง 2 ตัวมีค่าความแม่นยำเท่ากับ 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

2.2 ความสอดคล้อง (Consistency) และความสามารถในการจำแนก (Discriminancy)

ความสอดคล้อง (Consistency) เมื่อพูดถึงการประเมินประสิทธิภาพทั้งสองแบบที่แตกต่างกันหนึ่งในสิ่งที่ควรคำนึงถึงคือความสอดคล้อง เพื่อระบุว่าสิ่งที่กำลังเปรียบเทียบกันนั้นจะมีการทำงานหรือ เปลี่ยนแปลงไปในทิศทางเดียวกัน หรือกลับกัน หรือไม่

ความสามารถในการจำแนก (Discriminancy) ความสามารถในการแยกรูปแบบที่แตกต่างกัน หากจะบอกว่าการประเมินประสิทธิภาพแบบใดมีความสามารถจำแนกสูงกว่าอีกแบบ นั้นจะต้องมีเหตุการณ์ที่การ ประเมิน ประสิทธิภาพหนึ่งไม่สามารถแยกรูปแบบสองชุดข้อมูลที่มีความต่างกันได้ แต่อีกการประเมินประสิทธิภาพสามารถทำได้

3 ทดลองเปรียบเทียบ

ในส่วนนี้จะแสดงให้เห็นอย่างชัดเจนบนข้อมูลที่จำลองขึ้นในทุกกรณี โดยวิธีการเรียงสับเปลี่ยนทางคณิตศาสตร์ โดยข้อมูลจะมีทั้งหมดสองคลาส กำหนดให้เป็นคลาสบวกและ คลาสลบการทดลองจะแบ่งเป็นสองกรณีคือ ข้อมูลที่สมดุลกันและ ไม่สมดุลกัน เพราะ ชุดข้อมูลในโลกแห่งความจริงส่วนใหญ่ไม่สมดุล

3.1 ข้อมูลสองคลาสที่สมดุล

ดังนั้นชุดข้อมูลในการทดลองนี้จึงสมดุลกับตัวอย่างบวกและลบจำนวนเท่ากัน (binary class) โดยการทดลองจะทดลองข้อมูลที่มีขนาด 4, 6, 8, 10, 12, 14, 16, 18 และ 20 ข้อมูล

โดยเมื่อข้อมูลมีขนาด $2n$ จะมีรูปแบบที่เป็นไปได้ทั้งหมด $\binom{2n}{n}$ และรูปแบบที่สนใจทั้งหมดมีดังนี้ ให้ a, b เป็นรูปแบบชุดข้อมูลที่แตกต่างกัน ถ้า $AUC(a) > AUC(b)$ และ $acc(a) > acc(b)$ ด้วย จะนับว่าค่า AUC มีความสอดคล้องกับ ค่าความแม่นยำ แต่เมื่อ $AUC(a) > AUC(b)$ แต่ $acc(a) < acc(b)$ จะถูกนับว่าค่า AUC ไม่มีความสอดคล้องกับค่าความแม่นยำ โดยให้ค่าความสอดคล้อง C คิดได้จากเหตุการณ์ที่ค่า AUC และค่าความแม่นยำ เป็นไปในทิศทางเดียวกัน ทารด้วย ผลรวมของทั้งสองเหตุการณ์

ถ้า $acc(a) = acc(b)$ แต่ $AUC(a) \neq AUC(b)$ ด้วย จะนับว่าค่า 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

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

ตารางที่3 ความสอดคล้องกันของ 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

ตารางที่4 ความสามารถในการจำแนกของค่า AUC และ ความแม่นยำ

ในการทดลองที่ข้อมูลเป็นมีจำนวนคลาสสองคลาสและ เป็นข้อมูลที่สมดุล จะเห็นได้อย่างชัดเจนว่า AUC นั้นมีความสอดคล้องกับ ค่าความแม่นยำ และ AUC นั้นสามารถจำแนกเหตุการณ์ที่แตกต่างกันที่ค่าความแม่นยำไม่สามารถจำแนกได้มากกว่า และ หากมองความสามารถในการจำแนกนั้นจะเห็นได้ว่ายิ่งจำนวนข้อมูลเยอะมากขึ้นนั้น ความสามารถในการจำแนกของ AUC จะดีขึ้นตามไปด้วย

3.2 ข้อมูลสองคลาสที่สมดุล

ข้อมูลไบนารี คลาสที่ไม่สมดุลโดยจะ กำหนดให้มี ตัวอย่าง คลาสบวก 25% และตัวอย่างคลาสลบ 75% โดยข้อมูลที่ใช้จะมีจำนวน 4, 8, 12 และ 16 ตัวอย่างและยังคงใช้สูตรการคำนวณค่า AUC เหมือนเดิมและเนื่องจากจำนวนตัวอย่างนั้นไม่สมดุลทำให้การคำนวณค่าความแม่นยำจะเปลี่ยนจากเดิมที่ 5 ตัวอย่างแรกเป็นคลาสลบและ 5 ตัวอย่างถัดไปเป็นคลาสบวกหรืออีกนัยหนึ่งคือแบ่งตรงกลางอย่างละครึ่ง แต่เมื่อข้อมูลนั้นมีขนาดไม่เท่ากันทำให้การแยกคลาสบวกและคลาสลบ เป็น 75% แรกเป็นคลาสลบ และ 25% เป็นต่อมาเป็นคลาสบวกตามอัตราส่วนของข้อมูลเข้าที่เปลี่ยนไป

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

ตารางที่5 ความสอดคล้องกันของ 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

ตารางที่6 ความสามารถในการจำแนกของค่า AUC และ ความแม่นยำ(ไม่สมดุล)

และสุดท้ายเป็นการทดลองในหลายๆ อัตราส่วนของคลาสบวกและคลาสลบ โดยกำหนดให้มีข้อมูลทั้งหมด 10 ตัวอย่าง โดยเริ่มจากสมดุลคือมีทั้งหมดอย่างละ 5 ตัวอย่างจากนั้นเพิ่มและ ลดคลาสใดคลาสหนึ่งไปเรื่อยๆ จนไม่สามารถลดได้ ในกรณีนี้คือเหลือตัวเดียว

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

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

จากการทดลองทั้งสองจะเห็นได้ว่าไม่ว่าเป็นข้อมูลทั้งแบบที่สมดุลและไม่สมดุลก็ตามผลการทดลองยังคงเป็นไปในทิศทางเดียวกัน ทั้งในมุมมองความสอดคล้องที่ยังคงสอดคล้องกันสูง และในมุมมองความสามารถในการจำแนกที่ AUC มีความสามารถในการจำแนกสูงขึ้นเรื่อยๆ ตามขนาดของข้อมูลและยังมีความสามารถในการจำแนกสูงมากขึ้นเมื่อข้อมูลเกิดความไม่สมดุลของทั้งสองคลาส

4 การประยุกต์ใช้

จากการทดลองที่ผ่านมาได้เปรียบเทียบตัวประเมินประสิทธิภาพทั้งสองคือค่า AUC และ ค่าความแม่นยำ โดยได้ทำให้เห็นแล้วว่า AUC สามารถทำได้ดีกว่าค่าความแม่นยำ แต่อย่างไรก็ตามในการใช้งานจริงทั้ง AUC และความแม่นยำไม่ใช่เป้าหมายสุดท้าย เช่นตัวอย่าง ธนาคาร หรือ บริษัท ประกันภัย อาจจะมีข้อมูลของลูกค้าอยู่มหาศาลโดยสิ่งที่เขาต้องการสุดท้ายคือการคาดการณ์การทำการใดให้กับ บริษัท

สมมติว่าลูกค้าที่ถูกเก็บในฐานข้อมูลมีการเก็บด้วยแอตทริบิวต์จำนวนหนึ่งและลูกค้าแต่ละรายอาจเป็นผู้ซื้อหรือไม่ใช่ผู้ซื้อผลิตภัณฑ์บางอย่างเนื่องจากปัญหานี้เป็นปัญหาการจำแนกแบบไบนารี ลูกค้าจะได้รับการติดต่อจากแคมเปญการส่งเสริมการขายสำหรับลูกค้าแต่ละราย โดย บริษัท ต้องการคาดการณ์ว่าในสินค้าชนิดๆ หนึ่งนั้นลูกค้าแต่ละรายมีความต้องการสินค้านั้นมากเพียงใด และ ต้องเพิ่มโอกาสการซื้อเล็กน้อยเพียงใด

อย่างไรก็ตาม การประยุกต์ใช้ในโลกแห่งความเป็นจริง บริษัท อาจต้องการโปรโมตเพียงเล็กน้อยให้กับลูกค้าที่มีแนวโน้มสูงที่สุดที่คาดการณ์ไว้ และต้องโปรโมตมากขึ้นสำหรับลูกค้าที่มีแนวโน้มลดลง ซึ่งทำให้การที่ได้ต่อลูกค้าแต่ละคนนั้นต่างกันไปด้วยซึ่งในความเป็นจริง ก่อให้เกิดผลดีต่อรายได้ของบริษัท เพราะ สามารถลดการโปรโมตเกินจำเป็นสำหรับลูกค้าที่มีแนวโน้มจะซื้อสินค้าสูงๆ อยู่แล้ว เช่นลูกค้าที่มีแนวโน้มจะซื้อสินค้าสูงสุด 10% แรกนั้นอาจจะจะเป็นลูกค้าที่มีการซื้อสินค้าเป็นประจำในการโปรโมต

สินค้าที่ลูกค้ากลุ่มนี้ซื้อเป็นประจำอยู่แล้วอาจไม่จำเป็น และเพิ่มโอกาสให้ลูกค้าที่มีแนวโน้มลดลงมีโอกาสซื้อสินค้ามากขึ้นด้วย

5 สรุป

ในบทความนี้ได้ให้คำจำกัดความอย่างเป็นทางการเกี่ยวกับความสอดคล้องและ ความสามารถในการจำแนก เพื่อใช้ประเมินผลสำหรับขั้นตอนวิธีการเรียนรู้ในปัญหาการจำแนก กำหนดรูปแบบและเกณฑ์ที่ใช้สำหรับการเปรียบเทียบตัวประเมินประสิทธิภาพทั้งสอง และแสดงให้เห็นอย่างชัดเจนว่า AUC นั้นเป็นตัวประเมินประสิทธิภาพที่ดีกว่าค่าความแม่นยำ และได้นำไปเปรียบเทียบกับเหตุการณ์จริงในธุรกิจเพื่อแสดงผลลัพธ์ที่น่าสนใจว่า AUC เกี่ยวข้องโดยตรงกับกำไรสุทธิมากกว่าค่าความแม่นยำในการตลาดทางตรง

การเพิ่มประสิทธิภาพ AUC นั้นเป็นสิ่งที่ต้องการมากกว่าการเพิ่มความแม่นยำในการนำขั้นตอนวิธีการเรียนรู้สำหรับปัญหาการจำแนกรวมถึงการทำให้ข้อมูลที่จะนำไปประยุกต์ใช้ในโลกความเป็นจริง

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

ข้อมูลอ้างอิง

[1] Cortes, Corinna and Mohri, Mehryar. *AUC Optimization vs. Error Rate Minimization*. Florham Park, NJ 07932, USA

[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