

การเลือกประโยคที่เหมาะสมจากผลลัพธ์ที่ได้จากการแก้ไขประโยค

เมื่อตัวแรงแสวนประโยคทำการแก้ไขประโยคเสร็จแล้วผลลัพธ์ที่ได้ออกมา มักจะเป็นประโยคที่ถูกแก้ไขแล้วจำนวนหลายประโยค เราต้องมีวิธีการที่จะช่วยผู้ใช้ในการเลือกประโยคที่เหมาะสมออกมาเพื่อนำมาใช้งานในขั้นต่อไป ในการเลือกประโยคที่เหมาะสมนี้ เราอาจจะใช้การแรงแสวนแบบที่ใช้ Probabilistic Context Free Grammar (PCFG) (Su et al., 1991; Wright and Felshin, 1991) เข้าช่วย การแรงแสวนแบบนี้ในตารางแอลอาร์นอกจากจะมี การกระทำในรูปแบบปกติแล้วยังมีค่าความน่าจะเป็นในการทำการกระทำแต่ละการกระทำด้วย ดังนั้นแต่ละครั้งที่ทำการกระทำแต่ละอย่าง ก็จะต้องนำค่าความน่าจะเป็นเหล่านี้มาคำนวณด้วย ท้ายที่สุดโครงสร้างสายใดมีความน่าจะเป็นสูงสุดก็เลือกประโยคจากโครงสร้างสายนั้น

การใช้ความน่าจะเป็นในการแรงแสวนประโยคนั้น อาจแสดงให้เห็นได้ในตัวอย่างต่อไปนี้ (Tanaka,1993) คือ ถ้าให้ไวยากรณ์ PCFG มีค่าดังแสดงในรูปที่ 9.1 เมื่อ  $p_1, p_2, p_3, \dots, p_7$  เป็นค่าความน่าจะเป็นของ กฎการผลิตข้อต่างๆ

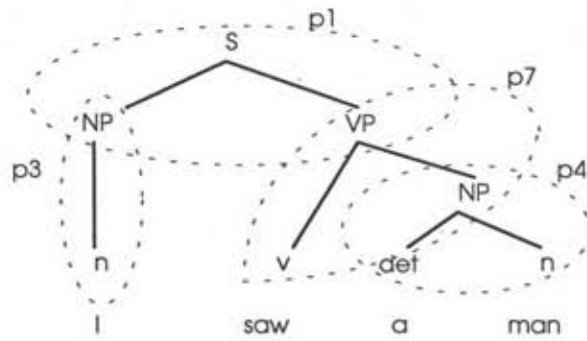
- |                                   |                                   |
|-----------------------------------|-----------------------------------|
| (1) S $\rightarrow$ NP VP, $p_1$  | (5) NP $\rightarrow$ NP PP, $p_5$ |
| (2) S $\rightarrow$ S PP, $p_2$   | (6) PP $\rightarrow$ p NP, $p_6$  |
| (3) NP $\rightarrow$ n, $p_3$     | (7) VP $\rightarrow$ v NP, $p_7$  |
| (4) NP $\rightarrow$ det n, $p_4$ |                                   |

รูปที่ 9.1 แสดงไวยากรณ์ PCFG ของกฎการผลิตที่ใช้

เงื่อนไขของความน่าจะเป็นของไวยากรณ์ PCFG เหล่านี้คือ

$$p_1 + p_2 = 1, p_3 + p_4 + p_5 = 1 \text{ และ } p_6 = 1, p_7 = 1$$

ถ้าประโยคที่ใช้ในการแรงแสวนคือ "I saw a man" และโครงสร้างต้นไม้ของประโยคนี้นี้ภายหลังการแรงแสวนแล้ว สามารถแสดงได้ดังรูปที่ 9.2



รูปที่ 9.2 แสดงโครงสร้างต้นไม้ของประโยค "I saw a man"

ในกรณีนี้ ความน่าจะเป็นของการเกิดโครงสร้างต้นไม้ในแบบนี้ก็คือ  $p1 \times p3 \times p7 \times p4$  เราสามารถนำหลักการนี้มาประยุกต์ใช้ ในการเลือกประโยคที่เหมาะสมจากประโยคผลลัพธ์ที่ได้จากการแก้ไขประโยคของตัวแฉงส่วน ได้โดยการเลือกประโยคที่มีความน่าจะเป็นสูงสุด

อย่างไรก็ตามการใช้ PCFG ในการแฉงส่วนประโยค จำเป็นต้องใช้ข้อมูลทางภาษาศาสตร์ประกอบ และเนื่องจากหัวข้อนี้ไม่ใช่จุดสำคัญในการทำวิทยานิพนธ์ ดังนั้นเราจึงได้ใช้วิธีการง่ายๆ คือตั้งสมมติฐานว่า ในการเลือกประโยคที่แก้ไขแล้วนั้น ประโยคที่ควรเลือกมากที่สุดก็คือประโยคที่มีการแก้ไขเปลี่ยนแปลงไปจากเดิมน้อยที่สุด ดังนั้นเราจึงใช้วิธีการเพิ่มค่าการแก้ไข (increasing penalty count) ลงไปในโครงสร้างมูลทุกครั้งที่มีการแก้ไขประโยค ดังนั้นเมื่อจบการประมวลผลแล้ว โครงสร้างข้อมูลสายใดที่มีค่าการแก้ไข น้อยที่สุดก็จะเป็นประโยคที่ถูกเลือก เช่น ตัวอย่างประโยคข้างล่างนี้ ผ่านการแก้ไขประโยคโดยวิธีการย้อนรอยกองซ้อนโครงสร้างกราฟ และมีการเรียงลำดับประโยคที่แก้ไขแล้ว จากประโยคที่มีค่าการแก้ไขประโยค จากน้อยไปมาก

ประโยคที่ถูกตัดอง : ฉัน เห็น เขา ใน สวน ตอน เทียง

ประโยคที่ผิดไวยากรณ์ : ฉัน เห็น เขา ใน สวน เทียง

ประโยคที่แก้ไขแล้ว : [ค่าการแก้ไข] ประโยค

- |      |                                  |
|------|----------------------------------|
| [01] | ฉัน เห็น เขา ใน สวน              |
| [01] | ฉัน เห็น เขา ใน สวน [p] เทียง    |
| [01] | ฉัน เห็น เขา ใน เทียง            |
| [02] | ฉัน [p] เขา ใน สวน [v] เทียง     |
| [03] | ฉัน [p] เขา [v] ใน สวน [p] เทียง |
| [03] | ฉัน [p] เขา [v] สวน              |
| [03] | ฉัน [p] เขา [v] สวน [p] เทียง    |
| [03] | ฉัน [p] เขา [v] [n] ใน สวน       |