

**รายงานสรุปการอบรม/สัมมนา/พัฒนาความรู้/ประชุมเชิงปฏิบัติการ/และเป็นวิทยากร
กองนโยบายและแผนการใช้ที่ดิน กรมพัฒนาที่ดิน**

| | |
|--|---|
| ส่วนที่ 1 ข้อมูลทั่วไป | <p>ชื่อ.....นายดิเรก.....นามสกุล.....คงแพ.....</p> <p>ตำแหน่ง.....นักวิเคราะห์นโยบายและแผนชำนาญการพิเศษ.....กลุ่ม/ฝ่าย.....กลุ่มวางแผนบริหารจัดการพื้นที่ชุ่มน้ำ.....</p> <p>หลักสูตร/หัวข้อเรื่องอบรม/สัมมนา/พัฒนาความรู้.....</p> <p>.....หลักสูตร การสร้างความรู้ความเข้าใจภาษาศาสตร์คอมพิวเตอร์เพื่อการนำไปประยุกต์ใช้กับปัญญาประดิษฐ์.....</p> <p>.....สถานที่อบรม/สัมมนา/พัฒนาความรู้.....</p> <p>.....ระบบการฝึกอบรมผ่านสื่ออิเล็กทรอนิกส์ ของ.....สถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล.....</p> <p>(https://e-learning.dga.or.th/xlms_ega/userroom/elearning.jsp).....</p> <p>หน่วยงานที่จัดฝึกอบรม/สัมมนา/พัฒนาความรู้.....</p> <p>.....สถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) (สพร.).....</p> <p>ตั้งแต่วันที่.....6.....เดือน.....กุมภาพันธ์.....พ.ศ.....2566..... ถึงวันที่.....6.....เดือน.....กุมภาพันธ์.....พ.ศ.....2566.....</p> <p>เพื่อ..... <input checked="" type="checkbox"/> อบรม..... <input type="checkbox"/> สัมมนา..... <input type="checkbox"/> อื่นๆ ระบุ.....</p> |
| ส่วนที่ 2 สิ่งที่ได้รับจากการอบรม/สัมมนา/พัฒนาความรู้ | <p>2.1 รายงานสรุปเนื้อหาสาระสำคัญในการอบรม/สัมมนา/พัฒนาความรู้.....</p> <p>.....การอบรมหลักสูตร การสร้างความรู้ความเข้าใจภาษาศาสตร์คอมพิวเตอร์เพื่อการนำไปประยุกต์ใช้กับ.....</p> <p>.....ปัญญาประดิษฐ์.....เรียนรู้เกี่ยวกับความหมายของภาษาศาสตร์คอมพิวเตอร์ ความรู้พื้นฐานที่เกี่ยวข้องกับการ.....</p> <p>.....เริ่มต้นเรียนรู้ภาษาศาสตร์คอมพิวเตอร์มีวัตถุประสงค์การเรียนรู้..... (1) เพื่อให้ผู้เรียนมีความรู้พื้นฐานเกี่ยวกับ.....</p> <p>.....ภาษาศาสตร์คอมพิวเตอร์ และ (2) เพื่อให้ผู้เรียนสามารถประยุกต์ใช้ความรู้ทางภาษาศาสตร์คอมพิวเตอร์ใน.....</p> <p>.....การพัฒนาปัญญาประดิษฐ์ประกอบด้วยเนื้อหาการเรียนคือ ภาษาศาสตร์คอมพิวเตอร์ Computational.....</p> <p>.....Linguistics คืออะไร Applications ของ Natural Language Processing การศึกษาคุณสมบัติทางสถิติของ.....</p> <p>.....ภาษา อยากรู้ภาษาอังกฤษต้องรู้ Vocab เท่าไร การพิสูจน์กฎของซีพพี กฎของซีพพีและโค้งซีพพี กฎของ.....</p> <p>.....ซีพพีสำหรับคลังข้อมูลภาษา ประยุกต์ใช้กฎของซีพพีเพื่อการพัฒนา Search Engine อยากรู้ Speech.....</p> <p>.....Recognizer เครื่องแปลงเสียงเป็นอักษร Language Mode เบื้องต้น การประยุกต์ Language Model ใน.....</p> <p>.....รูปแบบอื่น ๆ อยากรู้ให้เก่งขึ้น ต้องฝึก Language Model ให้ดีขึ้น.....</p> <p>.....อธิบายแยกตามหัวข้อการเรียน แบ่งได้เป็น.....</p> <p>.....1. ภาษาศาสตร์คอมพิวเตอร์ (Computational Linguistics) คืออะไร</p> <p>.....1) Computational Linguistics คือการใช้คอมพิวเตอร์วิเคราะห์ข้อมูลทางภาษา พุดให้เข้าใจง่าย.....</p> <p>.....ก็คืออยากรู้ให้คอมพิวเตอร์สามารถเข้าใจภาษาและทำหน้าที่ทางด้านภาษาต่าง ๆ แทนคนได้ เช่น การตรวจ.....</p> <p>.....เรียงความ จากที่ต้องให้ครูมานั่งอ่านและตรวจเองก็ใช้เทคโนโลยีเพื่อดูว่านักเรียนใช้คำถูกไหม มีการสะกดผิด.....</p> <p>.....ไหม เขียนแล้วอ่านรู้เรื่องไหม มีการสลับตำแหน่งประโยคใหม่ เป็นการวิเคราะห์ทั้งในระดับคำและระดับ.....</p> <p>.....ประโยคเพื่อจะประเมินได้ว่างานมีคุณภาพดีมากน้อยแค่ไหน หรือการแปลหนังสือ การแปลภาพยนตร์ ที่จะ.....</p> <p>.....สามารถนำเทคโนโลยีเข้ามาปรับใช้เพื่อการตรวจดูความหมายของการแปล และการสนทนาพูดคุยจากเดิมที่.....</p> <p>.....เราต้องโทรหาใครสักคนก็อาจเปลี่ยนไปสู่การใช้ Chatbot แทนได้ ซึ่งภาษาศาสตร์คอมพิวเตอร์ส่วนใหญ่ จะ.....</p> <p>.....ถูกแบ่งออกเป็น 5 กลุ่มคือ (1)ภาษาศาสตร์คลังข้อมูล โดยใช้คอมพิวเตอร์ช่วยในการวิเคราะห์ (computer.....</p> |

aided corpus linguistics) (2) การออกแบบโปรแกรมแจกแจงประโยค (parser) ให้รองรับภาษาธรรมชาติ (3) การออกแบบตัวกำกับ (tagger) เช่น ตัวกำกับชนิดคำ ฯลฯ (4) การนิยามตรรกศาสตร์แบบพิเศษ เช่น ตรรกศาสตร์ทรัพยากร เพื่อการประมวลผลภาษาธรรมชาติ (Natural language processing หรือ NLP) ฯลฯ (5) การวิจัยความสัมพันธ์ระหว่างภาษาฟอร์มอลกับภาษาธรรมชาติในสภาวะปกติ

..... 2) การประมวลผลภาษาธรรมชาติหรือ Natural language processing (NLP) เป็นวิทยาการแขนงหนึ่งในหมวดหมู่ของเทคโนโลยีปัญญาประดิษฐ์หรือ artificial intelligence ซึ่งช่วยให้คอมพิวเตอร์สามารถเข้าใจ ตลอดจนตีความและใช้งานภาษาปกติที่มนุษย์ใช้สื่อสารได้ โดยเทคโนโลยี NLP นี้ มีรากฐานจากวิทยาการหลากหลายสาขาคือ โดยเฉพาะด้านวิทยาการคอมพิวเตอร์ (computer science) และภาษาศาสตร์เชิงคำนวณ (computational linguistics) เพื่อวัตถุประสงค์ในการปิดช่องว่างทางการสื่อสารระหว่างมนุษย์และระบบคอมพิวเตอร์ เทคโนโลยี NLP ได้รวมภาษาศาสตร์คอมพิวเตอร์ (computational linguistics) หรือโมเดลจำลองแบบ ruled-base ของภาษามนุษย์ เข้ากับโมเดลทางสถิติ, machine learning และ deep learning เมื่อรวมทั้งหมดนี้เข้าด้วยกันด้วยเทคโนโลยี NLP จะทำให้คอมพิวเตอร์สามารถประมวลภาษาที่อยู่ในรูปของข้อความ และเสียง และสามารถเข้าใจความหมายได้อย่างสมบูรณ์ พร้อมเข้าใจเจตนา และความรู้สึกของผู้พูด หรือผู้เขียนได้เช่นกัน

..... 2. Applications ของ Natural Language Processing

..... 1) Machine translation เครื่องแปลภาษา คือ โปรแกรม หรือ เครื่องแปลภาษา เป็นส่วนหนึ่งของงานด้านการประมวลผลภาษาธรรมชาติ (Natural Language Processing) เป็นงานที่ท้าทายข้อจำกัดทางด้านภาษาของมนุษย์ เพื่อให้ผู้คนที่สามารถติดต่อสื่อสารกันได้โดยไม่มีกำแพงทางภาษามาวางกั้นไว้ เครื่องแปลภาษา เริ่มต้นโดยใช้ความรู้ทางด้านภาษาศาสตร์มาประยุกต์กับทางคอมพิวเตอร์เขียนกฎสำหรับแปลภาษาในคอมพิวเตอร์ เรียกว่าการแปลแบบ rule-based machine translation (RBMT) ซึ่งพอเวลาผ่านไปสักกระยะผู้คนเริ่มเห็นข้อจำกัดการแปลด้วยเทคนิคด้วยกฎ เพราะภาษามนุษย์มีไวยากรณ์มากมายและยากที่จะเขียนให้ครอบคลุม ต่อมาจึงมีเทคนิค Statistical machine translation (SMT) แปลด้วยสถิติ และล่าสุดเป็น Neural machine translation (NMT) ที่ใช้เทคนิคการเรียนรู้เชิงลึก (Deep learning) สามารถแปลได้แม่นยำที่สุดในปัจจุบัน

..... 2) Text Classification โจทย์ส่วนใหญ่ของ NLP นั้นสามารถแก้ได้ด้วยการสร้างเครื่องจำแนกประเภทข้อความ (Text Classifier) ซึ่งมี machine learning algorithm เป็นแกนหลัก (สามารถศึกษาพื้นฐานของ machine learning จาก module sentiment analysis) ปัจจุบันนี้คนหันมาใช้ Neural Network กันมากขึ้นโดยใช้พื้นฐานของความหมายของคำที่กำหนดโดยบริบทและการใช้ภาษา (Distributed semantic model หรือ distributional semantic model) ในคลังข้อมูลเป็นหลัก

..... 3) Chatbot คือ โปรแกรมคอมพิวเตอร์หรือปัญญาประดิษฐ์ ที่สร้างขึ้นเพื่อดูแลการสนทนาของผู้ใช้ ทั้งในรูปแบบตัวอักษร (Text) เสียง (Speech) แบบ Real-Time โดยใช้เทคโนโลยี Artificial Intelligent (AI) ที่ถูกพัฒนาขึ้น ให้มีบทบาทในการตอบกลับการสนทนาแบบอัตโนมัติผ่าน Messaging Application เสมือนการโต้ตอบของคนจริงๆ หรืออาจเรียกได้ว่าเป็นโปรแกรมตอบกลับอัตโนมัติ ซึ่งตัวโปรแกรมนี้จะถูกฝังอยู่บน Server, Application, หรือโปรแกรม Chat ต่าง ๆ ในปัจจุบัน Chatbot ได้รับความนิยมมาก ทั้งในกลุ่มผู้ประกอบการ เจ้าของเพจ นักการตลาด รวมถึงบริษัทใหญ่ เช่น IBM, Microsoft, Google, Facebook, LINE, Amazon เนื่องจากช่วยการทำงานให้มีประสิทธิภาพและคล่องตัวมากยิ่งขึ้น ทั้งในเรื่องการสนทนากับลูกค้าเป็นจำนวนมาก รวมถึงการทำ Digital Marketing ด้วย

..... 4) อื่นๆ เช่น Automatic Essay Grading เครื่องสำรวจการแจ้งเตือนภัย เครื่องแจ้งเตือนระบบจราจร เครื่องตรวจหาข่าวจริงข่าวปลอม เครื่องการวิเคราะห์วรรณกรรมและการตรวจสอบผู้เขียนวรรณกรรม

3. การศึกษาคุณสมบัติทางสถิติของภาษา

การศึกษาคุณสมบัติทางสถิติของภาษา (Statistical Properties of Language) เน้นการกระจายตัวของภาษาหน่วยคำ, ประโยค, คำ, นาม, กริยา, คุณสมบัติทางสถิติ, เช่นการกระจายตัว, การเกิดขึ้นของคำต่าง ๆ

THE INTERNATIONAL PHONETIC ALPHABET (2005)

CONSONANTS (PULMONIC)

| | Bilabial | Labio-dental | Dental | Alveolar | Post-alveolar | Retroflex | Palatal | Velar | Uvular | Pharyngeal | Epi-glottal | Glottal |
|---------------------|----------|--------------|--------|----------|---------------|-----------|---------|-------|--------|------------|-------------|---------|
| Nasal | m | ɱ | | n | | ɳ | ɲ | ŋ | ɴ | | | |
| Plosive | p b | ɸ β | | t d | | ʈ ɖ | c ɟ | k g | q ɢ | | ʔ | ʕ |
| Fricative | ɸ β | f v | θ ð | s z | ʃ ʒ | ʂ ʐ | ç ʝ | x ɣ | χ ʁ | ħ ʕ | ħ ʕ | h ɦ |
| Approximant | | ʋ | | ɹ | | ɻ | j | ɰ | | | | |
| Trill | ʙ | | | r | | | | | ʀ | | ʀ | |
| Tap, Flap | | ⱱ | | ɾ | | ɽ | | | | | | |
| Lateral fricative | | | | ɬ ɮ | | ɮ̺ | ç̺ | ʂ̺ | | | | |
| Lateral approximant | | | | l | | ɭ | ʎ | ʟ | | | | |
| Lateral flap | | | | ɭ | | ɮ̺ | | | | | | |

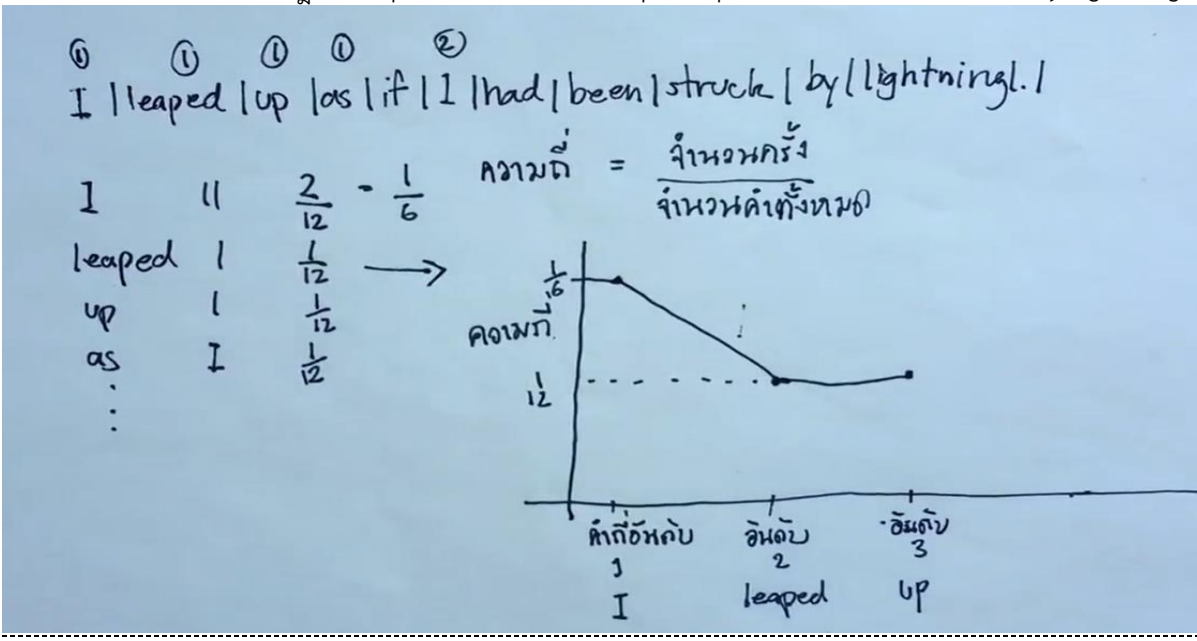
Where symbols appear in pairs, the one to the right represents a modally voiced consonant, except for murmured *f*. Shaded areas denote articulations judged to be impossible. Light grey letters are unofficial extensions of the IPA.

4. อยากรู้ภาษาอังกฤษต้องรู้ Vocab เท่าไหร่

การกระจายตัวของคำในภาษาสามารถอธิบายได้ด้วยกฎของซีฟฟ์ (Zipf Law) ซึ่งเป็นสูตรเดิมในแง่ของภาษาศาสตร์เชิงปริมาณที่ระบุว่าได้รับบางคลังของภาษาธรรมชาติคำพูดความถี่ของคำใด ๆ เป็นสัดส่วนผกผันในการจัดอันดับในตารางความถี่ ดังนั้นคำที่ใช้บ่อยที่สุดจะเกิดขึ้นประมาณสองเท่าของคำที่ใช้บ่อยที่สุด คำที่สอง บ่อยสามเท่าของคำที่มีความถี่สูงสุดที่สาม ฯลฯ ตัวอย่างเช่นใน Brown Corpus ของข้อความภาษาอังกฤษแบบอเมริกัน คำว่า "the" คือ คำที่เกิดขึ้นบ่อยที่สุด และโดยตัวมันเองคิดเป็นเกือบ 7% ของคำที่เกิดขึ้นทั้งหมด (69,971 จากมากกว่า 1 ล้านเล็กน้อย) ตามกฎของ Zipf คำที่สอง "ของ" คิดเป็น 3.5% ของคำเล็กน้อย (36,411 ครั้ง) ตามด้วย "และ" (28,852) คำศัพท์เพียง 135 รายการเท่านั้นที่จำเป็นในการคำนวณครึ่งหนึ่งของ Brown Corpus

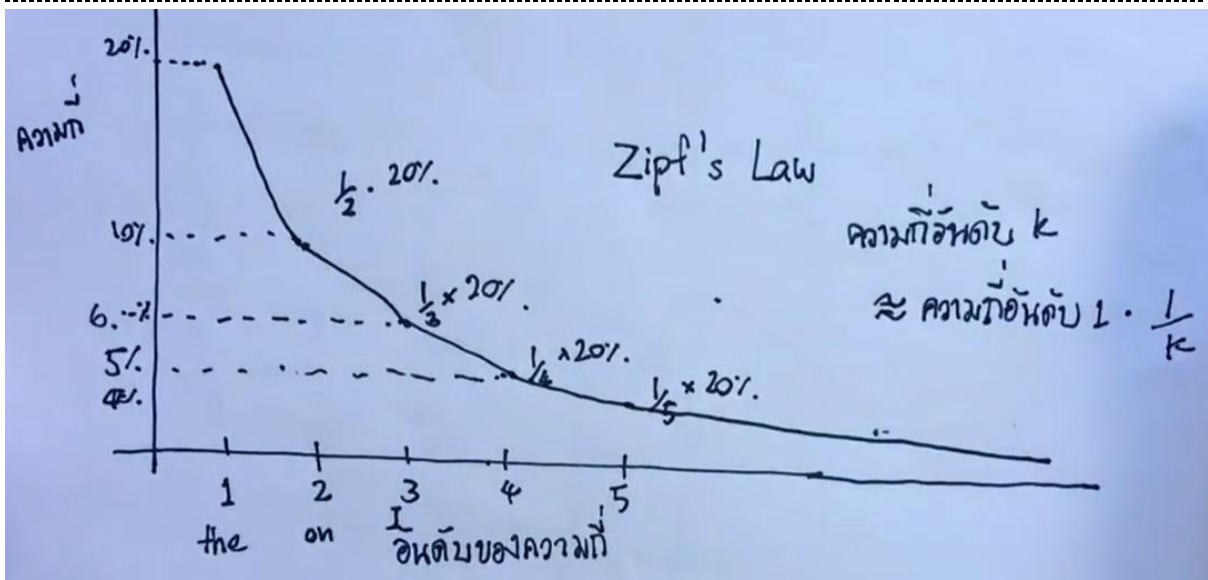
5. วิธีการพิสูจน์กฎของ Zipf

กฎของ Zipf เป็นโมเดลทางคณิตศาสตร์ โดยการ การตัดข้อความหรือประโยคออกเป็นคำ (word) และนำมาหาความถี่ ตามกฎของ Zipf ประโยคที่ว่า "I leaped up as if I had been struck by lightning"



6. กฎของ Zipf Zipf และเส้นโค้ง Zipf โดยนับจำนวนคำที่พบมาก และลดลงไปตามจำนวนที่นับได้

- A few words occur very often
 - 2 most frequent words can account for 10% of occurrences
 - top 6 words are 20%, top 50 words are 50%
- Many words are infrequent
- “Principle of Least Effort”
- Rank \times Frequency = Constant
 - $p_r = (\text{Number of occurrences of word of rank } r)/N$
 - N total word occurrences
 - probability that a word chosen randomly from the text will be the word of rank r
 - for D unique words $\sum p_r = 1$
 - $r \cdot p_r = A$
 - $A \approx 0.1$



7. กฎของ Zipf มักจะเป็นจริงสำหรับคลังข้อมูลภาษาทุกชนิด

สร้าง Log-log พล็อตชี้ให้เห็นว่าการกระจายตัวของคำในคลังข้อมูลต่าง ๆ เป็นไปตามกฎของ Zipf

Zipf's Law

$$f \approx \frac{c}{k}$$

ค่าคงตัว (แล้วแต่ภาษา) → c
 อันดับ → k

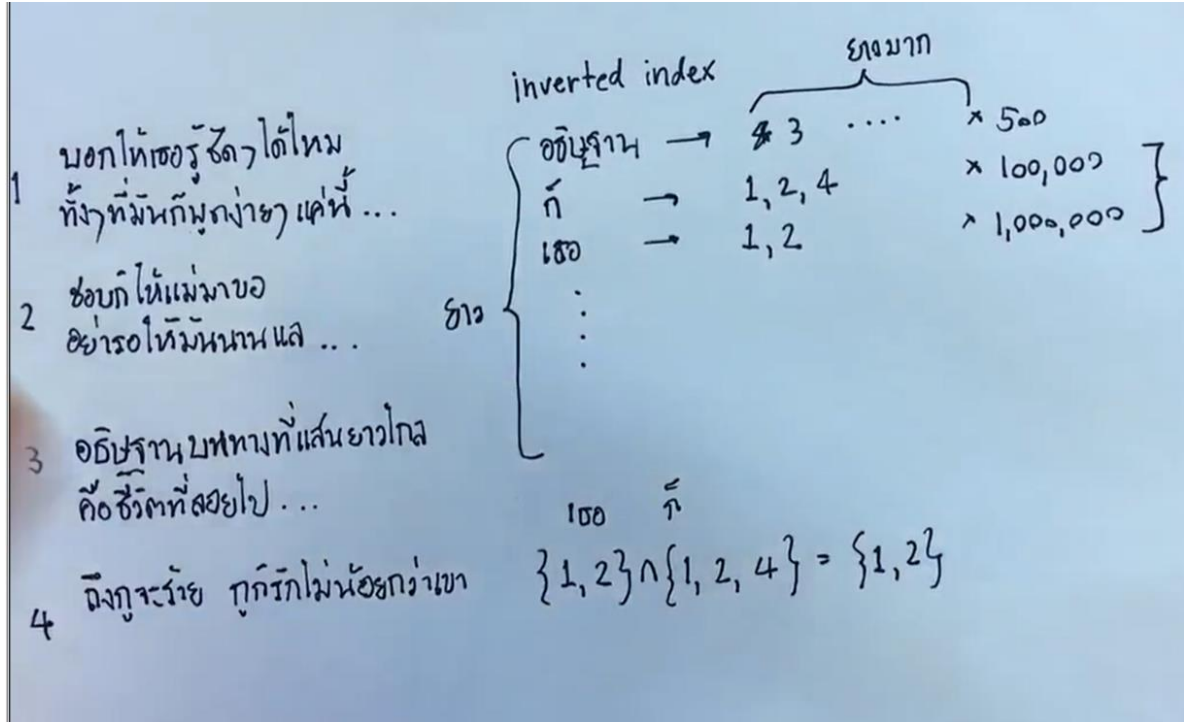
$$\log f = \log \left(\frac{c}{k} \right)$$

$$\log f = \frac{\log c - \log k}{\text{linear เชิงเส้น}}$$

$y = b + ax$
 $\uparrow \quad \quad \uparrow \quad \uparrow$
 $\log f \quad \log c \quad -1 \quad \log k$

8. การประยุกต์กฎของ Zipf เพื่อการพัฒนา Search Engine อย่างง่าย

สร้าง Search ค้นหาเพลง มีข้อดี (1) Search ได้เร็วขึ้น (2) ทำให้ inverted index เล็กลง เก็บได้ง่าย



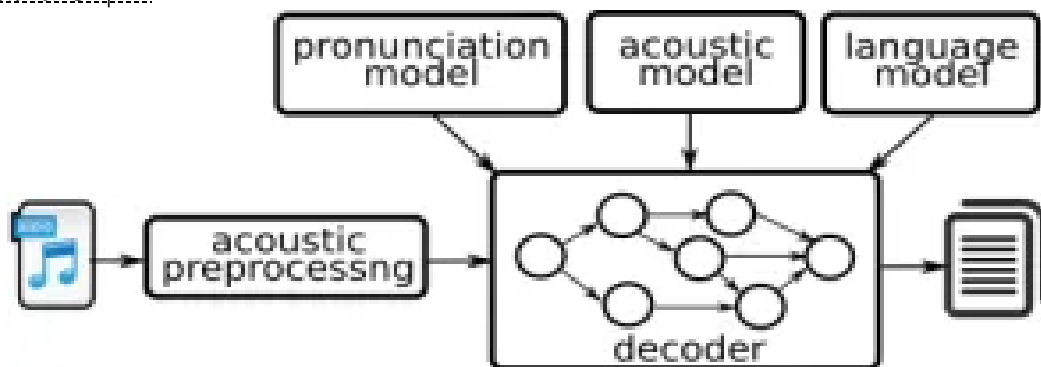
9. Speech Recognizer เครื่องแปลงเสียงเป็นอักษร

เทคโนโลยีรู้จำเสียงพูด (Automatic Speech Recognition: ASR) เป็นสาขาย่อยของ วิชาภาษาศาสตร์คอมพิวเตอร์ ที่พัฒนาวิธีการและเทคโนโลยีที่ช่วยให้การรับรู้และการแปลภาษาพูดเป็นข้อความโดยคอมพิวเตอร์ ซอฟต์แวร์รู้จำเสียงพูดขั้นพื้นฐานมีคำศัพท์ที่จำกัด ซึ่งประเภทของระบบรู้จำเสียงพูดสามารถแบ่งได้เป็น 3 ประเภทดังนี้ (1) เทคโนโลยีรู้จำเสียงพูดแบบคำโดด (Isolated speech) คือระบบที่รู้จำคำสั้น ๆ เพียงไม่กี่คำสั่ง เพื่อให้ระบบสามารถตอบโต้ได้อย่างรวดเร็ว (2) เทคโนโลยีรู้จำเสียงพูดแบบต่อเนื่อง (Continuous speech) คือระบบรู้จำคำจากเสียงอย่างต่อเนื่อง แล้วทำการพิจารณาตัดเสียงพูดและ (3) เทคโนโลยีรู้จำที่จำเสียงเพียงบางส่วน (Spontaneous speech) คือระบบที่จดจำเสียงที่ตรวจหาคำสำคัญเพียงคำเดียวในประโยคเพื่อหาใจความสำคัญ

1) การกำหนดแบบจำลองเสียง (Acoustic Model) แสดงถึงความสัมพันธ์ระหว่างสัญญาณเสียงและหน่วยเสียงหรือหน่วยภาษาอื่น ๆ ที่ใช้ในการพูด แบบจำลองนี้เรียนรู้จากชุดการบันทึกเสียงและการถอดเสียงที่สอดคล้องกัน ถูกสร้างขึ้นโดยการบันทึกเสียงการพูดและการถอดความข้อความและการใช้ซอฟต์แวร์เพื่อสร้างการแทนค่าทางสถิติของเสียงที่ประกอบกันเป็นแต่ละคำ

2) Pronunciation Model การทำรูปแบบการออกเสียง เป็นขั้นตอนที่เชื่อมโยงคำ

3) รูปแบบของภาษา (Language Model) เป็นขั้นตอนสำหรับทำนายลำดับของคำที่เป็นไปได้มากที่สุด ในหลาย ๆ คำ



4) การประมวลผลเสียงล่วงหน้า (Acoustic Preprocessing) เป็นขั้นตอนการแปลงจากคำพูดให้กลายเป็นลำดับของพารามิเตอร์อคูสติก

5) การถอดรหัสครั้งสุดท้าย (Finally a Decoder) เป็นขั้นตอนในการคาดการณ์ของอะคูสติกและแบบจำลองภาษาและถอดเสียงข้อความที่เป็นไปได้มากที่สุดสำหรับคำพูดออกมา

10. Language Model ถือเป็นหัวใจสำคัญของงานส่วนใหญ่ในด้าน NLP. เลยก็น่าจะได้ เพราะมันคือการหาสถิติของ string ในภาษา เช่น เวลาเรา search google ว่า "วันนี้" มันก็จะเดาคำถัดไปให้ครั้นว่าควรจะเป็นคำว่าอะไร เป็นต้น มีเทคนิคง่าย ๆ 3 แบบ คือ

1) Unigram model การตัดคำ เป็นแต่ละคำ และแทนค่าทางสถิติของเสียง

2) Bigram model การตัดคำ ใช้ 2 คำ และแทนค่าทางสถิติของเสียง

3) Trigram model การตัดคำ ใช้ 3 คำ และแทนค่าทางสถิติของเสียง

Unigram
LM (1) = 1000
LM (2) = 800
LM (3) = 100

$LM(1) = \# \text{Yesterday} + \# \text{I} + \# \text{led} + \# \text{her} + \# \text{to} \dots$
 $LM(2) = \# \text{Yesterday} + \# \text{I} + \# \text{let} + \# \text{her} + \# \text{to} \dots$
 $LM(3) = \# \text{Yesterday} + \# \text{I} + \# \text{letter} + \# \text{to} \dots$

Bigram model
2 word
 $LM(1) = \# (\text{Yesterday}, \text{I}) + \# (\text{I}, \text{led}) + \# (\text{led}, \text{her}) + \# (\text{her}, \text{to}) + \dots$
 $LM(2) = \# (\text{I}, \text{let}) + \# (\text{let}, \text{her}) + \# (\text{her}, \text{to})$
 $LM(3) = \# (\text{I}, \text{letter}) + \# (\text{letter}, \text{to})$

Trigram model
3 word
 $LM(1) = \# (\text{Yesterday}, \text{I}, \text{led}) + \# (\text{I}, \text{led}, \text{her}) + \# (\text{led}, \text{her}, \text{to}) + \dots$
 $LM(2) = \# (\text{I}, \text{let}, \text{her}) + \# (\text{let}, \text{her}, \text{to}) + \dots$

11. การประยุกต์ Language Model

N-Gram Language Model มี Technical detail ของการสร้าง Language Model เพื่อนำไปใช้ใน Applications อื่น ๆ เช่น text generator, grammar checker, spell checker, speech recognition, machine translation. Language model เป็นการรวมเทคนิคทางสถิติ ความน่าจะเป็น การคาดเดาคำจากบริบทรอบ ๆ โดยอ้างอิงจากสิ่งที่เรารู้มา. Language model จะแบ่งประเภทหลัก ๆ ได้ 2 ประเภท (1) Count-based คือการนับความถี่ของคำ เพื่อหาค่าความเป็นไปได้ของคำต่อไป (2) Continuous-space คือการนำ Neural network มาช่วยคำนวณเพื่อหาคำต่อไป โดยดูจากบริบทรอบข้างตัวอย่างจะเรียกว่า masking language คือการมาสกิ้งจุดที่จะให้ Language model สร้างคำ (หรือการเติมคำในช่องว่าง) จุดเด่นของ Count-based คือ language model จะมีความเบาและเร็วกว่า แบบ Continuous-space ที่มีเบื้องหลังเป็น Neural network เลยทำให้มีขนาดที่ใหญ่และประมวลผลได้ช้ากว่า จุดเด่นของ Continuous-space จะมีความสมบูรณ์ของประโยคมากกว่าเนื่องจากการดูบริบทของประโยคด้วย

12. อยากฟังให้เก่งขึ้น ต้องฝึก Language Model ให้มากขึ้น (สรุป)

การฝึก language model มีปัญหาจุกจิกทั้งด้านการ implement (การเขียนโปรแกรมเพื่อนำเอาโมเดลไปใช้จริง) และทางด้านการเลือกโมเดล เพราะต้องรับมือกับข้อมูลขนาดใหญ่ และภาษาทุกภาษามักมีความหลากหลายสูง NLP คือ Language Model หรือ โมเดลของภาษา ซึ่งถ้าโมเดลของเรามีความสามารถที่จะเข้าใจภาษาโดยภาพรวมได้ระดับหนึ่งแล้ว ก็จะส่งผลให้โมเดลนั้นทำงานเฉพาะทาง เช่น Classification, Sentiment Analysis, Machine Translation, Question-Answer ได้ดีขึ้นไปด้วยในตัว เพราะการประมวลผลภาษาธรรมชาติ หรือ Natural Language Processing (NLP) คือ หนึ่งในสาขาของวิทยาศาสตร์คอมพิวเตอร์ ที่เกี่ยวกับปัญญาประดิษฐ์ Artificial Intelligence และภาษาศาสตร์คอมพิวเตอร์

Computational Linguistics เป็นศาสตร์ที่ศึกษาเกี่ยวกับการทำให้คอมพิวเตอร์สามารถสื่อสารโต้ตอบด้วยภาษาของมนุษย์ และทำให้คอมพิวเตอร์เข้าใจภาษามนุษย์มากขึ้น ตัวอย่าง เช่น Siri, Google Assistant และ Alexa

2.2 ประสบการณ์/ประโยชน์ที่ได้รับ/การประยุกต์ใช้กับหน่วยงาน

ต่อตนเอง

ทำให้สามารถเรียนรู้และเข้าใจ การประมวลผลภาษาธรรมชาติ และ ภาษาศาสตร์คอมพิวเตอร์ กฎของ Zipf มักจะเป็นจริงสำหรับคลังข้อมูลภาษาทุกชนิด ที่เกี่ยวกับปัญญาประดิษฐ์ Artificial Intelligence เป็นต้น.....

ต่อหน่วยงาน / การนำมาประยุกต์ใช้กับหน่วยงาน

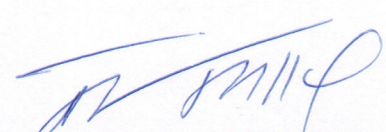
..... เป็นแนวทางการพัฒนา Language Model ถือเป็นหัวใจสำคัญของงานส่วนใหญ่ในด้าน NLP ในการพัฒนา Chatbot โปรแกรมคอมพิวเตอร์ บริการเกษตรกร ในอนาคต.....

2.3 ปัญหาและอุปสรรคในการอบรม/สัมมนา/พัฒนาความรู้ฯ

..... - การใช้เครื่องมือเป็นการเรียนรู้โมเดลภาษาใหม่ ๆ จึงต้องมีการศึกษาเพิ่มเติมในโปรแกรมภาษาคอมพิวเตอร์เช่น ภาษา Python C# เป็นต้น.....

2.4 ข้อคิดเห็นและข้อเสนอแนะ

- - ควรพัฒนาเครื่องมือ coding ภาษา Python -....
- - ควรเซิร์ฟเวอร์กลางในการทำงาน GPU มากพอทำระบบ coding ด้วย ภาษา Python -....
- - ควรพัฒนา Chatbot โปรแกรมคอมพิวเตอร์ พัฒนาขึ้นมาให้มีบทบาทในการตอบกลับการสนทนาด้วยตัวอักษรแบบอัตโนมัติผ่าน Messaging Application ให้ผู้ช่วยอัจฉริยะเพื่อบริการเกษตรกร ในอนาคต -....

ลงชื่อ..... 

(นายดิเรก คงแพ)

ตำแหน่ง นักวิเคราะห์นโยบายและแผนชำนาญการพิเศษ.....

ผู้รายงาน

วันที่ ๗ เดือน ก.พ. พ.ศ. ๖๖

ส่วนที่ 3 ความเห็นของผู้บังคับบัญชา

() ทราบ

ลงชื่อ..... 

(นายเชษฐจร จันทรเปล่ง)

ตำแหน่ง ผู้อำนวยการกองนโยบายและแผนการใช้ที่ดิน.....

วันที่ ๘ เดือน ก.พ. พ.ศ. ๖๖

ประกาศนียบัตร

ให้ไว้เพื่อแสดงว่า

ดิเรก คงแพ

ได้ผ่านการอบรมด้วยระบบการเรียนออนไลน์ในบทเรียน
การสร้างความรู้ความเข้าใจภาษาศาสตร์คอมพิวเตอร์
เพื่อการนำไปประยุกต์ใช้กับปัญญาประดิษฐ์

รวมระยะเวลาทั้งสิ้น 1 : 30 ชั่วโมง

โดยสถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล
ภายใต้การดำเนินงานของสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)
ให้ไว้ ณ วันที่ 6 ก.พ. 2566



(นางไอรดา เหลืองวิไล)

รองผู้อำนวยการสำนักงานพัฒนารัฐบาลดิจิทัล
รักษาการแทนผู้อำนวยการสถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล



1fc9aeb4