



Natural language processing (NLP)

Crash Course (Week 3)

Attawut Nardkulpat

Faculty of Geoinformatics

Burapha University

attawut@buu.ac.th



Outline

- **Week 3 :**
 - Sentiment Analysis
 - Logistic Regression
 - Naïve Bayes
 - Named-Entity Recognition
 - Geolocation



Sentiment Analysis



Sentiment Analysis คืออะไร ?

- คือ “การวิเคราะห์ความรู้สึก”
- Positive = เป็นในทางที่ดี
- Negative = เป็นในทางที่ไม่ดี
- Neutral = เป็นกลาง



Logistic Regression

- <https://www.ilog.ai/logistic-regression-%e0%b8%84%e0%b8%b7%e0%b8%ad/>
- <https://www.ilog.ai/logistic-regression-%e0%b8%84%e0%b8%b7%e0%b8%ad-ep-2/>



Metric

- <https://datarockie.com/2019/03/30/top-ten-machine-learning-metrics/>



Named-Entity Recognition (NER)



Information Extraction

- เป็นเทคนิคที่สำคัญในการสกัดความรู้ออกจากข้อมูลที่เป็นข้อความที่มีขนาดใหญ่เกินที่คนจะสามารถอ่านเองได้หมด ข้อมูลที่สกัดนี้มักจะถูกจัดอยู่ในรูปแบบที่เป็นมาตรฐานที่สามารถนำไปวิเคราะห์ต่อไปได้ เป็นเทคนิคที่สำคัญในการสกัดความรู้ออกจากข้อมูลที่เป็นข้อความที่มีขนาดใหญ่เกินที่คนจะสามารถอ่านเองได้หมด ข้อมูลที่สกัดนี้มักจะถูกจัดอยู่ในรูปแบบที่เป็นมาตรฐานที่สามารถนำไปวิเคราะห์ต่อไปได้



NER

- การรู้จำชื่อเป็นการสกัด ชื่อบุคคล ชื่องค์กร ชื่อสถานที่ และอื่นๆ จากข้อความ
- ทำการดึงข้อมูลที่เราอยากได้ออกจากข้อความ เหมือนกับการให้เครื่องหา highlight ส่วนของข้อความที่เราสนใจโดยอัตโนมัติ

I hear Place Berlin is wonderful in the Time winter



In fact, the Chinese NORP market has the three CARDINAL most influential names of the retail and tech space – Alibaba GPE, Baidu ORG, and Tencent PERSON (collectively touted as BAT ORG), and is betting big in the global AI GPE in retail industry space. The three CARDINAL giants which are claimed to have a cut-throat competition with the U.S. GPE (in terms of resources and capital) are positioning themselves to become the ‘future AI PERSON platforms’. The trio is also expanding in other Asian NORP countries and investing heavily in the U.S. GPE based AI GPE startups to leverage the power of AI GPE. Backed by such powerful initiatives and presence of these conglomerates, the market in APAC AI is forecast to be the fastest-growing one CARDINAL, with an anticipated CAGR PERSON of 45% PERCENT over 2018 - 2024 DATE.

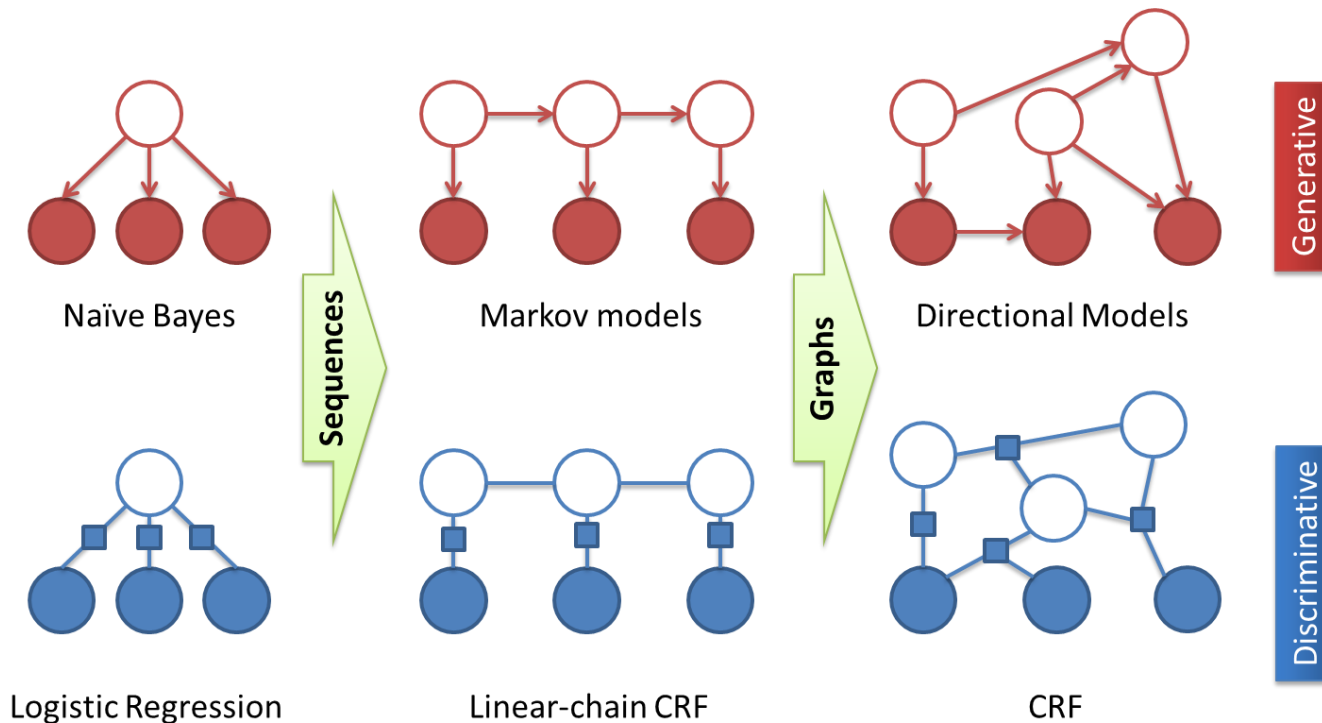
To further elaborate on the geographical trends, North America LOC has procured more than 50% PERCENT of the global share in 2017 DATE and has been leading the regional landscape of AI GPE in the retail market. The U.S. GPE has a significant credit in the regional trends with over 65% PERCENT of investments (including M&As, private equity, and venture capital) in artificial intelligence technology. Additionally, the region is a huge hub for startups in tandem with the presence of tech titans, such as Google ORG, IBM ORG, and Microsoft ORG.



How NER works

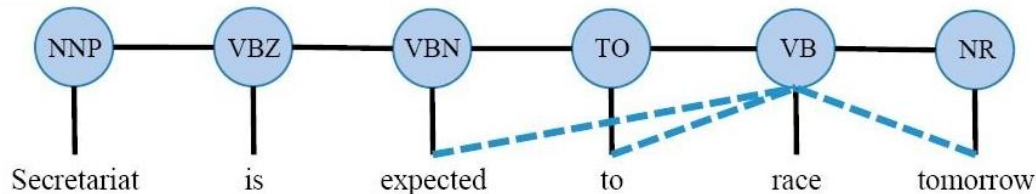
- Detect a named entity / Categorize the entity
- Person : E.g., Elvis Presley, Audrey Hepburn, David Beckham
- Organization : E.g., Google, Mastercard, University of Oxford
- Time : E.g., 2006, 16:34, 2am
- Location : E.g., Trafalgar Square, MoMA, Machu Picchu
- Work of art : E.g., Hamlet, Guernica, Exile on Main St.

Algorithm





CRF



CRF

$$P(\bar{s} | \bar{o}) \propto \frac{1}{Z_{\bar{o}}} \prod_{t=1}^{|\bar{o}|} \exp \left(\sum_j \lambda_j f_j(s_t, s_{t-1}) + \sum_k \mu_k g_k(s_t, x_t) \right)$$

