Opinion Mining using Text Representing Centriods

Maleerat Sodanil
Faculty of Information Technology, KMUTNB
Email: maleerat.s@it.kmutnb.ac.th

# Social Media – User Generated Contents

By David Ndichu

User Generated Contents (UGC)

+ User Generated Contents (UGC) is any content—text, videos, images, reviews, etc.—created by people

+ It may effected to the decision maker who need to listen to the sound of social media

Natural Language Processing-NLP

+ NLP is a field in machine learning with the ability of a computer to understand, analyze, manipulate, and potentially generate human language. NLP is used in several area such as:

+ Information Retrieval

+ Information Extraction

+ Machine Translation

+ Sentiment Analysis

+ Text Summarization

+ Etc...
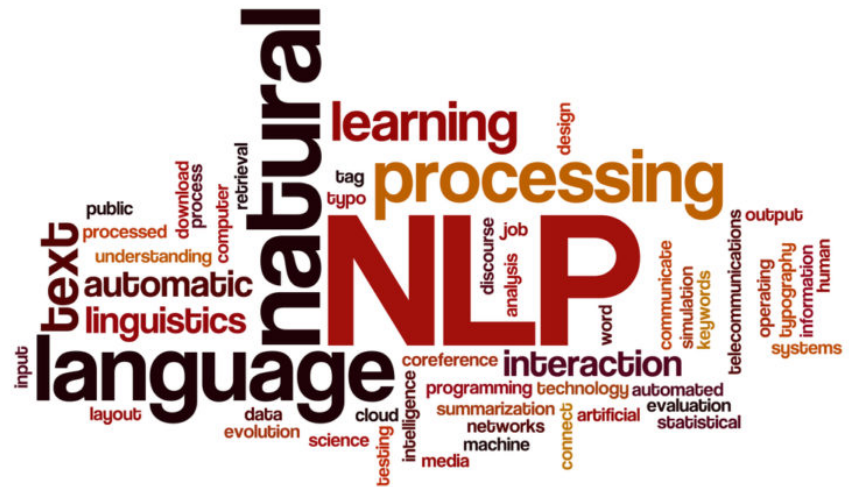
What people think?

+ *"Which phone should I buy?"*

+ *"Which Professor to work for?"*

+ *"Whom should I vote for?"*

What people think about Phone?

👍 **Positive**

👎 **Negative**

Example of reviews

รีวิว iphone XR จาก ตปท. มันดีย์นะ ไม่ใช่แค่พอใช้นะแกร~

`iPhone` `iOS`

สรุปจาก the verge นะครับ
https://www.theverge.com/2018/10/23/18011306/apple-iphone-xr-review-camera-scre

(บ้านเรามีของตาอู๋ spin9 รีวิวแล้วนะ)

สรุป
- ชิป A12 bionic ,กล้องหน้า,กล้องหลัง(ยกเว้นเทเลใช้ซูม) HW แทบจะยกมาเหมือน xs,xs max
- ประสบการณ์การใช้งานแทบไม่ต่างจาก xs ในราคาถูกกว่า หมื่นนึง
- จอด้อยกว่า xs แต่ไม่ห่วย คือจอเท่า iphone 7 ใครใช้จอ iphone 7 มาก่อนก็คงเฉยๆหรือไม่ส้
ความดำไม่เท่า เปิด true tone สีจะ worm กว่า xs
ไม่ support HDR กับ dolby vision ถ้าแคร์เรื่อง จอมาก ไป xs ดีกว่า
(xs max สี HDR เวลาดูหนังแจ่มมาก-จขกท)

Basic components of an opinion

1. **Opinion holder**: The person or organization that holds a specific opinion on a particular object

2. **Object**: item on which an opinion is expressed

3. **Opinion**: a view, attitude, or appraisal on an object from an opinion holder.

Type of opinions

+ **Direct Opinions**: direct sentiment expressions on some target objects, e.g., products, events, topics, persons.

  + E.g., "the picture quality of this camera is great."

+ **Comparative Opinions**: Comparisons expressing similarities or differences of more than one object. Usually stating an ordering or preference.

  + E.g., "car A is cheaper than car B"

Opinion Mining- Direct Opinion

Opinion Mining $\rightarrow$ Sentiment classification computational study of opinions, sentiments and emotions expressed in text.

+ Document Level

+ Sentence Level

+ Feature Level

What is an opinion?

- An *opinion* is a quintuple

$$(o_j, f_{jk}, so_{ijkl}, h_i, t_l),$$

where

- ❑ o_j is a target object.
- ❑ f_{jk} is a feature of the object o_j.
- ❑ so_{ijkl} is the sentiment value of the opinion of the opinion holder h_i on feature f_{jk} of object o_j at time t_l. so_{ijkl} is +ve, -ve, or neu, or a more granular rating.
- ❑ h_i is an opinion holder.
- ❑ t_l is the time when the opinion is expressed.

Handbook of Natural Language Processing

Sentiment Classification
(Pang and Lee, et al 2002 and Turney 2002)

+ Classify a document based on the overall sentiment expressed by opinion holder
  + Classes: Positive, or negative (and neutral)
+ In the model, (o_j, f_{jk}, so_{ijkl}, h_i, t_l),

Text representative

+ The standard representation methods : Bag-of-Words (BOW)

+ BOW:
    + text is represented as a vector of words weights
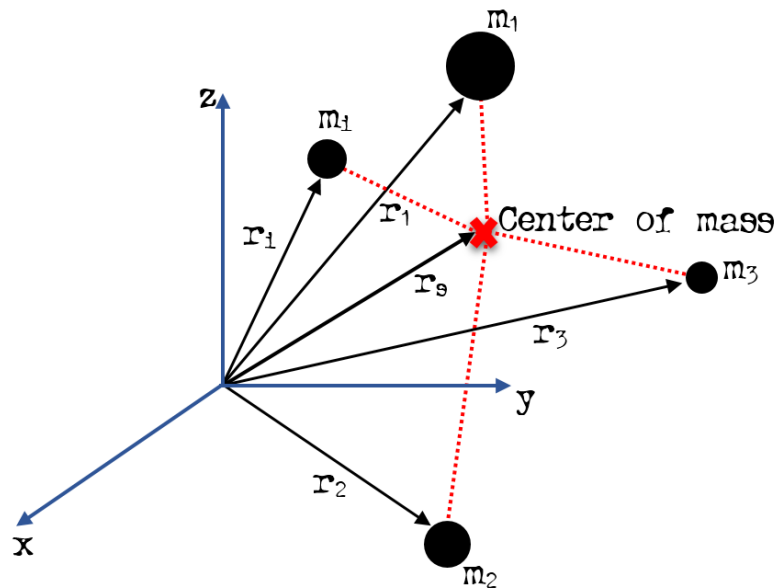   + techniques based on the term frequency information

Source: https://bzdww.com/article/159625/

Text representative

+ Problem: BOW
    + No sequence of words considered (ignores the context of words)
    + No structure of document is considered
    + Return only rough estimations regarding the relatedness of texts

What is Centroid?

The physical analogon: **-> centre of mass**

M. M. Kubek and H. Unger, "Centroid Terms as Text Representatives,
" *Proceedings of the 2016 ACM Symposium on Document Engineering - DocEng 16*, 2016.

Text Representing Centroids

+ Text representation model is an important element in determining the performance and accuracy of the categorization.

+ The suitable representation of the textual information is representing the semantic information of the text.

M. M. Kubek and H. Unger, "Centroid Terms as Text Representatives, " *Proceedings of the 2016 ACM Symposium on Document Engineering - DocEng 16*, 2016.

Text Representing Centroids (TRC)

- words = mass point
- distance vector = distance in co-occ. graph

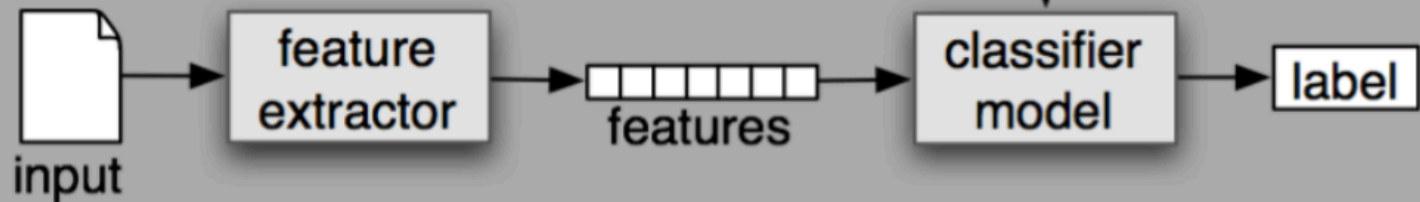→ e.g. school is the centroid of a document containing classroom, students, teacher but also computer
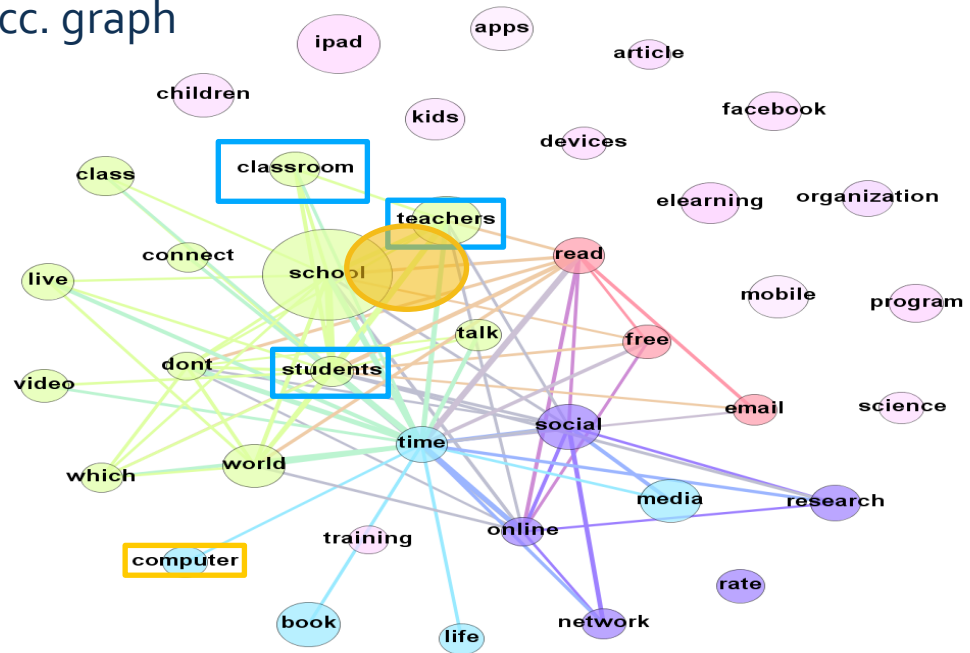
→ The centroid of a document is the term with the minimum average distance to all words of the respective document in the co-occ. graph.
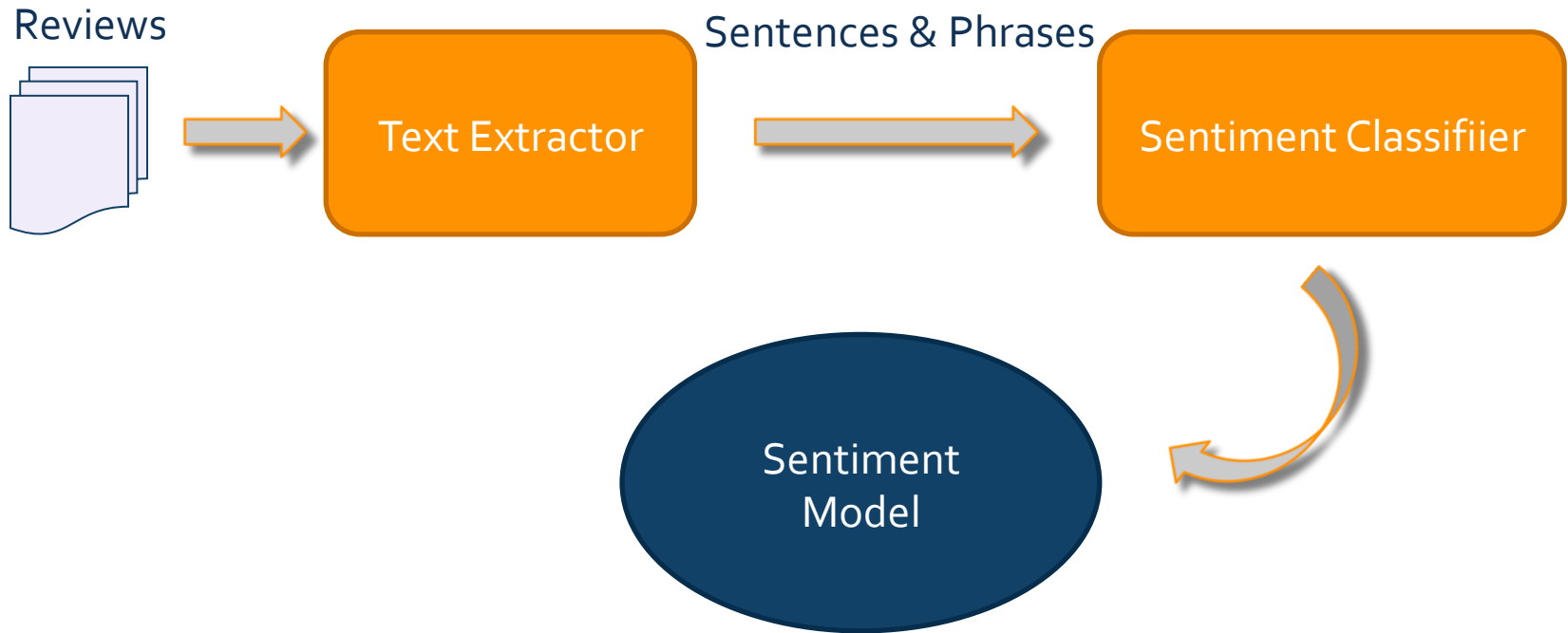
Algorithm for Centroid term

+ Step 1: Calculate Weight and co-occurrence frequency

+ Step 2: Calculate a distance

+ Step 3: Calculate centroid-term of documents

The centroid of a document $\rightarrow$ the minimum average distance to all words of the respective document in the co-occ. graph.

Classification Framework
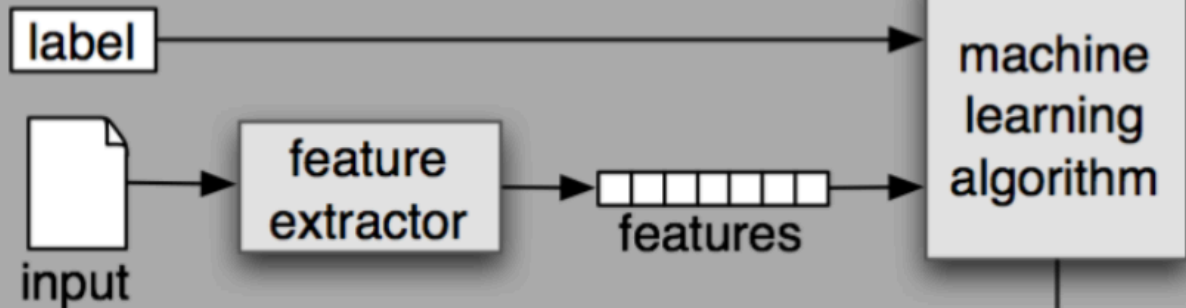
Opinion Mining Framework

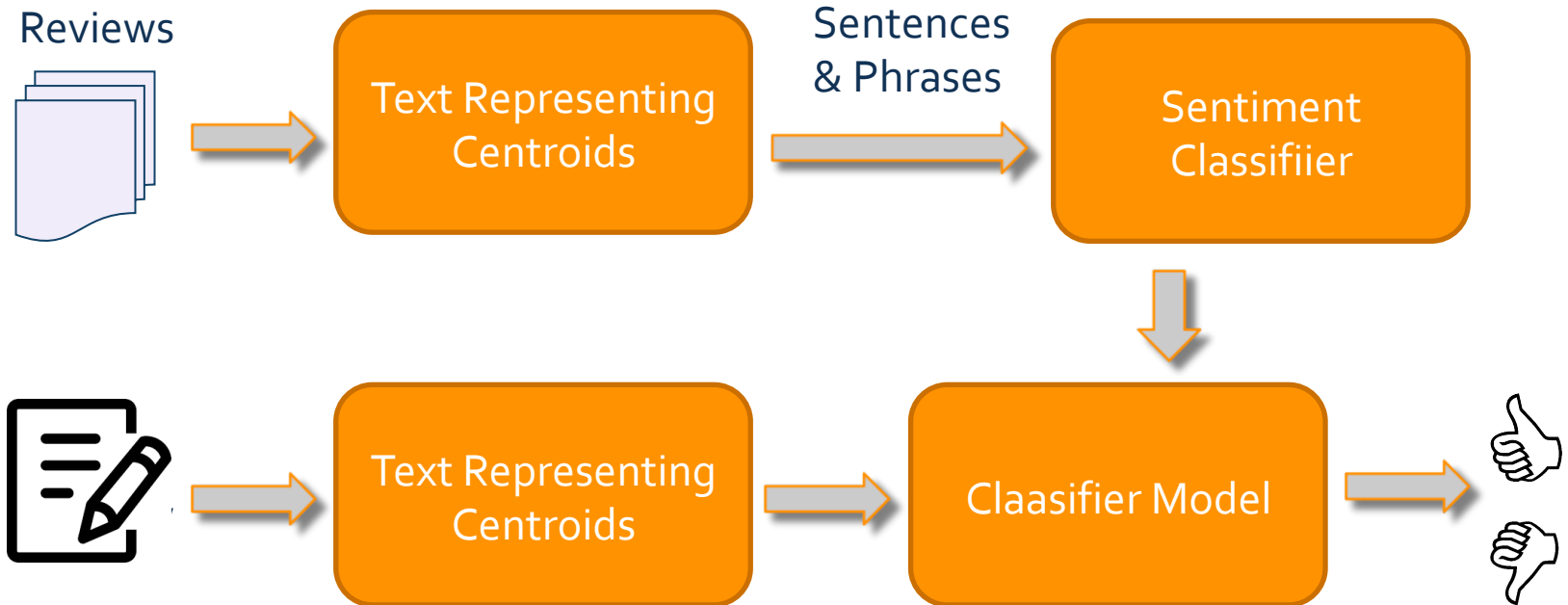Reviews

Sentences & Phrases

Text Extractor

Sentiment Classifiier

Sentiment Model

Opinion Mining using Centroids

Reviews

Text Representing Centroids

Sentences & Phrases

Sentiment Classifiier

Text Representing Centroids

Claasifier Model

Summary

+ Generally modeled as classification
    + predict a binary or ordinal label (e.g. Positive, Negative, Neutral)

+ Text Extractor:
    + Using centroid terms: Single words that semantically and topically characterize text documents

Example of Sentiment Analysis App.
S-Sense (NECTEC)

+ http://ssense.in.th/tourism/

Thank you