COMP7630 – Web Intelligence and its Applications

Sentiment Analysis

Valentino Santucci

(valentino.santucci@unistrapg.it)

Opinion Mining or Sentiment Analysis

- Computational study of opinions and sentiments expressed in text
  - Reviews, blogs, discussions, news, comments, feedback, or any other documents

- Businesses always want to find public or consumer opinions about their products and services.
- Potential customers want to know the opinions of existing users before they use a service or purchase a product.

- Also applicable to other sectors (e.g., politics)
- Opinions are important because they are key influencers of our behaviors.

Opinion Mining (Illustration)

- Consider the following review segment on iPhone:
- *"I bought an iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. However, my mother was mad with me as I did not tell her before I bought it. She also thought the phone was too expensive, and wanted me to return it to the shop. … "*

Opinion Mining (Illustration)

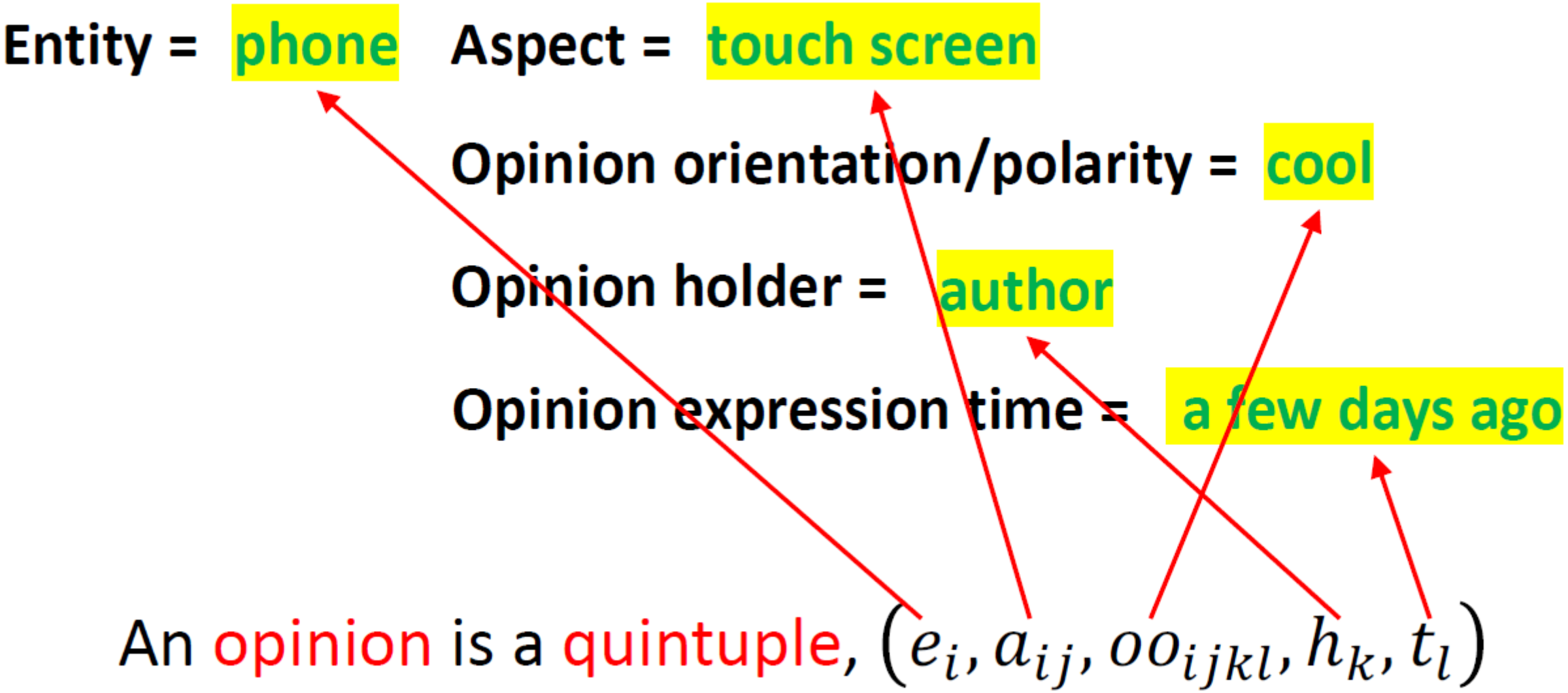
- Consider the following review segment on iPhone:
- *"(1) I bought an iPhone a few days ago. (2) It was such a nice phone. (3) The touch screen was really cool. (4) The voice quality was clear too. (5) However, my mother was mad with me as I did not tell her before I bought it. (6) She also thought the phone was too expensive, and wanted me to return it to the shop. …"*

Opinion Target

Opinion Holder = Author

Opinion Holder = Mother

Semantics – What do we mean?



An opinion is a quintuple, $(e_i, a_{ij}, oo_{ijkl}, h_k, t_l)$

Opinion (Formal) Representation

- An opinion is a quintuple, $(e_i, a_{ij}, oo_{ijkl}, h_k, t_l)$
  - e_i - entity (e.g., phone)
  - a_{ij} - aspect of e_i (e.g., touch screen)
    - specify *GENERAL* if the opinion is on e_i instead of its aspect
  - oo_{ijkl} - orientation (e.g., positive/negative/neutral/)
  - h_k - opinion holder (e.g., author)
  - t_l - the time when the opinion is expressed by h_k.
- Model of an opinionated document
  - contains opinions on a set of entities $\{e_1, e_2, \ldots, e_r\}$ from a set of opinion holders $\{h_1, h_2, \ldots, h_p\}$.
  - The opinions on each entity e_i are expressed on the entity itself and some aspects.

Opinion Mining – Key Tasks

1. Entity Extraction
2. Aspect Extraction
3. Opinion Holder and Time Extraction
4. Aspect Sentiment Classification

Task 1: Entity Extraction

Posted by: bigXyz on Nov-4-2010: (1) I bought a Motorola phone and my girlfriend bought a Nokia phone yesterday. (2) We called each other when we got home. (3) The voice of my Moto phone was unclear, but the camera was good. (4) My girlfriend was quite happy with her phone, and its sound quality. (5) I want a phone with good voice quality. (6) So I probably will not keep it.

(Motorola, voice_quality, negative, bigXyz, Nov-4-2010)

(Motorola, camera, positive, bigXyz, Nov-4-2010)

(Nokia, GENERAL, positive, bigXyz's girlfriend, Nov-4-2010)

(Nokia, voice_quality, positive, bigXyz's girlfriend, Nov-4-2010)

Task 2: Aspect Extraction

Posted by: bigXyz on Nov-4-2010: (1) I bought a Motorola phone and my girlfriend bought a Nokia phone yesterday. (2) We called each other when we got home. (3) The voice of my Moto phone was unclear, but the camera was good. (4) My girlfriend was quite happy with her phone, and its sound quality. (5) I want a phone with good voice quality. (6) So I probably will not keep it.

(Motorola, voice_quality, negative, bigXyz, Nov-4-2010)

(Motorola, camera, positive, bigXyz, Nov-4-2010)

(Nokia, GENERAL, positive, bigXyz's girlfriend, Nov-4-2010)

(Nokia, voice_quality, positive, bigXyz's girlfriend, Nov-4-2010)

Task 3: Holder & Time Extraction

Posted by: bigXyz on Nov-4-2010: (1) I bought a Motorola phone and my girlfriend bought a Nokia phone yesterday. (2) We called each other when we got home. (3) The voice of my Moto phone was unclear, but the camera was good. (4) My girlfriend was quite happy with her phone, and its sound quality. (5) I want a phone with good voice quality. (6) So I probably will not keep it.

(Motorola, voice_quality, negative, bigXyz, Nov-4-2010)

(Motorola, camera, positive, bigXyz, Nov-4-2010)

(Nokia, GENERAL, positive, bigXyz's girlfriend, Nov-4-2010)

(Nokia, voice_quality, positive, bigXyz's girlfriend, Nov-4-2010)

Task 4: Aspect Sentiment Classification

Posted by: bigXyz on Nov-4-2010: (1) I bought a Motorola phone and my girlfriend bought a Nokia phone yesterday. (2) We called each other when we got home. (3) The voice of my Moto phone was unclear, but the camera was good. (4) My girlfriend was quite happy with her phone, and its sound quality. (5) I want a phone with good voice quality. (6) So I probably will not keep it.

(Motorola, voice_quality, **negative**, bigXyz, Nov-4-2010)

(Motorola, camera, **positive**, bigXyz, Nov-4-2010)

(Nokia, GENERAL, **positive**, bigXyz's girlfriend, Nov-4-2010)

(Nokia, voice_quality, **positive**, bigXyz's girlfriend, Nov-4-2010)

Overall Expected Results:

Posted by: bigXyz on Nov-4-2010: (1) I bought a Motorola phone and my girlfriend bought a Nokia phone yesterday. (2) We called each other when we got home. (3) The voice of my Moto phone was unclear, but the camera was good. (4) My girlfriend was quite happy with her phone, and its sound quality. (5) I want a phone with good voice quality. (6) So I probably will not keep it.

(Motorola, voice_quality, **negative**, bigXyz, Nov-4-2010)

(Motorola, camera, **positive**, bigXyz, Nov-4-2010)

(Nokia, GENERAL, **positive**, bigXyz's girlfriend, Nov-4-2010)

(Nokia, voice_quality, **positive**, bigXyz's girlfriend, Nov-4-2010)

How to carry out the tasks?

- NLP pipeline is very useful here!

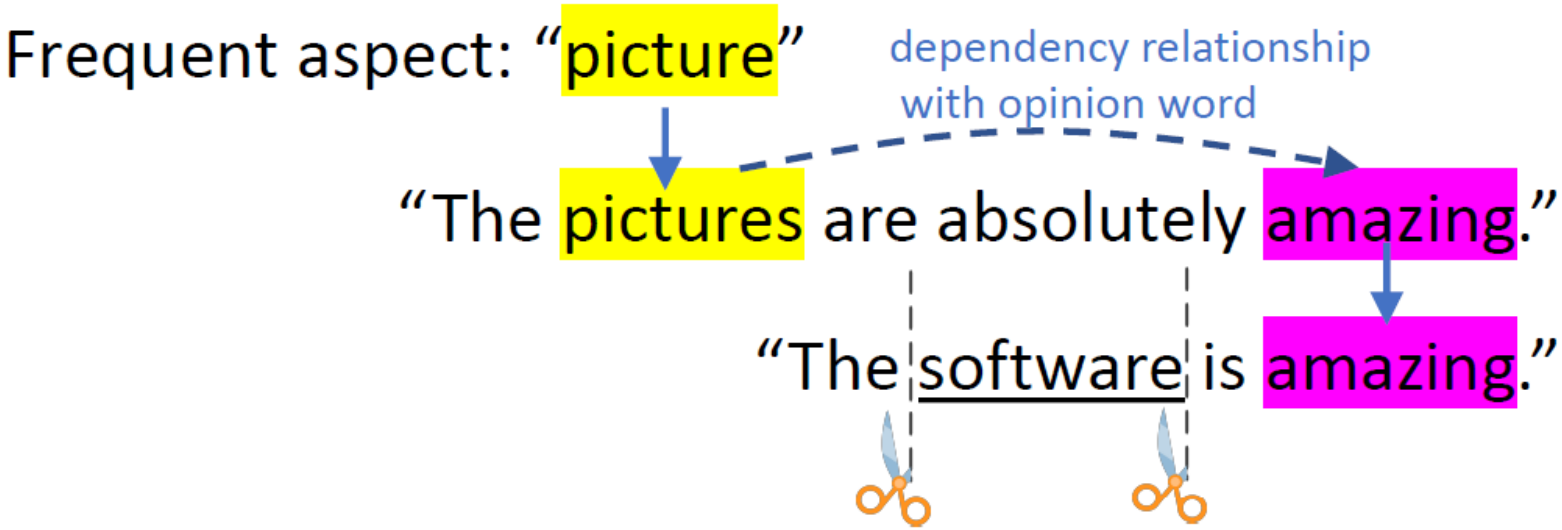
Entity/Aspect Extraction

- Extraction Rules are often designed to locate and extract the target information based on …

  - 1) Specific words/forms representing the targets
    - E.g., [{"Notebook", "Labtop", "portable", …}]
  - 2) Specific words/forms before and after the target
    - E.g., [{"good", "excellent"} follow-by [NOUN]] where the noun extracted will be the aspect.

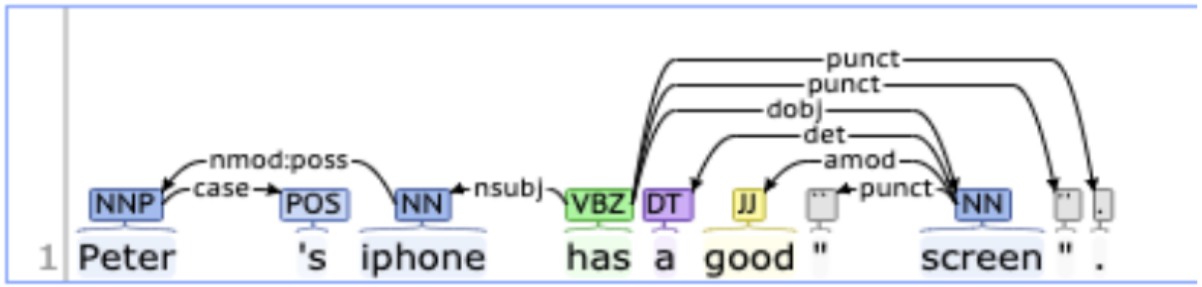
Unsupervised Aspect Extraction

- Step 1: Find frequent nouns and noun phrases by a POS tagger.
  - E.g., What could be extracted for camera reviews?
- Step 2: Find infrequent aspects based on frequent aspects and opinion words:



Aspect Extraction Rule based on Linguistic Dependency Relationships

Basic Dependencies:



O ----> *O-Dep* ----> *A*

Opinion words (e.g., "good", "amazing")

Dependency relation, e.g., adjectival modifier (amod), modifier (mod), etc.

Noun (NN)

OUTPUT:

A = newly extracted aspect

Commonly used rules for Aspect Extraction

	Observations	Output	Examples
$R1_1$ (OA-Rel)	$O \rightarrow O\text{-}Dep \rightarrow A$ s.t. $O \in \{O\}$, $O\text{-}Dep \in \{MR\}$, $POS(A) \in \{NN\}$	$a = A$	*The phone has a good "screen".* *good→**mod**→screen*
$R1_2$ (OA-Rel)	$O \rightarrow O\text{-}Dep \rightarrow H \leftarrow A\text{-}Dep \leftarrow A$ s.t. $O \in \{O\}$, $O/A\text{-}Dep \in \{MR\}$, $POS(A) \in \{NN\}$	$a = A$	*"iPod" is the best mp3 player.* *best→**mod**→player←**subj**←iPod*
$R2_1$ (OA-Rel)	$O \rightarrow O\text{-}Dep \rightarrow A$ s.t. $A \in \{A\}$, $O\text{-}Dep \in \{MR\}$, $POS(O) \in \{JJ\}$	$o = O$	same as $R1_1$ with *screen* as the known word and *good* as the extracted word
$R2_2$ (OA-Rel)	$O \rightarrow O\text{-}Dep \rightarrow H \leftarrow A\text{-}Dep \leftarrow A$ s.t. $A \in \{A\}$, $O/A\text{-}Dep \in \{MR\}$, $POS(O) \in \{JJ\}$	$o = O$	same as $R1_2$ with *iPod* is the known word and *best* as the extract word.
$R3_1$ (AA-Rel)	$A_{i(j)} \rightarrow A_{i(j)}\text{-}Dep \rightarrow A_{j(i)}$ s.t. $A_{j(i)} \in \{A\}$, $A_{i(j)}\text{-}Dep \in \{CONJ\}$, $POS(A_{i(j)}) \in \{NN\}$	$a = A_{i(j)}$	*Does the player play dvd with audio and "video"?* *video→**conj**→audio*
$R3_2$ (AA-Rel)	$A_i \rightarrow A_i\text{-}Dep \rightarrow H \leftarrow A_j\text{-}Dep \leftarrow A_j$ s.t. $A_i \in \{A\}$, $A_i\text{-}Dep = A_j\text{-}Dep$ OR ($A_i\text{-}Dep = subj$ AND $A_j\text{-}Dep = obj$), $POS(A_j) \in \{NN\}$	$a = A_j$	*Canon "G3" has a great len.* *len→**obj**→has←**subj**←G3*
$R4_1$ (OO-Rel)	$O_{i(j)} \rightarrow O_{i(j)}\text{-}Dep \rightarrow O_{j(i)}$ s.t. $O_{j(i)} \in \{O\}$, $O_{i(j)}\text{-}Dep \in \{CONJ\}$, $POS(O_{i(j)}) \in \{JJ\}$	$o = O_{i(j)}$	*The camera is amazing and "easy" to use.* *easy→**conj**→amazing*
$R4_2$ (OO-Rel)	$O_i \rightarrow O_i\text{-}Dep \rightarrow H \leftarrow O_j\text{-}Dep \leftarrow O_j$ s.t. $O_i \in \{O\}$, $O_i\text{-}Dep = O_j\text{-}Dep$ OR ($O_i/O_j\text{-}Dep \in \{pnmod, mod\}$), $POS(O_j) \in \{JJ\}$	$o = O_j$	*If you want to buy a sexy, "cool", accessory-available mp3 player, you can choose iPod.* *sexy→**mod**→player←**mod**←cool*

Aspect Sentiment Classification

- Problem Definition: Given a sentence s, two subtasks are performed:
  - Subtask 1 Subjectivity classification: Determine whether s is a subjective sentence or an objective sentence
  - Subtask 2 Sentiment classification: If s is subjective, determine whether it expresses a positive, negative, or neutral opinion

Methods for Aspect Sentiment Classification

- If labels are available: any Supervised Classification model
  - Naive Bayes
  - Logistic Regression
  - SVM
  - Neural Networks
  - …
- Lexicon-based approach
  - It involves the use of a sentiment lexicon, which is a collection of words (usually, adjectives and/or adverbs) that have been labeled with their corresponding sentiment polarity (positive, negative, or neutral).
  - Assign a sentiment score to each word in the text based on its presence in the sentiment lexicon.
  - Combine the scores to obtain an overall sentiment score for the text.

References

- Liu, Bing. Web data mining: exploring hyperlinks, contents, and usage data. Berlin: springer, 2011. Chapter 11.