第十二章：

文本情感分析技术

杨建武

北京大学计算机科学技术研究所

Email:yangjianwu@icst.pku.edu.cn

情感计算的概念

- 情感计算（Affective Computing）
  - 通过计算机技术，自动分析文本、图像或视音频等对象所包含的情感倾向及其强度
    - 例如：正面或负面、喜欢或讨厌、快乐或悲伤、愤怒和恐惧等
  - 情感计算的分类
    - 主观性（Subjectivity）
      - 主观性、客观性和中性
    - 情感倾向（Orientation）
      - 正面（褒义）、负面（贬义）和中性

情感计算的应用

- Businesses and organizations: product and service benchmarking. Market intelligence.
  - Business spends a huge amount of money to find consumer sentiments and opinions.
    - Consultants, surveys and focused groups, etc
- Individuals: interested in other's opinions when
  - Purchasing a product or using a service,
  - Finding opinions on political topics,
- Ads placements: Placing ads in the user-generated content
  - Place an ad when one praises a product.
  - Place an ad from a competitor if one criticizes a product.
- Opinion retrieval/search: providing general search for opinions.

文本情感计算

- 词或短语的情感倾向
- 文档与句子的情感倾向
- 观点挖掘
  - 基于特征的观点挖掘
  - 比较式观点挖掘

词语的情感倾向

- **Opinion Words or Phrases** (also called polar words, opinion bearing words, etc). E.g.,
  - Positive: beautiful, wonderful, good, amazing
  - Negative: bad, poor, terrible
- **Important to note:**
  - **Some opinion words are context independent (e.g., good).**
  - **Some are context dependent (e.g., long)**.
- **Three main ways to compile such a list:**
  - Manual approach: not a bad idea, only an one-time effort
  - Corpus-based approaches
  - Dictionary-based approaches

词语的情感倾向

- 1997年，Hatzivassiloglou等人通过连词的语义约束计算形容词的情感倾向
- 2002年，Turney等人提出利用搜索引擎查询词之间的互信息(PMI)：AltaVista的Near操作符, “excellent”and “poor”
- 2003年，Turney等人又提出基于潜在语义分析（LSA）计算词语的语义倾向
- 2004年，Kamps等人提出基于WordNet 的方法，通过计算词与“good”and“bad”之间的语义距离来作为分类标注

SO-PMI

- Measuring Praise and Criticism: Inference of Semantic Orientation from Association (TURNEY 2003)
- SO-PMI (Semantic Orientation from Pointwise Mutual Information)

$Pwords$ = a set of words with positive semantic orientation

$Nwords$ = a set of words with negative semantic orientation

$\mathrm{A}(word_1, word_2)$ = a measure of association between $word_1$ and $word_2$

$$\text{SO-A}(word) = \sum_{pword \in Pwords} \mathrm{A}(word, pword) - \sum_{nword \in Nwords} \mathrm{A}(word, nword).$$

SO-PMI

$$\mathrm{PMI}(word_1, word_2) = \log_2\left(\frac{\frac{1}{N}\,\mathrm{hits}(word1\ \mathrm{NEAR}\ word2)}{\frac{1}{N}\,\mathrm{hits}(word1)\ \ \frac{1}{N}\,\mathrm{hits}(word2)}\right)$$

N is the total number of documents indexed by the search engine.

$$\mathrm{SO\text{-}PMI}(word)$$

$$= \log_2\left(\frac{\prod_{pword \in Pwords} \mathrm{hits}(word\ \ \mathrm{NEAR}\ \ pword) \cdot \prod_{nword \in Nwords} \mathrm{hits}(nword)}{\prod_{pword \in Pwords} \mathrm{hits}(pword) \cdot \prod_{nword \in Nwords} \mathrm{hits}(word\ \ \mathrm{NEAR}\ \ nword)}\right)$$

Dictionary-based approaches

- Typically use WordNet's synsets and hierarchies to acquire opinion words
  - Start with a small seed set of opinion words.
  - Use the set to search for synonyms and antonyms in WordNet (Hu and Liu, KDD-04; Kim and Hovy, COLING-04).
  - Manual inspection may be used afterward.
- Use additional information (e.g., glosses 注释) from WordNet and learning
  - (Andreevskaia and Bergler, EACL-06)
  - (Esuti and Sebastiani, CIKM-05)
- Weakness of the approach
  - Do not find context dependent opinion words,
    - e.g., small, long, fast.
- 中文资源： HowNet、同义词词林

Documents Sentiment classification

- Classify documents (e.g., reviews) based on the overall sentiments expressed by opinion holders (authors),
  - Positive, negative, and (possibly) neutral
- Similar but different from topic-based text classification.
  - In topic-based text classification, topic words are important.
  - In sentiment classification, sentiment words are more important, e.g., great, excellent, horrible, bad, worst, etc.

文章的倾向分析

- 2003年，Turney用评论中出现的词语的倾向的平均值来代表整篇评论的倾向；
- 2003年，Dave等用词的倾向代表文章的倾向，考虑了词的倾向强度；
- 2002年，Bo Pang等人首先在情感分析领域引入了机器学习的方法，利用Naïve Bayes、Max Entropy、SVM等分类，在文档级别上对文档进行自动的情感分类；（作者通过IMDB收集了具有标注的电影评论）
- 2004年，Bo Pang等人又提出通过机器学习和图中最小割的方法对文档中的句子进行主观性判断；
- 2005年，Bo Pang等人进一步拓展了他们的工作，通过机器学习的方法对电影评论进行3级或4级打分。

Unsupervised review classification

- (Turney, ACL-02)
- Data: reviews from epinions.com on automobiles, banks, movies, and travel destinations.
- The approach: Three steps
- Step 1:
  - Part-of-speech tagging
  - Extracting two consecutive words (two-word phrases) from reviews if their tags conform to some given patterns, e.g., (1) JJ, (2) NN.

Unsupervised review classification

- Step 2: Estimate the semantic orientation (SO) of the extracted phrases
  - Use Pointwise mutual information

$$PMI(word_1, word_2) = \log_2\left(\frac{P(word_1 \wedge word_2)}{P(word_1)P(word_2)}\right)$$

  - Semantic orientation (SO):

$$SO(phrase) = PMI(phrase, \text{“excellent”}) - PMI(phrase, \text{“poor”})$$

  - Using AltaVista near operator to do search to find the number of hits to compute PMI and SO.

Unsupervised review classification

- Step 3: Compute the average SO of all phrases
  - classify the review as recommended if average SO is positive, not recommended otherwise.
- Final classification accuracy:
  - automobiles - 84%
  - banks - 80%
  - movies - 65.83
  - travel destinations - 70.53%

Sentiment classification using machine learning methods

- (Pang et al, EMNLP-02)
- This paper directly applied several machine learning techniques to classify movie reviews into positive and negative.
- Three classification techniques were tried:
  - Naïve Bayes
  - Maximum entropy
  - Support vector machine
- Pre-processing settings: negation tag, unigram (single words), bigram, POS tag, position.
  - SVM: the best accuracy 83% (unigram)

Sentence-level sentiment analysis

- Document-level sentiment classification is too coarse for most applications.
- Much of the work on sentence level sentiment analysis focuses on identifying subjective sentences in news articles.
  - Classification: objective and subjective.
  - All techniques use some forms of machine learning.
  - E.g., using a naïve Bayesian classifier with a set of data features/attributes extracted from training sentences (Wiebe et al. ACL-99).

Let us go further?

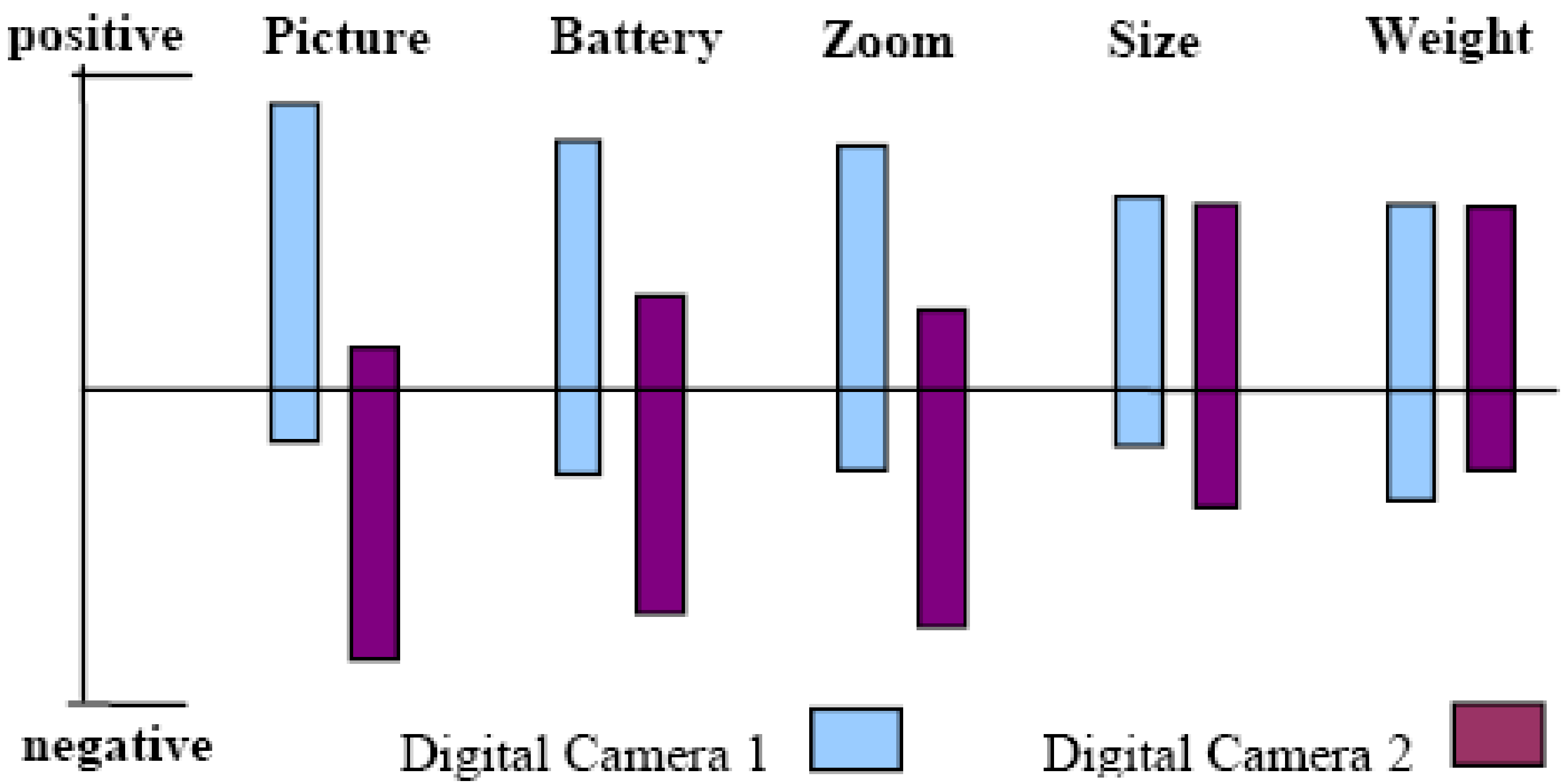
- Sentiment classification at both document and sentence (or clause) levels are useful, but
  - They do not find what the opinion holder like and dislike.
- An negative sentiment on an object
  - does not mean that the opinion holder dislikes everything about the object.
- A positive sentiment on an object
  - does not mean that the opinion holder likes everything aboutthe object.
- **We need to go to the feature level.**

观点挖掘

- 观点挖掘
  - 目的：从文档或者文档集合中挖掘出评论对象以及对该对象的观点；
  - 与文档级别的情感分类相比较：观点挖掘需要在更细的粒度上对文档进行情感分析；
  - 观点挖掘往往采用紧密相关的信息抽取技术，用来发现文章内的对象、及其相应观点。

观点挖掘

- Bing Liu：原型系统Opinion Observer



Opinion mining

- (Hu and Liu, KDD-04; Liu, Web Data Mining book 2007)
- Basic components of an opinion
  - Opinion holder: The person or organization that holds a specific opinion on a particular object.
  - Object: on which an opinion is expressed
  - Opinion: a view, attitude, or appraisal on an object from an opinion holder.

Opinion mining tasks

- Opinion holders:
- At the document (or review) level
- At the sentence level
- At the feature level

Opinion mining tasks

- Opinion holders:
  - identify holders is useful, e.g., in news articles, etc,
  - but they are usually known in the user generated content, i.e., authors of the posts.
- At the document (or review) level:
  - Task: sentiment classification of reviews
    - Classes: positive, negative, and neutral
    - Assumption: each document (or review) focuses on a single object (not true in many discussion posts) and contains opinion from a single opinion holder.

Opinion mining tasks

- At the sentence level:
  - Task 1: identifying subjective/opinionated sentences
    - Classes: objective and subjective (opinionated)
  - Task 2: sentiment classification of sentences
    - Classes: positive, negative and neutral.
    - Assumption: a sentence contains only one opinion
      - not true in many cases.
    - Then we can also consider clauses or phrases.

Opinion Sentence

- 美國各大報紙 認為 這個判決很可能敲響美國菸草工業的喪鐘

不特定發表者 發表的動作 發表的(負面)意見 負面意見詞彙

標記結果：

<OPINION_SRC type="IMP">美國各大報紙</OPINION_SRC>
<OPINION_OPR type="PSV">認為</OPINION_OPR>，
<SEN_ATTITUDE type="NSP">
    這個判決很可能敲響美國菸草工業的
    <SENTIMENT_KW type="NEG">喪鐘</SENTIMENT_KW>
</SEN_ATTITUDE>

Opinion Sentence

- 阿莫迪歐 說：「多少錢都不能改變我進食的方式。」

特定發表者 發表的動作 否定詞彙 發表的負面意見

- 標記結果

<OPINION_SRC type="EXP">阿莫迪歐</OPINION_SRC>

<OPINION_OPR type="PSV">說</OPINION_OPR>：「<SEN_ATTITUDE type="NSP">多少錢都不能改變我進食的方式</SEN_ATTITUDE>。」

目前否定詞彙不在標記集中，而是以人工收集的方式獲得。

Opinion Extraction Opinion Extraction at Sentence Level

Algorithm: *Opinion Sentence Extraction*

1. **For** every sentence p
2. **For** every sentiment word in p
3. **If** a negation operator appears before, then reverse the sentiment tendency.
4. **Decide** the opinionated tendency of p by the function of sentiment words and the opinion holder as follows.

$$S_p = S_{opinion-holder} \times \sum_{j=1}^{n} S_{w_j}$$

Where S_p, $S_{opinion\text{-}holder}$, and S_{wj} are the sentiment score of sentence p, the weight of *opinion holder*, and the sentiment score of word w_j, respectively, and n is the total number of sentiment words in p.

Opinion mining tasks

- At the feature level:
  - *Task* 1: Identifying and extracting object features that have been commented on in each review.
  - *Task* 2: Determining whether the opinions on the features are positive, negative or neutral.
  - *Task* 3: Grouping synonyms of features.
  - Produce a feature-based opinion summary of multiple reviews (more on this later).

More at the feature level

- **Problem 1**: Both *F* and *W* are unknown.
  - We need to perform all three tasks:
- **Problem 2**: *F* is known but *W* is unknown.
  - All three tasks are still needed. Task 3 is easier. It becomes the problem of matching the discovered features with the set of given features *F*.
- **Problem 3**: *W* is known (*F* is known too).
  - Only task 2 is needed.

F: the set of features

W: synonyms of each feature

Feature-based opinion extraction and summarization

- Object (Definition): (Liu's Web mining book 2006)
  - An object O is an entity which can be a product, person, event, organization, or topic.
  - O is represented as a tree or taxonomy of components (or parts), sub-components, and so on.
    - Each node represents a component and is associated with a set of attributes.
    - O is the root node (which also has a set of attributes)
- An opinion can be expressed on any node or any attribute of the node.
- To simplify our discussion, we use "**features**" to represent both components and attributes
  - The term "feature" should be understood in a ***broad sense***,
    - Product feature, topic or sub-topic, event or sub-event, etc
  - Note: the object *O* itself is also a feature.

Review Format

- Format 1 - Pros, Cons and detailed review: The reviewer is asked to describe Pros and Cons separately and also write a detailed review.

My SLR is on the shelf

by camerafun4. Aug 09 '04

Pros: Great photos, easy to use, very small
Cons: Battery usage; included memory is stingy.

I had never used a digital camera prior to purchasing this Cannon A70. I have always used a SLR … **Read the full review**

Review Format

- Format 2 - Pros and Cons: The reviewer is asked to describe Pros and Cons separately.

Pros:
It's small in size, and the rotatable lens is great. It's very easy to use, and has fast response from the shutter. The LCD has increased from 1.5 in to 1.8, which gives bigger view. It has lots of modes to choose from in order to take better pictures.

Cons:
It almost has no cons, it would be better if the LCD is bigger and it's going to be best if the model is designed to a smaller size.

Review Format

- Format 3 - free format: The reviewer can write freely, i.e., no separation of Pros and Cons.

GREAT Camera., Jun 3, 2004

Reviewer: **jprice174** from Atlanta, Ga.

I did a lot of research last year before I bought this camera... It kinda hurt to leave behind my beloved nikon 35mm SLR, but I was going to Italy, and I needed something smaller, and digital.

The pictures coming out of this camera are amazing. The 'auto' feature takes great pictures most of the time. And with digital, you're not wasting film if the picture doesn't come out. …

Feature extraction from Pros and Cons of Format 1

- (Liu et al WWW-03, Hu and Liu 2005)
- **Observation**:
  - Each sentence segment in Pros or Cons contains only one feature.
  - Sentence segments can be separated by commas, periods, semi-colons, hyphens, ‘&’’s, ‘and’’s, ‘but’’s, etc.
- Pros in Example 1 can be separated into 3 segments:
  - great photos <photo>
  - easy to use <use>
  - very small <small> ⇒ <size>
- Cons can be separated into 2 segments:
  - battery usage <battery>
  - included memory is stingy <memory>

My SLR is on the shelf

by camerafun4. Aug 09 ‘04

Pros: Great photos, easy to use, very small

Cons: Battery usage; included memory is stingy.

Extraction using label sequential rules

- Liu's Web mining book 2006
- Label sequential rules (LSR) are a special kind of sequential patterns, discovered from sequences.
- LSR Mining is supervised.
- The training data set is a set of sequences, e.g.,

  "*Included memory is stingy*"

  is turned into a sequence with POS tags.

  ⟨{included, VB}{memory, NN}{is, VB}{stingy, JJ}⟩

  then turned into

  ⟨{included, VB}{$feature, NN}{is, VB}{stingy, JJ}⟩

Using LSRs for extraction

- Based on a set of training sequences, we can mine label sequential rules, e.g.,

$$\langle\{easy, JJ\}\{to\}\{*, VB\}\rangle \rightarrow \langle\{easy, JJ\}\{to\}\{\$feature, VB\}\rangle$$

[sup = 10%, conf = 95%]

- Feature Extraction
  - Only the right hand side of each rule is needed.
  - The word in the sentence segment of a new review that matches $feature is extracted.
  - We need to deal with conflict resolution also (multiple rules are applicable).

Some results

	Pros		Cons	
	recall	prec	Recall	prec
data1	0.862	0.857	0.865	0.794
data2	0.937	0.937	0.824	0.806
data3	0.817	0.817	0.730	0.741
data4	0.919	0.914	0.745	0.708
data5	0.911	0.904	0.883	0.900
Avg.	**0.889**	**0.886**	**0.809**	**0.790**

Extraction of features of formats 2 and 3

- Reviews of these formats are usually complete sentences, e.g.:
  - "the pictures are very clear."
    - Explicit feature: picture
  - "It is small enough to fit easily in a coat pocket or purse."
    - Implicit feature: size
- Extraction: Frequency based approach
  - Frequent features
  - Infrequent features

Frequency based approach

- (Hu and Liu, KDD-04; Liu's Web mining book 2006)
- Frequent features: those features that have been talked about by many reviewers.
- Use sequential pattern mining (association mining is less suitable): co-location
- Why the frequency based approach?
  - Different reviewers tell different stories (irrelevant)
  - When product features are discussed, the words that they use converge (集中).
  - Sufficient for practical use: They are main features.
- Sequential pattern mining finds frequent phrases.
- "Google Product Search" implemented this approach.

Improvement

- (Popescu and Etzioni, 2005)
- The algorithm tries to remove those frequent noun phrases that may not be product features. To improve precision (with a small drop in recall).
- It tries to identify part-of relationship
  - Each noun phrase is given a PMI (pointwise mutual information) score between the phrase and **part discriminators** associated with the product class, e.g., a scanner class.
  - The part discriminators for the scanner class are, "of scanner", "scanner has", "scanner comes with", etc, which are used to find components or parts of scanners by searching on the Web (the KnowItAll approach).

Infrequent features extraction

- How to find the infrequent features?
- Observation: the same opinion word can be used to describe different features and objects.
  - "The pictures are absolutely amazing."
  - "The software that comes with it is amazing."

Identify feature synonyms

- Liu et al (2005) made an attempt using only WordNet.
- Carenini et al (2005) proposed a more sophisticated method based on several similarity metrics, but it requires a taxonomy of features to be given.
  - The system merges each discovered feature to a feature node in the taxonomy.
  - The similarity metrics are defined based on string similarity, synonyms and other distances measured using WordNet.
  - Experimental results based on digital camera and DVD reviews show promising results.
- Many ideas in information integration are applicable.

Identify opinion orientation of feature

- For each feature, we identify the sentiment or opinion orientation expressed by a reviewer.
- We work based on sentences, but also consider,
  - A sentence can contain multiple features.
  - Different features may have different opinions.
  - E.g., The battery life and picture quality are *great* (+), but the view founder is *small* (-).
- Almost all approaches make use of opinion words and phrases. But notice again:
  - Some opinion words have context independent orientations, e.g., “great”.
  - Some other opinion words have context dependent orientations, e.g., “small”
- Many ways to use them.

Aggregation of opinion words

- (Hu and Liu, KDD-04)
- Input: a pair (*f*, *s*), where *f* is a product feature and *s* is a sentence that contains *f*.
- Output: whether the opinion on *f* in *s* is positive, negative, or neutral.
- Two steps:
  - Step 1: split the sentence if needed based on BUT words (but, except that, etc).
  - Step 2: work on the segment s_f containing *f*.
    - Let the set of opinion words in s_f be w_1, .., w_n.
    - Sum up their orientations (1, -1, 0),
    - and assign the orientation to (*f*, *s*) accordingly.

Aggregation of opinion words

- In (Ding and Liu, 2007)
  - step 2 is changed to

$$\sum_{i=1}^{n} \frac{w_i.o}{d(w_i, f)}$$

with better results.

$w_i.o$ is the opinion orientation of w_i.

$d(w_i, f)$ is the distance from f to w_i.

Context dependent opinions

- Popescu and Etzioni (EMNLP-05) used
  - constraints of connectives in (Hazivassiloglou and McKeown, ACL-97), and some additional constraints, e.g., morphological relationships, synonymy and antonymy, and
  - relaxation labeling to propagate opinion orientations to words and features.
- Ding and Liu (2007) used
  - constraints of connectives both at intra-sentence and intersentence levels, and
  - additional constraints of, e.g., TOO, BUT, NEGATION, ….
- to directly assign opinions to (*f*, *s*) with good results (>0.85 of F-score).

Sample results

- Feature extraction: Recall $\approx$ 0.80%, precision $\approx$ 70+% of feature extraction (Hu and Liu 2004)
- The precision result was improved subsequently in (Popescu and Etzioni, 2005), but with drop in recall.
- Opinion orientation classification: 70+ to 80+% in Fscore.
- Many other researchers have worked on releated problems, e.g.,
  - Carenini et al (2005), Hu and Liu (2004, 2006), Kim and Hovy (2004), Liu et al. (2005), Kobayashi et al. (2005), Ku et al. (2005) and Morinaga et al. (2002), Popescu and Etzioni (2005), Yi et al. (2003), etc.

Two types of evaluations

- Direct Opinions: sentiment expressions on some objects/entities, e.g., products, events, topics, individuals, organizations, etc
  - E.g., “the picture quality of this camera is great”
  - Subjective
- Comparisons: relations expressing similarities, differences, or ordering of more than one objects.
  - E.g., “car x is cheaper than car y.”
  - Objective or subjective

Types of Comparatives: Gradable

- *Gradable*
  - *Non-Equal Gradable*: Relations of the type *greater* or *less than*
    - *Keywords like better, ahead, beats, etc*
    - *Ex: "optics of camera A is better than that of camera B"*
  - *Equative*: Relations of the type *equal to*
    - Keywords and phrases like *equal to*, *same as*, *both*, *all*
    - Ex: "*camera A and camera B both come in 7MP*"
  - *Superlative*（最高级）: Relations of the type *greater* or *less than all others*
    - Keywords and phrases like *best, most, better than all*
    - Ex: "*camera A is the cheapest camera available in market*"

Types of comparatives: non-gradable

- Non-Gradable: Sentences that compare features of two or more objects, but do not grade them. Sentences which imply:
  - Object A is similar to or different from Object B with regard to some features.
  - Object A has feature F1, Object B has feature F2 (F1 and F2 are usually substitutable).
  - Object A has feature F, but object B does not have.

Comparative Relation: gradable

- **Definition**: A gradable comparative relation captures the essence of a gradable comparative sentence and is represented with the following: (relationWord, features, entityS1, entityS2, type)
  - *relationWord*: The keyword used to express a comparative relation in a sentence.
  - *features*: a set of features being compared.
  - *entityS1* and *entityS2*: Sets of entities being compared.
  - *type*: *non-equal gradable*, *equative* or *superlative*.

Examples: Comparative relations

- Ex1: "*car X has better controls than car Y*"

  (relationWord = better, features = controls, entityS1 = car X, entityS2 = car Y, type = non-equal-gradable)
- Ex2: "*car X and car Y have equal mileage*"

  (relationWord = equal, features = mileage, entityS1 = car X, entityS2 = car Y, type = equative)
- Ex3: "*Car X is cheaper than both car Y and car Z*"

  (relationWord = cheaper, features = null, entityS1 = car X, entityS2 = {car Y, car Z}, type = non-equal-gradable)
- Ex4: "company X produces a variety of cars, but still best cars come from company Y"

  (relationWord = best, features = cars, entityS1 = company Y, entityS2 = null, type = superlative)

Tasks

- Given a collection of evaluative texts
  - Task 1: Identify comparative sentences.
  - Task 2: Categorize different types of comparative sentences.
  - Task 3: Extract comparative relations from the sentences.

Identify comparative sentences

- (Jinal and Liu, SIGIR-06)

Keyword strategy

- An observation:
  - It is easy to find a small set of keywords that covers almost all comparative sentences, with a very high recall and a reasonable precision
- We have compiled a list of **83 keywords** used in comparative sentences, which includes:
  - Words with POS tags of JJR, JJS, RBR, RBS
    - POS tags are used as keyword instead of individual words.
    - Exceptions: more, less, most and least
  - Other indicative words like *beat*, *exceed*, ahead, etc
  - Phrases like *in the lead*, *on par with*, etc

Identify comparative sentences

- 2-step learning strategy:
- **Step1**: Extract sentences which contain at least a keyword (recall = 98%, precision = 32% on our data set for gradables)
- **Step2**: Use the naïve Bayes (NB) classifier to classify sentences into two classes
  - comparative and non-comparative.
  - Attributes: **class sequential rules** (CSRs)
    - Use words within radius *r* of a keyword to form a sequence (words are replaced with POS tags)

Classify different types of comparatives

- Classify comparative sentences into three types: non-equal gradable, equative, and superlative
  - SVM learner gave the best result.
  - Attribute set is the set of keywords.
  - If the sentence has a particular keyword in the attribute set, the corresponding value is 1, and 0 otherwise.

Extraction of comparative relations

- (Jindal and Liu, AAAI-06; Liu's Web mining book 2006)
- Assumptions
  - There is only one relation in a sentence.
  - Entities and features are nouns (includes nouns, plural nouns and proper nouns) and pronouns.
    - Adjectival comparatives
    - Does not deal with adverbial comparatives
- 3 steps
  - Sequence data generation
  - Label sequential rule (LSR) generation
  - Build a sequential cover/extractor from LSRs

Sequence data generation

- Label Set = {$entityS1, $entityS2, $feature}
- Three labels are used as pivots to generate sequences.
  - Radius of 4 for optimal results
- Following words are also added
  - Distance words = {*l*1, *l*2, *l*3, *l*4, *r*1, *r*2, *r*3, *r*4}, where
    - "*li*" means distance of *i* to the left of the pivot (中心点).
    - "*ri*" means the distance of *i* to the right of pivot.
  - Special words #*start* and #*end* are used to mark the start and the end of a sentence.

Sequence data generation example

- The comparative sentence
- *"Canon/NNP has/VBZ better/JJR optics/NNS"*
  - *$entityS1* "*Canon*" and *$feature* "*optics*".
- **Sequences are:**
  - ⟨{#start}{*l*1}{**$entityS1**, NNP}{*r*1}{has, VBZ }{r2} {better, JJR}{*r*3}{$Feature, NNS}{r4}{#end}⟩
  - ⟨{#start}{*l*4}{**$entityS1**, NNP}{*l*3}{has, VBZ}{*l*2} {better, JJR}{*l*1}{**$Feature**, NNS}{*r*1}{#end}⟩

Build a sequential cover from LSRs

- LSR: $\langle\{*, NN\}\{VBZ\}\rangle \rightarrow \langle\{\$entityS1, NN\}\{VBZ\}\rangle$
  - Select the LSR rule with the highest confidence. Replace the matched elements in the sentences that satisfy the rule with the labels in the rule.
  - Recalculate the confidence of each remaining rule based on the modified data from step 1.
  - Repeat step 1 and 2 until no rule left with confidence higher than the *minconf* value (we used 90%).

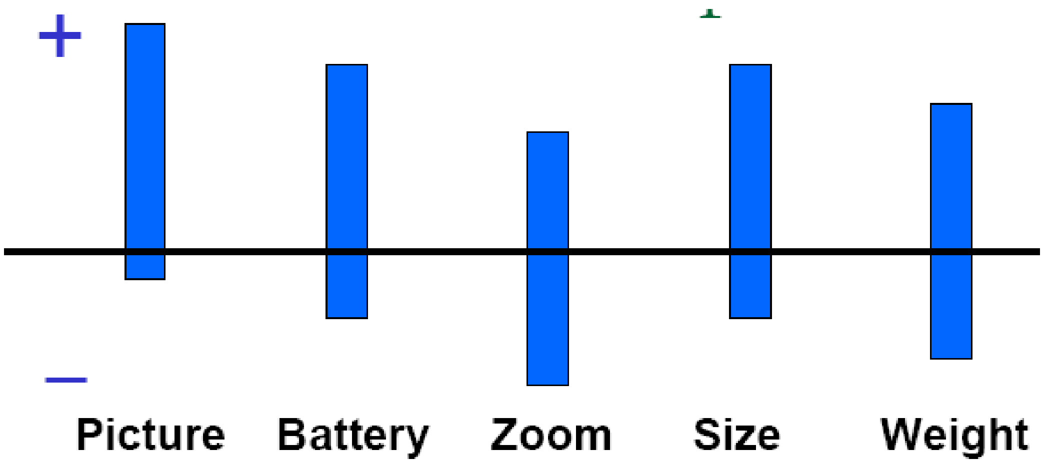
(Details skipped)

Experimental results

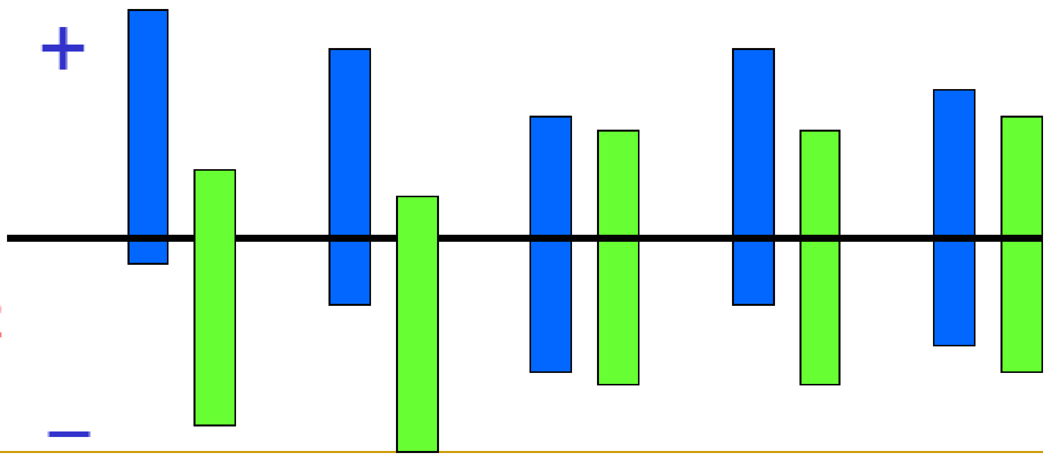
- Identifying Gradable Comparative Sentences
  - precision = 82% and recall = 81%.
- Classification into three gradable types
  - SVM gave accuracy of 96%
- Extraction of comparative relations
  - LSR (label sequential rules): F-score = 72%

Visual Summarization & Comparison

- Summary of reviews of
  Digital camera 1



- Comparison of reviews of
  Digital camera 1
  Digital camera 2



观点比较

正方观点:6968票

PK

反方观点:2184票

我继续使用XP	我将投向Vista
XP优势: 1. WinXP可是说是微软最成功的家用平台，系统非常成熟完善 2. 非高配本情况下，XP运行软件速度快，低端机也能流畅运行 3. XP对软件和驱动程序的兼容性更宽广，不会影响日常使用 4. 在同一个测试环境下 运行XP系统的笔记本续航时间更长久 5. WinXP覆盖率高，系统早已深入身心，熟练操作不会成为障碍 投上一票	Vista的优势: 1. Vista独有的ReadyBoost技术，可轻松加速系统运行速度 2. 更全面的搜索功能，操作过程随时搜索，无须调用浏览器 3. 支持DX 10图像技术，让游戏娱乐画面更炫更酷（XP无法移植） 4. Vista全面铺货，购买本本基本上全免费预装正版Vista系统 5. 应用Aero，更加华丽的操作界面和插件使一切变的得心应手 投上一票

支持正方观点：(我继续使用XP)

▸ 我来说两句

7/8 的人觉得此评论有用

ZOL网友

在现阶段，我还是会使用XP的，一、对XP熟悉，二、我是一个游戏爱好者，VISTA占用内存过多，影响电脑性能，三、现在MS还是继续推出XP补丁，XP还没有走到尽头。等哪天XP淘汰了，VISTA肯定会具有亲和力的姿态来对待大家...[全文]

12-06 17:44 | 10回应 | 此评论对你 有用 无用

支持反方观点：(我将投向Vista)

▸ 一个菜鸟对VISTA的评价

5/10 的人觉得此评论有用

ZOL网友

首先声明我只是个菜鸟!对与VISTA的出现我并不觉得反感，正如楼上所说\"时代在进步\"有些人认为VISTA系统所消耗的空间太大(系统的大小)，说句实在点的话，我们用电脑的人，80G硬盘不够你用的吗? 何必在乎这点呢?最重要的其实还是系统与应用软件的不兼容性~毕竟现在我们用的好多软件是在 XP 系统下开发与研制的.我相信微软会很快解决著个问题!这个过渡期还是

对比检索

爱搜车首页

爱搜车·众评　众评　经验　问答　热帖

车型搜索　isoche有什么不一样?

众评首页 > 车型对比 > 雅阁 vs 凯美瑞

雅阁

浏览此车

满意度 点击查看详情

统计各大论坛网友的发言，
自动计算获得，仅作参考

油耗	67	满意：167 不满：81
安全性	78	满意：98 不满：28
空间	79	满意：184 不满：49
动力	79	满意：268 不满：72
操控	63	满意：109 不满：64
外观	84	满意：410 不满：78
内饰	80	满意：210 不满：52
性价比	70	满意：218 不满：92
配置	85	满意：157 不满：28

贴身对比

凯美瑞跟雅阁哪个好啊
凯美瑞跟雅阁哪个好啊 我明天就要买了啊~`希望大家能给我个答案 此帖出自：http://bbs.pcauto.com.cn。我个人..喜欢...KMR多点..哈哈。KMR的外型YG的内在！看楼主 搜...
太平洋汽车网　发布日:2008-07-03　浏览:138　回复:9 - 快照

凯美瑞跟雅阁哪个好啊 急哦！　希望大家能给我个答...
我明天就要买车了 不知道 凯美瑞跟雅阁哪个好啊 此帖出自：http://bbs.pcauto.com.cn。都不錯 我個人是選擇 凱美瑞。Re:[zhangfuliang,1楼][quote]...
太平洋汽车网　发布日:2008-07-03　浏览:204　回复:10 - 快照

新天籁即将上市 PK雅阁凯美瑞仅售21万?
雅阁、凯美瑞、天籁是国内三大主流日系中级车。从销量上看，雅阁、凯美瑞遥遥领先天籁，为了扭转目前的不利局面，东风日产在北京车展上高调推出全新一代天籁，它也将成为2008年东风日产最重要的新车。 按照计划...
汽车之家　发布日:2008-07-03　浏览:191　回复:8 - 快照

《致胜，凯美瑞，雅阁3款自然吸气中型车全方位对比...
雅阁加速成绩优异得益于先进的发动机和较窄的轮胎，致胜的2.3升发动机对于拉动这么庞大的车身毫无心得，外加偏重操控的235mm轮胎的使用，致胜在加速过程中给人肉呼呼的感觉。 凯美瑞给人也是肉...
西祠胡同　发布日:2008-07-02　浏览:33　回复:4 - 快照

［凯美瑞］ 和八代雅阁的竞争！！！！
凯美瑞4S店。现车销售颜色齐全，可销售全国各地上牌凯美瑞4S店。现车销售颜色齐全，可销售全国各地凯美瑞2.0E自动 19.78万 销售价格18.68万凯美瑞2.0G自动 2...
汽车之家　发布日:2008-07-02　浏览:253　回复:5 - 快照

凯美瑞雅阁花冠被福布斯评为10大少年和学生用车
凯美瑞，丫阁，花冠被福布斯评为10大少年和学生用车。在我国可是中高级车。我们都是老年用车。哈哈。偷偷的打酱油 BO22JQ。...
汽车之家　发布日:2008-07-02　浏览:262　回复:11 - 快照

铂锐 雅阁 凯美瑞三个车买哪个好！
雅阁 凯美瑞三个车买哪个好！ 此帖出自：http://bbs.pcauto.com.cn。Re:[xiatq,1楼]那就看你自己喜欢哪个了，铂锐，有点美国车的感觉，就是不够力雅阁，想变得运动一点凯美瑞...
太平洋汽车网　发布日:2008-06-28　浏览:100　回复:6 - 快照

八代雅阁就是没凯美瑞强
关于凯美瑞PK八代雅阁话题不同看法请各楼主发表评论:handshake:。能说说八代雅阁为

凯美瑞

浏览此车

满意度 点击查看详情

统计各大论坛网友的发言，
自动计算获得，仅作参考

油耗	58	满意：90 不满：66
安全性	81	满意：80 不满：18
空间	74	满意：113 不满：40
动力	56	满意：115 不满：89
操控	46	满意：72 不满：83
外观	89	满意：389 不满：47
内饰	74	满意：126 不满：43
性价比	49	满意：116 不满：120
配置	83	满意：121 不满：25

话题比较

朝鲜炸毁核设施
落实六方会谈第二阶段成果
美朝就全面检查核计划达成原则共识
相关专题
朝鲜宣布拥有核武器
2005年2月
视频新闻
朝鲜炸毁核设施冷却塔
分析评论
六方会谈后的朝核进程
背景资料与往日回顾
往日专题回顾
历次六方会谈回顾
第六轮六方会谈
第五轮六方会谈
第四轮六方会谈
第三轮六方会谈
第二轮六方会谈
第一轮六方会谈
朝核问题大事记

伊朗拒绝停止核活动
滚动新报 | 最新消息 | 图片报道 | 专题图集 | 视频汇总 | 评论分析 | 相关专题：伊朗重启核燃料设施
伊朗拟在纳坦兹核设施装配五万台离心机
伊朗核问题示意图
最新进展
内贾德称伊朗将为其核计划抗争到底（图）
伊朗军队已提高戒备等级防止核设施遭袭
美抨击伊朗核活动违反安理会决议
潘基文敦促伊朗执行安理会决议
伊朗称其已运行离心机数量不久将揭晓
俄官员称伊朗在核问题上向国际社会挑衅
伊朗宣布已能进行工业规模核燃料生产
伊朗官员不顾安理会制裁决议访俄
请点击播放视频
伊朗核计划发展状况
各方反应
伊朗：威胁将退出核不扩散条约
美国：抨击伊朗核活动违反联合国安理会决议
俄罗斯：官员称伊朗在核问题上向国际社会挑衅
英国：拟推动安理会就伊朗核决议草案强行表决
德国：伊朗重返核谈判大门仍然敞开
法国：建议进一步对伊朗实施制裁
各领域发展状况
事件背景 | 资料
分析 | 评论
伊朗核问题关键期的强硬与弹性
专家称俄撤走在伊朗核人员可能是战争前兆
伊朗核问题在新决议草案推出后走向何方
外国媒体称美国拟效仿科索沃战争攻打伊朗
伊朗军事演示
查看更多图片>>

小结

- 情感计算的概念
- 词或短语的情感倾向
- 文档与句子的情感倾向
- 观点挖掘
  - 基于特征的观点挖掘
  - 比较式观点挖掘

- 主要参考:
  - Bing Liu “From Web Content Mining to Natural Language Processing”(4. Opinion Mining and Summarization), ACL-2007 Tutorial

Any Question?