Mining and Summarizing Opinions on the Web

Bing Liu
Department of Computer Science
University of Illinois at Chicago

Introduction

- Word-of-mouth on the Web
  - The Web has dramatically changed the way that people express their views and opinions.
  - One can express opinions on almost anything at review sites, forums, discussion groups, blogs ..., which are called user generated contents or media.
- Benefits of opinion analysis
  - Businesses and organizations: marketing intelligence, product and service benchmarking and improvement.
  - Individuals: seeing a summary of opinions, rather than reading a lot of reviews, discussions, etc.
  - Research: Intellectually challenging.

Introduction (contd …)

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
- We will use **consumer reviews of products** as examples in this talk

Roadmap

- Sentiment classification
- Feature-based opinion extraction and summarization
  - Problems
  - Some existing techniques
- Comparative sentence and relation extraction
  - Problems
  - Some existing techniques

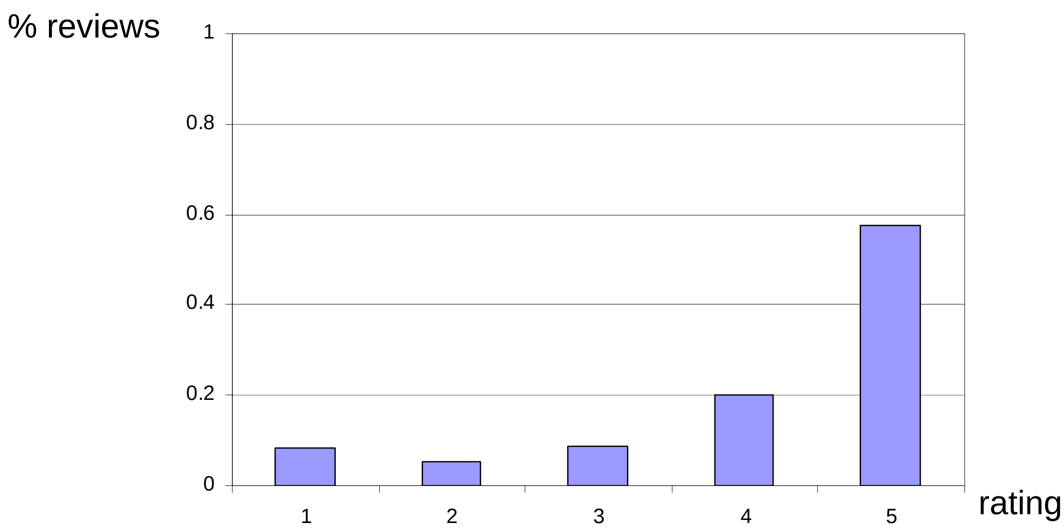
Sentiment classification

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

Sentiment classification

- Many researchers have studied the problem.
- (Turney, 2002), (Pang, Lee and Vaithyanathan 2002), (Dave, Lawrence and Pennock, 2003), (Hatzivassiloglou and Wiebe 2000) (Wiebe and Riloff 2005) (Hearst 1992), (Tong, 2001), (Das and Chen, 2001), (Gamon 2004), (Riloff and Wiebe, 2003), (Wilson, Wiebe and Hwa, 2004), and many more ...

FYI: Review rating distribution



Review rating distribution (Amazon.com)

Can we go further?

- Sentiment classification is useful, but it does not find what the reviewer liked and disliked.
- An negative sentiment on an object does not mean that the reviewer does not like anything about the object.
- A positive sentiment on an object does not mean that the reviewer likes everything.
- Go to the sentence level and feature level.

Roadmap

- Sentiment classification
- Feature-based opinion extraction and summarization
  - Problems
  - Some existing techniques
- Comparative sentence and relation extraction
  - Problems
  - Some existing techniques.

Feature-based opinion extraction and summarization

- We are interesting in what reviewers liked and disliked,
  - features and components
- Since the number of reviews for an object can be large, we want to produce a simple summary of opinions.
- The summary can be easily visualized and compared.

Object/entity (Liu's Web mining book 2006)

- **Definition** (**object**): 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, i.e., we omit the hierarchy.

Model of user evaluation

- An object is represented with a finite set of features, $F = \{f_1, f_2, \ldots, f_n\}$.
  - Each feature f_i in F can be expressed with a finite set of words or phrases W_i, which are **synonyms**.

  That is to say:
  - we have a set of corresponding synonym sets $W = \{W_1, W_2, \ldots, W_n\}$ for the features.
  - since each feature in F has a name, we assume $f_i \in W_i$. Each **opinion holder** j comments on a subset of the features $S_j \subseteq F$.
  - For each feature $f_k \in S_j$ that opinion holder j comments on, he/she
    - chooses a word or a phrase from W_k to describe the feature, and
    - expresses a positive, negative or neutral opinion on f_k.

Three main tasks

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, which is simple after the above three tasks are performed.

Three main problems

- **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 needed. Task 3 is easier. It becomes the problem of matching discovered features with the set of given features *F*.
- **Problem 3**: *W* is known (F is known too).
  - Only task 2 is needed.

Is task 2 needed? (Liu et al WWW-05)

Depending on the 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. Epinions.com uses this format.
- Format 2 - Pros and Cons: The reviewer is asked to describe Pros and Cons separately. C|net.com uses this format.
- Format 3 - free format: The reviewer can write freely, i.e., no separation of Pros and Cons. Amazon.com uses this format.

Example 1: Format 1

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**

Example 2: Format 2

User rating
Perfect
10
out of 10

"It is a great digitbal still camera for this century"
September 1, 2004

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.

Example 3: Format 3 (with summary)

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 Based Summary:

Feature1: **picture**
Positive: 12

- The pictures coming out of this camera are amazing.
- Overall this is a good camera with a really good picture clarity.

…
Negative: 2

- The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
- Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.

Feature2: **battery life**
…

Example of a summary (Hu and Liu, KDD-04)

*Digital_camera_*1:

Feature: picture quality

Positive: 123 <individual review sentences>

Negative: 6 <individual review sentences>

Feature: size

Positive: 82 <individual review sentences>

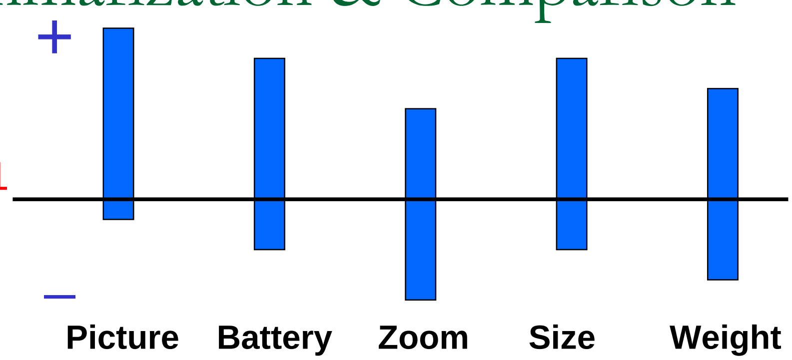
Negative: 10 <individual review sentences>

Feature: …

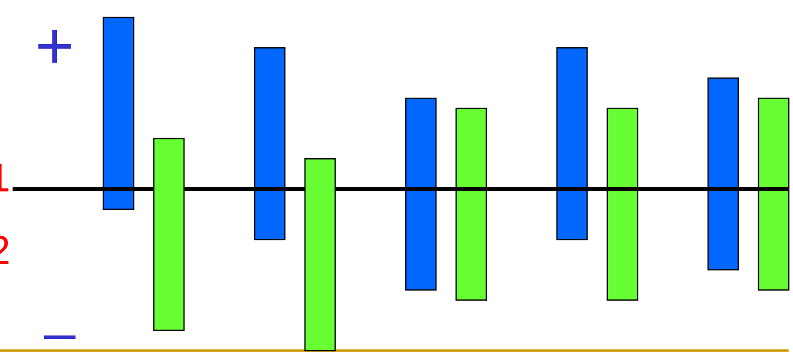
…

Visual Summarization & Comparison

- Summary of reviews of Digital camera 1



- Comparison of reviews of Digital camera 1 Digital camera 2



Roadmap

- Sentiment classification
- Feature-based opinion extraction
  - Problems
  - Some existing techniques
- Comparative sentence and relation extraction
  - Problems
  - Some existing techniques.

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> $\Rightarrow$ <size>

- Cons can be separated into 2 segments:

battery usage	<battery>
included memory is stingy	<memory>

Extraction using label sequential rules

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

$$\langle\{\text{easy, JJ}\}\{\text{to}\}\{*, \text{VB}\}\rangle \rightarrow \langle\{\text{easy, JJ}\}\{\text{to}\}\{\$\text{feature, VB}\}\rangle \quad [\text{sup} = 10\%, \text{conf} = 95\%]$$

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.
  - Sufficient for practical use: They are main features.
- Sequential pattern mining finds frequent phrases.
- We can further restrict their POS tags.
- Froogle implemented this approach (no POS restriction).

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.”

- Frequent features
- Infrequent features
- Opinion words

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 features

Using sentiment words and phrases (Hu and Liu 2004; Kim and Hovy 2004)

- Identify words that are often used to express positive or negative sentiments
- There are many ways.

- Use dominate orientation of opinion words as the sentence orientation, e.g.,
  - Sum: a negative word is near the feature, -1, a positive word is near a feature, +1

(Yu and Hatzivassiloglou, 2003; Popescu and Etzioni, 2005) used some different methods.

Text classification methods can be employed too.

Sample results

- Feature extraction: Recall ≈ 0.80%, precision ≈ 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 F-score.
- 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.

Roadmap

- Sentiment classification
- Feature-based opinion extraction
  - Problems
  - Some existing techniques
- Comparative sentence and relation extraction
  - Problems
  - Some existing techniques.

Extraction of Comparatives

(Jinal and Liu SIGIR-06, AAAI-06; Liu's Web mining book 2006)

- Two types of evaluation
  - Direct opinions: "I don't like this car"
  - Comparisons: "Car X is not as good as car Y"
- They use different language constructs.
- Comparative Sentence Mining
  - Identify comparative sentences, and
  - extract comparative relations from them.

Linguistic Perspective

- Comparative sentences use morphemes like
  - *more/most*, *-er/-est*, *less/least* and *as*.
- *than* and *as* are used to make a 'standard' against which an entity is compared.

Limitations

- Limited coverage
  - *Ex: "In market capital, Intel is way ahead of Amd*"
- Non-comparatives with comparative words
  - Ex1: "*In the context of speed, faster means better*"
  - *Ex2: "More men than James like scotch on the rocks*" (meaningless comparison)
- For human consumption; no computational methods

Comparative sentences

- An *Object* (or *entity)* is the name of a person, a product brand, a company, a location, etc, under comparison in a comparative sentence.
- A *feature* is a part or property (attribute) of the object/entity that is being compared.

Definition: A *comparative sentence* expresses a relation based on similarities, or differences of more than one objects/entities.

- It usually orders the objects involved.

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*: 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 F_1, Object B has feature F_2 (F_1 and F_2 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. Entities in entityS1 appear to the left of the relation word and entities in entityS2 appear to the right of the relation word.
  - *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 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 2: Extract comparative relations from the sentences.

Focus on gradable comparatives in this talk.

Roadmap

- Sentiment classification
- Feature-based opinion extraction
  - Problems
  - Some existing techniques
- Comparative sentence and relation extraction
  - Problems
  - Some existing techniques.

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, i.e., 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

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 sentences.
  - using **class sequential rules** (CSRs) generated from sentences in step1 as attributes, e.g.,

    $\langle\{1\}\{3\}\{7, 8\}\rangle \rightarrow \text{class}_i$ [sup = 2/5, conf = 3/4]

1. Sequence data preparation

- Use words within radius *r* of a keyword to form a sequence (words are replaced with POS tags)
- ….

2. CSR generation

- Use different minimum supports for different keywords
- 13 manual rules, which were hard to generate automatically.

3. Learning using a NB classifier

- Use CSRs and manual rules as attributes to build a final classifier.

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.

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.
    "*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*" has *$entityS1* "*Canon*" and *$feature* "*optics*".

Sequences are:

- ⟨{#start}{l1}{**$entityS1,** NNP}{*r*1}{has, VBZ }{r2 }{better, JJR}{*r*3}{$Feature, NNS}{r4}{#end}⟩
- ⟨{#start}{l4}{**$entityS1**, NNP}{l3}{has, VBZ}{*l*/2}{better, JJR}{*l*/1}{**$Feature**, NNS}{*r*1}{#end}⟩

Build a sequential cover from LSRs

LSR: ⟨{*, NN}{VBZ}⟩ → ⟨{$entityS1, NN}{VBZ}⟩

- 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

- NB using CSRs and manual rules as attribute
  precision = 82% and recall = 81%.
- NB using CSRs alone:
  precision = 76% and recall = 74%.
- SVM: precision = 71% and recall = 69%

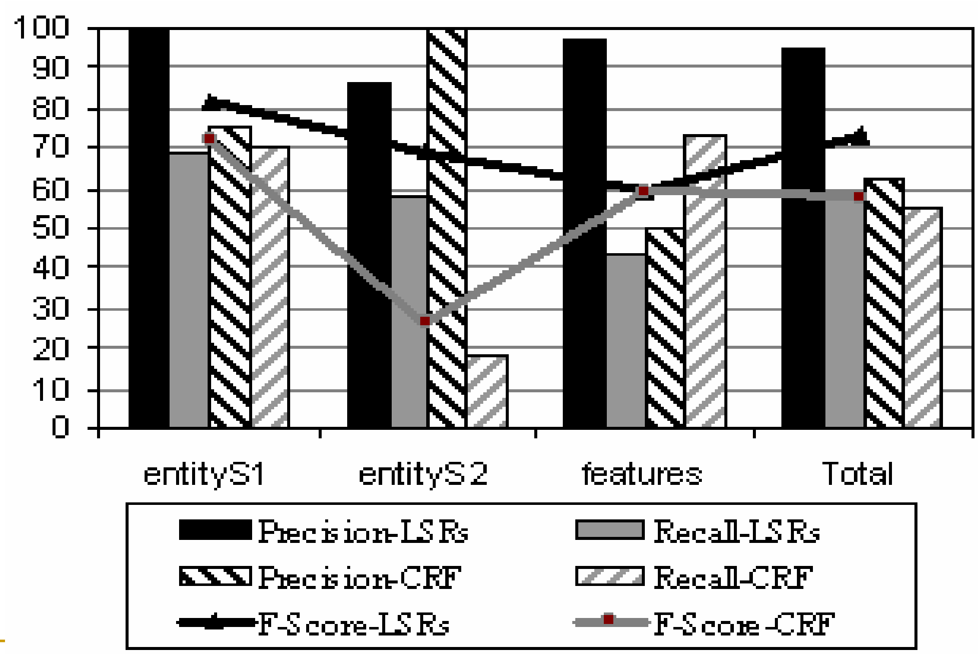
Classification into three different gradable types

- SVM gave accuracy of 96%
- NB gave accuracy of 87%

Extraction of comparative relations

- LSR gave F-score = 72%
- CRF gave F-score = 58%
- LSR extracted
  - 32% of complete relations
  - 32% relations where one item was not extracted
- Extracting relation words:
  - Non-Equal Gradable Precision = 97%. Recall = 88%
  - Equative: Precision = 93%. Recall = 91%
  - Superlative: Precision = 96%. Recall = 89%

LSR vs. CRF on relation item extraction



Conclusion

- Two types of evaluations are discussed
- **Direct opinions**: A lot of interesting work to do: Accuracy is the key:
  1. Feature extraction
  2. Opinion orientations on features
- **Comparison extraction**: a lot of work to do too,
  1. identify comparative sentences
  2. Group them into different types
  3. Extraction of relations
- Current techniques are preliminary, not deep or refined enough. More NLP is needed …
- Industrial applications are coming soon …