oiioiioii oxford
oiioiioii internet institute
oiioiioii university of oxford

ÅBO AKADEMI
Information Studies

Adapting systems for different domains and languages

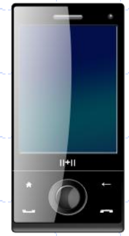
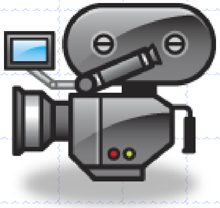
Mike Thelwall
University of Wolverhampton, UK

Statistical Cybermetrics
Research Group

CYBEREMOTIONS

UNIVERSITY OF WOLVERHAMPTON

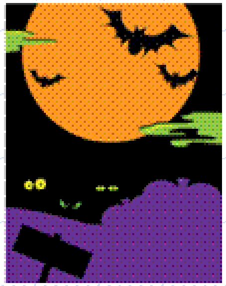
Contents


- The importance of domain in sentiment analysis
- Techniques for transferring from one domain to another
- Domain-independent methods
- Language issues and methods to translate a sentiment analysis method from one language to another

Are these positive or negative, and what are the sentiment terms?

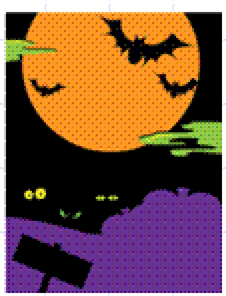
- The Maxtor hard disk was very large and could fit all of my data.
- The cellphone was very large and could not fit in my pocket.
- The horror movie was very frightening.
- The car brakes were very frightening.



Two issues

1. Non-sentiment words may express clearly positive or negative opinions in a particular context
   - E.g., large (memory, screen), heavy (computer, cellphone), long (travel time, film duration), reliable (brakes), unpredictable (movie plot)

2. Non-sentiment words (and even some sentiment words) vary in typical meaning according to context
   - Typically positive words in horror movie reviews
     - Frightening, scary, unexpected, entertaining, Depp(?), good, excellent, recommend

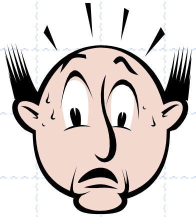


   - Typically positive words in family car reviews
     - Reliable, economical, roomy, large, strong, Honda(?), good, cheap

     - What could be typically negative words in these contexts?

Domain dependence

- The polarity and strength of words may vary by domain
- This is typically opinion or judgement words rather than direct sentiment words – e.g.,
  - “frightening” is typically positive for a horror movie review, but negative for a car review
  - “large” is typically positive for hard disk reviews, negative for cellphones



Domain context dependence

- The polarity and strength of words may also vary by context within a domain, e.g.,
  
  - "large" is typically positive for a cellphone screen but negative for cellphone size
  
  - "cheap" may be positive for a car overall but negative for the seats (or individual parts)

  - Any other examples??

Domain dependence – implications for sentiment analysis

- Generic sentiment algorithms will miss some domain-specific expressions
- Domain-specific algorithms can therefore be more accurate
- But domain-specific algorithms need domain-specific training data – often time consuming and difficult to produce

Cross-domain sentiment analysis

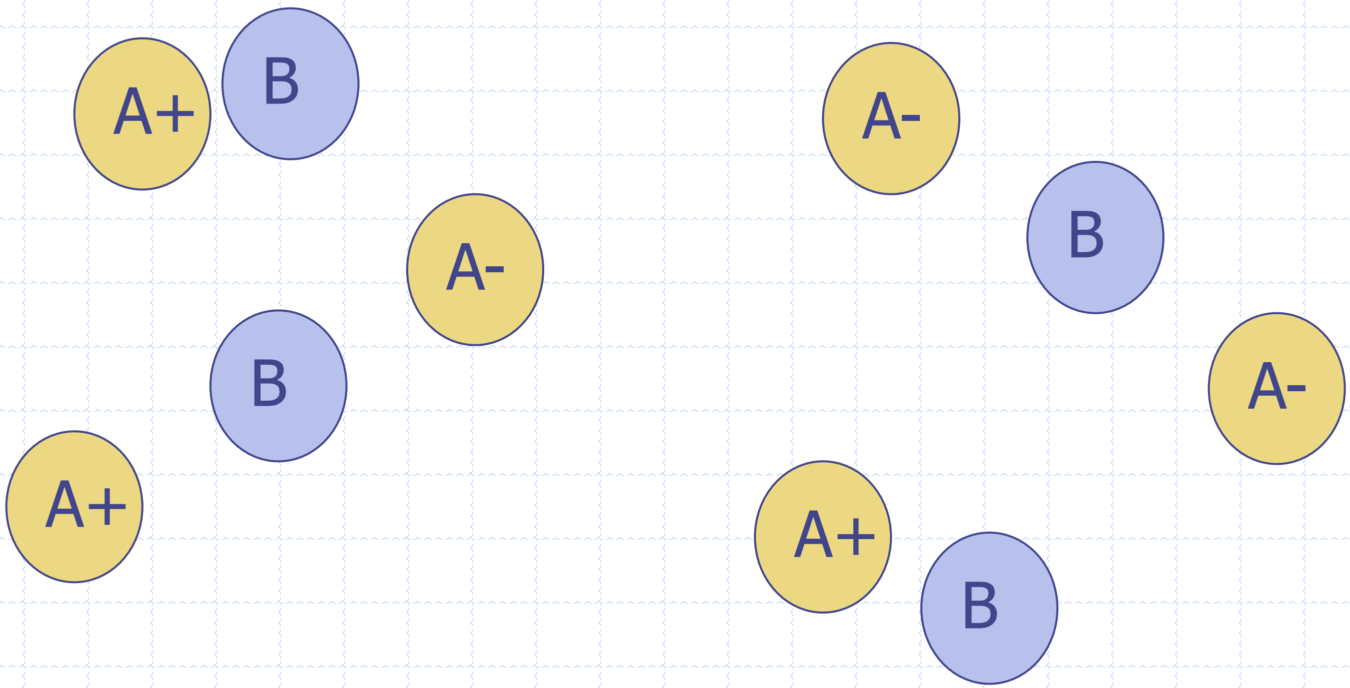
- Domain transfer is the problem of taking a sentiment classifier trained on one domain and adapting it to work on another domain
- Why not train a new classifier on the new domain?
  - Because of the problem creating training data
  - Especially if *many* new domains need sentiment classifiers

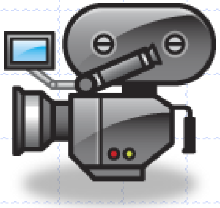
Cross domain A -> B: 5 methods

1. Train classifier on A, apply to B (no domain transfer)
2. Use ensembles of classifiers trained on different domain As
3. Train classifier on A, classify texts in B with the classification of texts in A that they are most similar to (e.g., TF-IDF)
4. Train classifier on A, annotate texts in B with the classification of texts in A that they are most similar to (e.g., TF-IDF) & retrain classifier on A & B
5. Train a classifier for B on the data from A but with only features found in B

How should the Bs be classified?



Summary



- Domain-specific sentiment analysis can be more accurate than general sentiment analysis
  - Especially for product reviews
- Domain-specific algorithms can be difficult to create due to the need for large gold standard text collections
- Cross-domain methods make it easier to generate domain-specific algorithms

Bibliography

- Aue, A., & Gamon, M. (2005). Customizing sentiment classifiers to new domains: A case study. *Proccedings of the International Conference on Recent Advances in Natural Language Processing*, http://research.microsoft.com/pubs/65430/new_domain_sentiment.pdf