OPINION MINING: A CASE STUDY FOR ADVANCED BOW & TRANSFER LEARNING

Vincent Guigue
vincent.guigue@agroparistech.fr

MIA PARIS-SACLAY EKINOCS

AGROPARISTECH Institut des Sciences et Industries du Vivant et de l'Environnement

Introduction

General position

Many keywords :

subjectivity, sentiment, opinion, emotion, affective state,...

2 main scientific communities :

- Affective Computing
- Sentiment Analysis

- Work on subjective problems
- Many industrial applications
  - Users do consider reviews !
  - We are able to analyse reviews...
  - Can we predict bahaviours ?

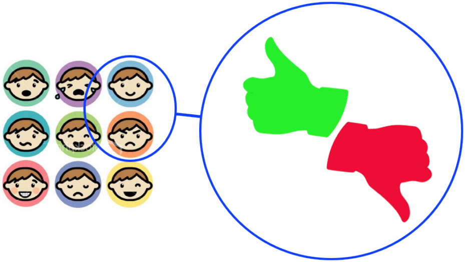
General position

Many keywords :

subjectivity, **sentiment**, **opinion**, emotion, affective state,...

2 main scientific communities :

- Affective Computing
- Sentiment Analysis



F. Zhu and X.M. Zhang Journal of Marketing, 2010 Impact of Online Consumer Reviews on Sales : The Moderating Role of Product and Consumer Characteristics

[D. Jurafsky] From emotion to sentiments

Scherer Typology of Affective States

- **Emotion** : brief, synchronized... evaluation of a major event
  - angry, sad, joyful, fearful, ashamed, proud, elated
- **Mood** : diffuse non-caused low-intensity long-duration change in subjective feeling
  - cheerful, gloomy, irritable, listless, depressed, buoyant
- **Interpersonal stances** : affective stance toward another person in a specific interaction
  - friendly, flirtatious, distant, cold, warm, supportive, contemptuous
- **Attitudes** : affectively colored beliefs, dispositions towards objects or persons
  - liking, loving, hating, valuing, desiring
- **Personality traits** : stable personality dispositions and typical behavior tendencies
  - nervous, anxious, reckless, morose, hostile, jealous

[D. Jurafsky] From emotion to sentiments

Scherer Typology of Affective States

- **Emotion** : brief, synchronized... evaluation of a major event
  - angry, sad, joyful, fearful, ashamed, proud, elated
- **Mood** : diffuse non-caused low-intensity long-duration change in subjective feeling
  - cheerful, gloomy, irritable, listless, depressed, buoyant
- **Interpersonal stances** : affective stance toward another person in a specific interaction
  - friendly, flirtatious, distant, cold, warm, supportive, contemptuous
- **Attitudes** : affectively colored beliefs, dispositions towards objects or persons
  - liking, loving, hating, valuing, desiring
- **Personality traits** : stable personality dispositions and typical behavior tendencies
  - nervous, anxious, reckless, morose, hostile, jealous

Sentiment Analysis = Opinion Mining

People

- B. Pang, L. Lee, Wiebe, Riloff, B. Liu...

Applications & conferences

- Survey, Reputation, IR
- WWW, EMNLP, ACL, ... (+IR / ML)

About mining user generated content on the web 2.0

Sentiment Analysis = Opinion Mining

People

- B. Pang, L. Lee, Wiebe, Riloff, B. Liu...

Applications & conferences

- Survey, Reputation, IR
- WWW, EMNLP, ACL, ... (+IR / ML)

Introducing users

Sentiment Analysis = Opinion Mining

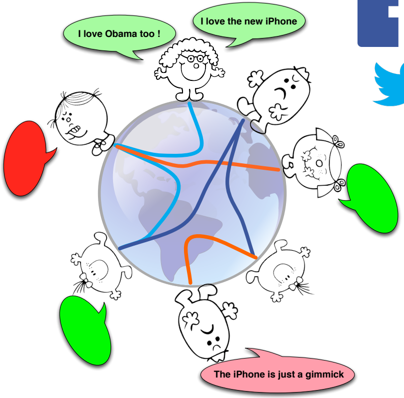
People

- B. Pang, L. Lee, Wiebe, Riloff, B. Liu...

Applications & conferences

- Survey, Reputation, IR
- WWW, EMNLP, ACL, ... (+IR / ML)

- Link with **Community Detection**
- Link with **Recommender Systems**



Introducing users & links

Three application clusters (at the document level)

1. E-reputation

   Companies looking for brand name protection

2. Survey

   Everybody trying to save money & conduct surveys on everything

3. (Opinionated) Information Retrieval

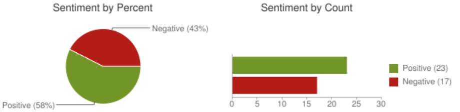
   Meeting consumer needs

⇒ Watching the whole web 24/7, detecting buzz, indexing...

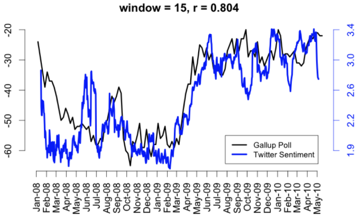
"*It takes 20 years to build a reputation and five minutes to ruin it.*"-
Warren Buffett

Survey & Monitoring : scientific issues

- Aggregating results from sentiment classification (new features)



- Building a time series prediction



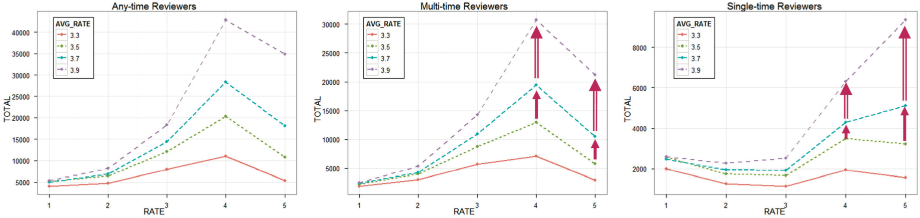
ICS (Index of Consumer Sentiment), Michigan monthly poll

O'Connor et al., ICWSM 2010
From Tweets to Polls : Linking Text Sentiment to Public Opinion Time Series.

IR : Adverse Search & Review Spam

Distinguishing real comments from review spams

- Natural distributions of marks (vs non-natural ones)



Feng et al., ICWSM 2012

- User/Author modeling
  - Special patterns
  - Social groups
  - Rating behaviors

"Positive opinions often mean profits and fame for businesses and individuals. This is a very strong incentive for people to game the system by posting fake opinions or reviews to promote or discredit certain products."- **Bing Liu**

Social Graph Mining

Combining influence modeling & Sentiment classification

⇒ who can give an efficient positive message about something ?

Issues : extracting

- Opinion Communities
- Opinion Influencers

Scientific issue :

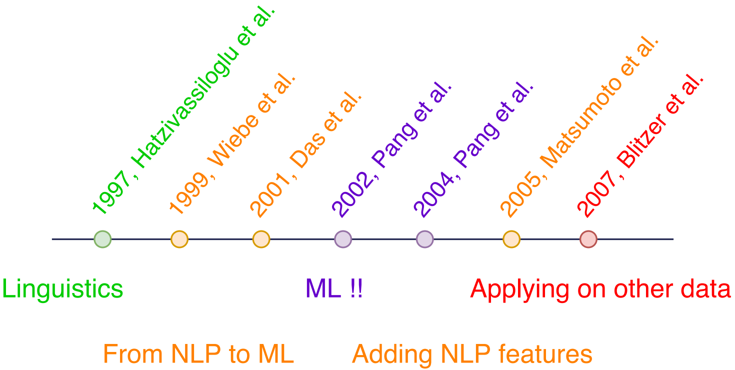
Dealing with a **lot of data**, taking into account both **graph structure** and **content** informations

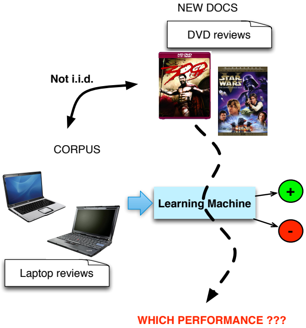
Advanced features & Bag of Words

When Machine learning & NLP merge

Web 2.0 : the users write some comments... But they also their ratings !

$\Rightarrow$ **Supervised Learning Techniques**





When Machine learning & NLP merge

Web 2.0 : the users write some comments... But they also their ratings !
⇒ **Supervised Learning Techniques**

1997 Hatzivassiloglou et al. : Resource building (adjectives)
1999 Wiebe et al. : Corpus building + Naive Bayes
2001 Das and Chen : BOW + NLP (Negation Coding) + ML : **It works !**
2002 Pang et al. : Simple Features (BOW) + SVM (or *Naive Bayes*)
It works very well !
2004 Pang et al. : ... but not at the sentence level
2005 Matsumoto et al. (and others) : complex NLP features are required
2007 Blitzer et al. : we want to apply those models on new (non rated) data

Evolution of the sentiment corpora

2002 Pang and Lee, Movie reviews : 1400 documents

2007 Blitzer et al. Amazon : 4x 2000 documents

2009 Whitehead : 10x [200, 2000] documents

... more and more corpora, with more and more documents

- Big Amazon : 20x [5k, 50k] documents
- Big IMDB : 2x 25k documents
- Trip advisor : 2x 50k documents
- Huge Amazon : 5.8 M reviews
- ...

⇒ **A very unusual task were the labeled data is almost infinite !**

Bag of words representation & coding

Most of the models rely on **Bag Of Words (BOW)** representation :

$$\begin{array}{ccc} \mathcal{D} & \rightarrow & \mathcal{X} \\ repr(doc) & = & d \in \mathbb{R}^V \end{array}$$

- Unigrams, Bigrams
- Additional features from NLP (POS...)

Since early studies [Das 2001, Pang 2002], a consensus on the coding :

- tf *vs* tf-idf *vs* presence coding
- Discriminant coding : Δ-tf : difference between frequencies in classes $+$ and $-$ ($\approx$ kind of NB in the coding)

$\Rightarrow$ **Presence Coding (= working in $\mathbb{B}^V$)**

Negation Coding, POS filtering

NLP vs ML

How to code the structure of the sentences ?
(Which is known to be important in the decision !)

- **N-grams**... But very high dimensionality above $N = 2$
- **POS filtering** : Reducing the complexity of the structure [Pang 2002]
- **Unigrammes + negation coding** [Das et al. 2001]
  Add `NOT_` to every word between negation and following punctuation :

```
[...] didn't like this movie , but I [...]
⇒ didn't NOT_like NOT_this NOT_movie but I
```

Sanjiv Das and Mike Chen. APFA, 2001
Yahoo ! for Amazon : Extracting market sentiment from stock message boards

Complex NLP features

More and more complex feature to catch the structure...

- **N-grams, subsequences, POS filtering** [Dave et al. 2003]

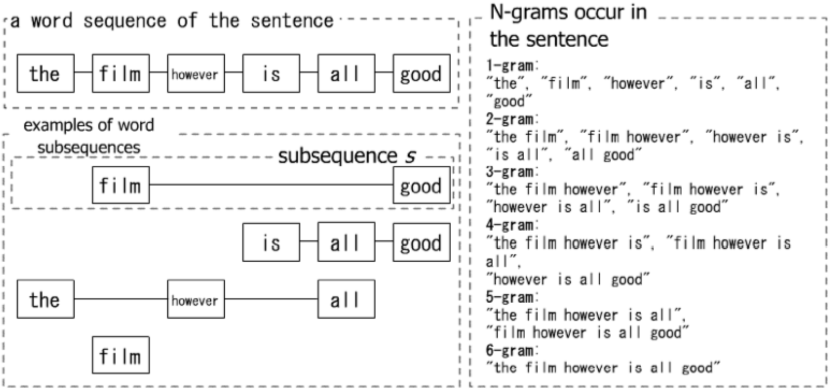


Fig. 2. A word sequence of a sentence "*The film however is all good*" and exampl
subsequences

$\Rightarrow$ filtering the noise & extracting frequent patterns

Complex NLP features

More and more complex feature to catch the structure...

- **N-grams, subsequences, POS filtering** [Dave et al. 2003]
- **Subsequences, tree-dependency** [Matsumoto et al. 2005]

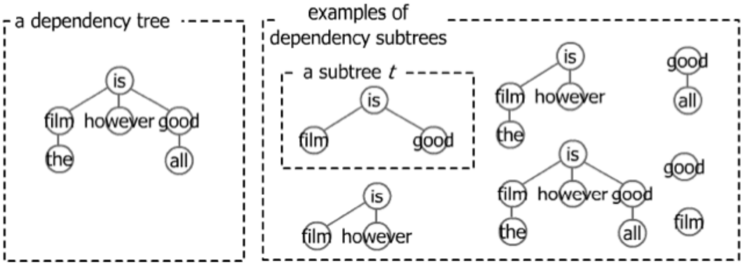


Fig. 3. A dependency tree of a sentence "*The film however is all good*" and exa
of subtrees

Representation & preprocessing

2008 Pang & Lee Survey, Lemmatization : open issue, depending on

- experiments, langage,
- size of the learning set

2006 Kennedy and Inkpen, Emotional Quantifiers :

- modifying the counting of terms depending on *valence shifters*

2010 Wiegand et al.,

- A Survey on the Role of Negation in Sentiment Analysis

2009 Potts : coding emoticons

```
[<>]?                         # optional hat/brow
[:;=8]                        # eyes
[\-o\*\']?                    # optional nose
[\)\]\(\[dDpP/\:\}\{@\|\\]    # mouth
|                             #### reverse orientation
[\)\]\(\[dDpP/\:\}\{@\|\\]    # mouth
[\-o\*\']?                    # optional nose
[:;=8]                        # eyes
[<>]?                         # optional hat/brow
```

High dimensionality representation

Raw extraction (no POS filtering)
Removing words with only one occurrence

Corpus	Corpus Size (N)	Review length	Vocabulary (V)		
			Unigr.	U+Bigr.	U+B+SSeq.
Books	2000	240	10536	45750	78664
Dvd	2000	235	10392	48955	89313
Electronics	2000	154	5611	30101	49994
Kitchen	2000	133	5314	26156	40773
Movie Rev.	2000	745	26420	148765	308564

Table 1 – Descriptions of 5 famous corpora (IMDB, Amazon).

Dimensionality increases quickly !

High dimensionality ⇒ many problems

- Cost of the dictionary (memory and computation time)
  - POS filtering & a priori knowledge...
- Optimization problem
  - Curse of dimensionality
  - Ill posed problems :
    - 10^3 docs *VS* 10^4 or 10^6 features
  - Risk of over fitting
- Engineering problems
  - Growing datasets ⇒ harder developments ⇒ bug/bias risk

Results [Pang et al. 2002]

- Very close performances for all algorithms
- [Pang et al. 2002] : SVM slightly better, Naive Bayes slightly worse.

	Features	# of features	frequency or presence?	NB	ME	SVM
(1)	unigrams	16165	freq.	**78.7**	N/A	72.8
(2)	unigrams	"	pres.	81.0	80.4	**82.9**
(3)	unigrams+bigrams	32330	pres.	80.6	80.8	**82.7**
(4)	bigrams	16165	pres.	77.3	**77.4**	77.1
(5)	unigrams+POS	16695	pres.	81.5	80.4	**81.9**
(6)	adjectives	2633	pres.	77.0	**77.7**	75.1
(7)	top 2633 unigrams	2633	pres.	80.3	81.0	**81.4**
(8)	unigrams+position	22430	pres.	81.0	80.1	**81.6**

Figure 3: Average three-fold cross-validation accuracies, in percent. Boldface: best performance for a gi
setting (row). Recall that our baseline results ranged from 50% to 69%.

Results [Matsumoto et al. 2005]

- Rich coding $\Rightarrow$ $\nearrow$ polarity recognition

Table 2. Results for dataset 1

Features	Acurracy(%)		
	test1	test2	test3
Pang et al. [1]	82.9	N/A	N/A
Mullen et al. [10]	**84.6**	N/A	N/A
word unigram (= *uni*)	83.0	83.7	83.0
lemma unigram (= uni_l)	82.8	83.8	83.2
word bigram (= *bi*)	79.6	80.4	80.1
lemma bigram (= bi_l)	80.4	80.9	80.7
$uni + bi$	83.8	84.6	84.0
$uni + bi_l$	83.6	84.2	83.5
$uni_l + bi$	**84.4**	84.8	**84.6**
$uni_l + bi_l$ (= *bow*)	84.0	**84.9**	84.2
$bow + seq$	84.1	85.3	84.9
$bow + seq_l$	84.4	85.7	84.9
$bow + dep$	86.6	87.6	87.5
$bow + dep_l$	**87.3**	**88.3**	**88.0**
$bow + seq + dep$	86.2	87.2	87.2

Table 3. Results for dataset 2

Features	Acurracy(%)		
	test1	test2	test3
Pang et al. [2]	**87.1**	N/A	N/A
word unigram (= *uni*)	87.1	88.1	87.0
lemma unigram (= uni_l)	86.4	86.9	85.9
word bigram (= *bi*)	84.2	85.3	85.1
lemma bigram (= bi_l)	84.3	85.2	84.7
$uni + bi$ (= *bow*)	**88.1**	**88.8**	**88.0**
$uni + bi_l$	87.8	88.6	87.8
$uni_l + bi$	87.3	88.2	87.3
$uni_l + bi_l$	87.7	88.3	87.9
$bow + seq$	88.2	89.4	88.3
$bow + seq_l$	88.5	89.8	88.5
$bow + dep$	92.4	**93.7**	92.7
$bow + dep_l$	92.8	**93.7**	92.9
$bow + seq + dep$	92.6	93.5	92.8

Linear models & decision making

In detail example with bigrams dictionary (Amazon Book Reviews) :

Given a trained model w, For a document x_i , decision is computed as follows :
$f(\mathrm{x}_i) = \langle \mathrm{x}_i, \mathrm{w} \rangle = \sum_V x_{ij} w_j$

With presence coding (namely $x_{ij} = 0/1$) : $f(\mathrm{x}_i) = \sum_{j \in \mathrm{x}_i} w_j$

- w_j is the contribution associated to the presence of the word j
  - detection of positive & negative markers
- $|w_j|$ can be seen as a subjectivity score ($\approx$ Sentiwordnet definition)

Linear models & decision making

In detail example with bigrams dictionary (Amazon Book Reviews) :

Linear models & presence coding : $f(x_i) = \sum_{j \in x_i} w_j$

```
This book is great
Score : 0.20597761553808144
   this_book : -0,0569
   be : 0,0760
   book_be : 0,0195
   great : 0,2507
   book : -0,0183
   be_great : 0,0057
   this : -0,0707
```

```
I felt disappointed
Score :-0.07434288371014953
   I : 0,0189
   disappointed : -0,0946
   I_feel : 0,0062
   feel : -0,0048
```

Linear models & decision making

In detail example with bigrams dictionary (Amazon Book Reviews) :

```
This book is not easy to read
Score : -0.029908342264773540
  this_book : -0,0569
  read : 0,0866
  book_be : 0,0195
  easy : 0,1688
  this : -0,0707
  be_not : -0,0661
  not : -0,3336
  be : 0,0760
  not_easy : 0,0348
  easy_to : 0,1046
  book : -0,0183
  to : 0,1163
  to_read : -0,0909
```

Negation examples :

```
could couldn't should shouldn't
  should : -0,0245
  could : -0,0272
  shouldn't : 0,0026
  couldn't : 0,0293
```

Linear models & decision making

In detail example with bigrams dictionary (Amazon Book Reviews) :

Inefficient at sentence level :

```
I don't like this book
Score : -0.38660630368246496
   this_book : -0,0569
   I_don't : -0,1249
   don't : -0,1555
   I : 0,0189
   don't_like : 0,0391
   book : -0,0183
   like : -0,0706
   this : -0,0707
   like_this : 0,0522
```

```
I like this book
Score : -0.06739329361414403
   this_book : -0,0569
   I : 0,0189
   book : -0,0183
   like : -0,0706
   this : -0,0707
   like_this : 0,0522
   I_like : 0,0780
```

Decision at the sentence level

Pang & Lee 2004 : same idea as Wilson et al. 05 :

Focusing on opinionated sentences to remove some noise and make an efficient decision about the polarity

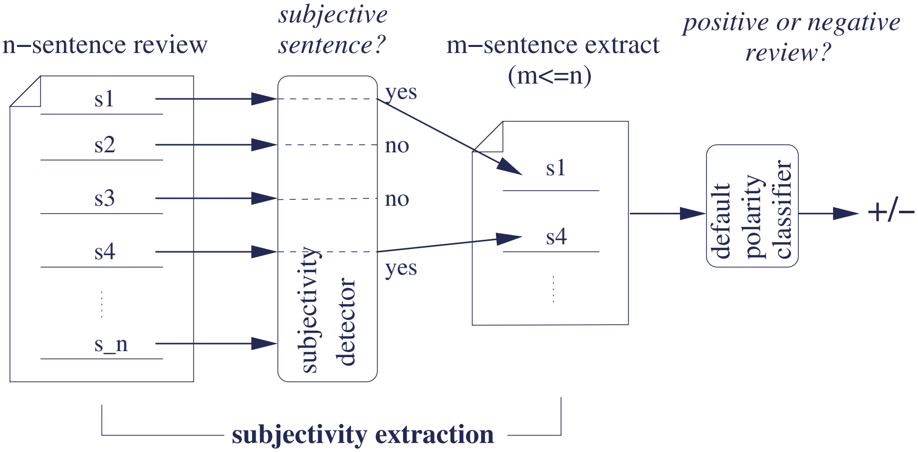


Figure 1: Polarity classification via subjectivity detec

Decision at the sentence level

Pang & Lee 2004 : same idea as Wilson et al. 05 :

Focusing on opinionated sentences to remove some noise and make an efficient decision about the polarity

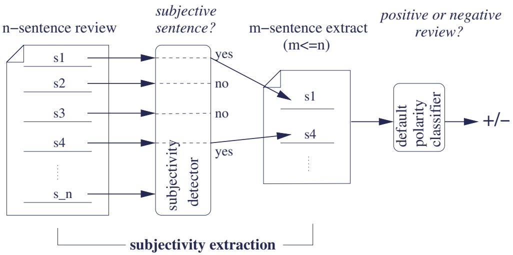


Figure 1: Polarity classification via subjectivity detection.

⇒ **Very hard to obtain satisfying results !**

Bo Pang and Lillian Lee ACL 2004 A Sentimental Education : Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts

Conclusion

+ Impressive performances on specific datasets
+ (Almost) no NLP knowledge required
+ Comprehensive model (at the corpus level)
– ... But mixed performance at the sentence level
+ Very efficient approach with existing SVM libraries & basic coding
– ... But computation time increases quickly when descriptors become complex

⇒ let's move to multi-domain sentiment classification

+ Good for the performance on a single dataset...
– Bad for the generalization abilities of the model !

Machine Learning, Multi-Domaines (ML-MD)

(Very) Brief History of Multi-domain sentiment classification

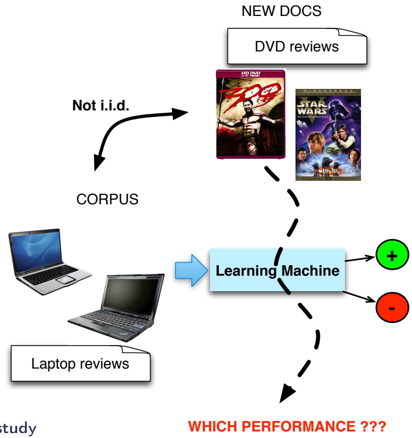
Aue 05 Description of the task

Blitzer 07 SCL algorithm & Amazon dataset release

>07 Many proposals

Vocabulary :

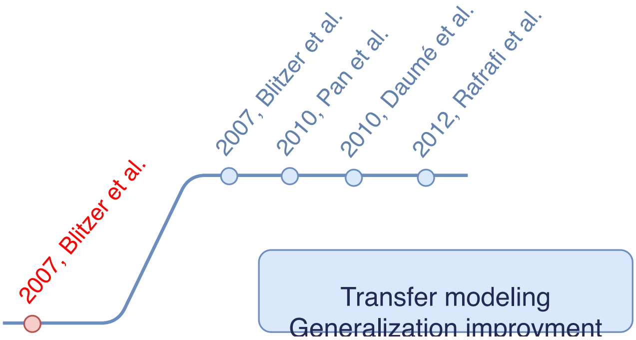
- **Source :** the labeled data used to train our model
- **Target :** the data we want to classify

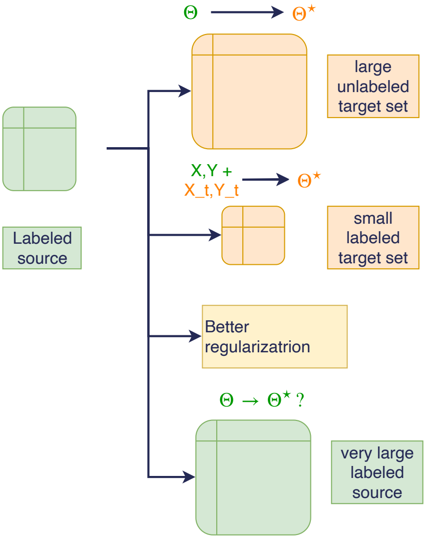


Aue, Gamon RANLP 2005 Customizing sentiment classifiers to new domains : A case study

Blitzer, Dredze, Pereira ACL 2007 Biographies, Bollywood, Boom-boxes and Blenders : Domain Adaptation for Sentiment Classification

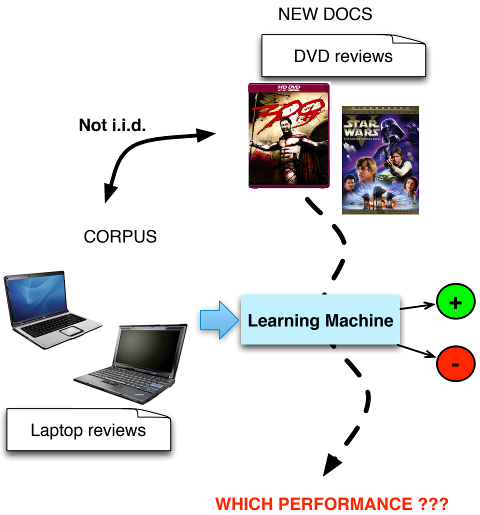
Algorithmic approaches





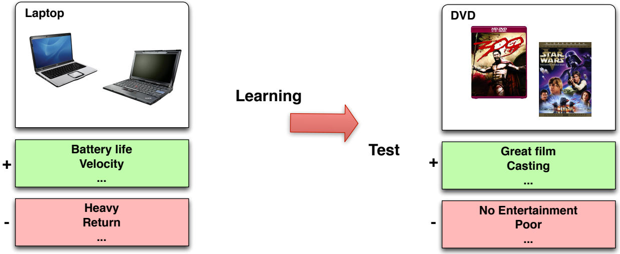
Blitzer Framework

- 4 domains from Amazon :
  - Books (2k docs),
  - Dvd (2k docs),
  - Kitchen (2k docs),
  - Electronics (2k docs)
  - >5k unlabeled docs in each domain
- Cross-domain experiments
  - Learning on one domain
  - Testing on another one
- Binary Classification (+/−)



Blitzer, Dredze, Pereira ACL 2007 Biographies, Bollywood, Boom-boxes and Blenders : Domain Adaptation for Sentiment Classification

Why is it difficult ?



- Vocabulary transfer
- Different context

$\Rightarrow$ Even with unigrams, linear model learn domain specific rules

- *summer* : good for books (summer book)
- *return* : very bad for all physical product (return to the seller)
- *loved* : good for a book, bad for a phone...

Many proposals

1. Engineering approach
   Adapting the learnt model to the test distribution
   - When we get few test labels
   - Without test labels
2. Theoretical optimization approach
   better generalization $\approx$ regularization process
   - Finding an efficient regularization for sentiment classification
3. Linguistic approach
   Better understanding of the messages $\Rightarrow$ Better generalization
   - Introducing a (probabilistic) semantic

Structural Correspondance Learning

Input: labeled source data $\{(\mathbf{x}_t, y_t)_{t=1}^T\}$,
unlabeled data from both domains $\{\mathbf{x}_j\}$

Output: predictor $f : X \to Y$

1. Choose m pivot features. Create m binary prediction problems, $p_\ell(\mathbf{x})$, $\ell = 1 \dots m$

2. For $\ell = 1$ to m

$$\hat{\mathbf{w}}_\ell = \underset{\mathbf{w}}{\operatorname{argmin}} \left(\sum_j L(\mathbf{w} \cdot \mathbf{x}_j, p_\ell(\mathbf{x}_j)) + \lambda ||\mathbf{w}||^2 \right)$$

end

3. $W = [\hat{\mathbf{w}}_1 | \dots | \hat{\mathbf{w}}_m]$, $[U\ D\ V^T] = \mathrm{SVD}(W)$, $\theta = U^T_{[1:h,:]}$

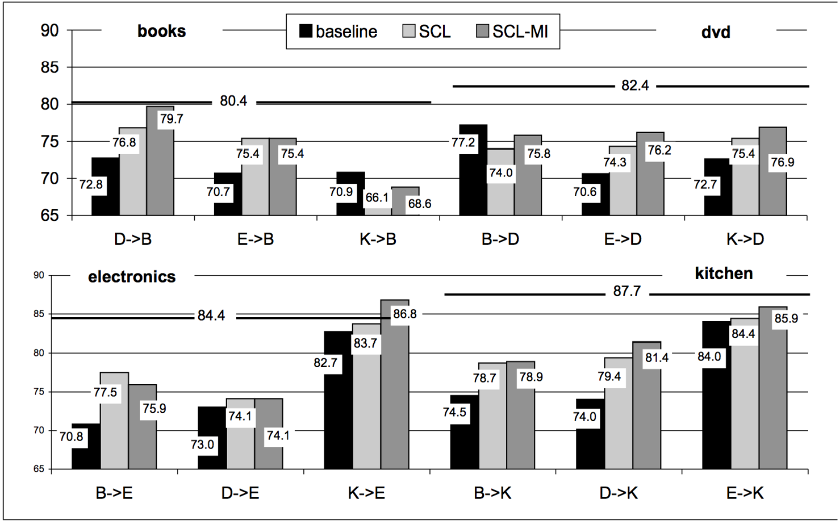
4. Return f, a predictor trained on $\left\{ \left(\begin{bmatrix} \mathbf{x}_t \\ \theta \mathbf{x}_i \end{bmatrix}, y_t \right)_{t=1}^T \right\}$

Figure 3: SCL Algorithm

- Labeled Source , Unlabeled Target data
- Pivot feature p = frequent sentiment word
- W = which words cooccure with p ?
- SVD on W = concept extraction
- Enriched description : Original + Concepts

Blitzer, Dredze, Pereira ACL 2007 Biographies, Bollywood, Boom-boxes and Blenders : Domain Adaptation for Sentiment Classification

Algo. [Blitzer et al. 2007], [Pan et al., 2010]



Simpler Algorithm

Hypothesis : we have a small labeled target dataset

- The target dataset is too small (*vs* the source) to impact the final model
- Proposal : high dimensionality coding

$$\Phi^s(x) = [x, x, 0], \quad \Phi^t(x) = [x, 0, x], \quad f(x) = \langle \Phi(x), [w^{gen}, w^s, w^t] \rangle$$

NB : implementation = copying the doc. representation

- regularization is required : either on $\|w\|^2$, or on $\|w^s - w^t\|^2$

$\Longrightarrow$ Very good performances... On non-sentiment applications
$\Longrightarrow$ Interesting performances for sentiments

NB : a semi-supervised version of the algo. called EasyAdapt++ has been developed afterwards.

Daumé, ACL 2007
Frustratingly Easy Domain Adaptation

CODA [Chen 2011]

Idea

Using a kind of *pseudo relevance feedback* to retrain our model and fit the target distribution

1. Training a model M on the source $\mathcal{S}$
2. Applying M on the target $\mathcal{T}$
3. Selecting documents x_c with the most confident decision in $\mathcal{T}$
4. Retrain M^{t+1} on $\mathcal{S}^{t+1} = \{\mathcal{S}, \mathrm{x}_c\}$
5. Repeat until $\mathcal{T}$ is empty

- Very good performance
- Heavy computation cost

M. Chen, K. Q. Weinberger and J. C. Blitzer, NIPS 2011
Co-Training for Domain Adaptation

Multiple sources [Whitehead, 2009]

Idea : using many sources to optimize the target performance

Table 1: Cross Domain Classification

Testing Dataset \ Training Dataset	camera	camp	doctor	drug	laptop	lawyer	movie	music	radio	rest	tv	Ave. Test
camera	**90**	64	67	57	64	63	51	55	60	61	62	63
camp	71	**85**	69	57	53	68	52	63	62	66	71	65
doctor	59	65	**84**	58	53	72	50	57	63	72	65	63
drug	57	59	59	**72**	53	62	50	54	57	59	56	58
laptop	74	56	56	53	**96**	63	57	50	55	61	52	61
lawyer	57	61	71	55	59	**83**	51	60	60	66	65	63
movie	57	63	56	52	59	59	**81**	55	53	58	59	59
music	58	58	65	61	50	53	50	**88**	61	63	56	60
radio	55	64	62	53	52	62	50	58	**73**	59	58	59
rest	64	67	76	64	53	62	56	64	64	**85**	65	65
tv	61	70	63	53	51	71	52	58	62	63	**82**	62
Ave. Train	64	65	66	58	58	65	55	60	61	65	63	

Whitehead, Yaeger IEEE 2009 Building a General Purpose Cross-Domain Sentiment Mining Model

Multiple sources [Whitehead, 2009]

Idea : using many sources to optimize the target performance

Table 3: Leave-One-Out vs. Single-Domain

Test Set	LOO	Ave. Other	Best Other	Same
camera	62	63	67	90
camp	66	65	71	85
doctor	61	64	72	84
drug	53	58	62	72
laptop	65	61	74	96
lawyer	58	63	71	83
movie	62	59	63	81
music	61	60	65	88
radio	56	59	64	73
restaurant	67	65	76	85
tv	60	62	71	82
Average	61	62	69	83

$\Rightarrow$ But it doesn't work !

Whitehead, Yaeger IEEE 2009 Building a General Purpose Cross-Domain Sentiment Mining Model

Multiple sources : other studies

Theoretical study [Dredze 2010]

Using a mixture of sources to model the target **should** improve the performance on the target

Experimental Results (Whitehead, 09) :

Mixture of sources $\approx$

Choosing best source (with an oracle) $>$

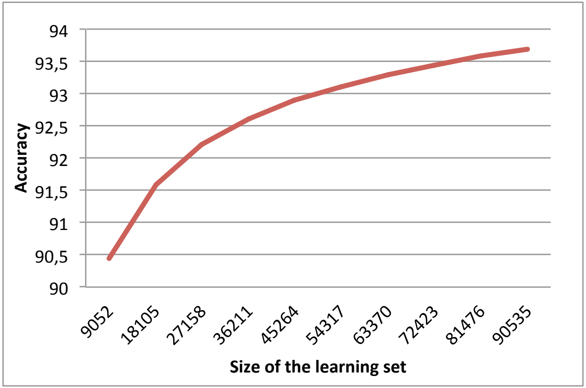
blending all sources

Mark Dredze, Alex Kulesza and Koby Crammer, Machine Learning 2010
Multi-domain learning by confidence-weighted parameter combination

Multiple sources & scaling [Rafrafi 2013]

(Very) empirical study :
what append when we enlarge the learning set ?

Classical cross-validation on blended Amazon :

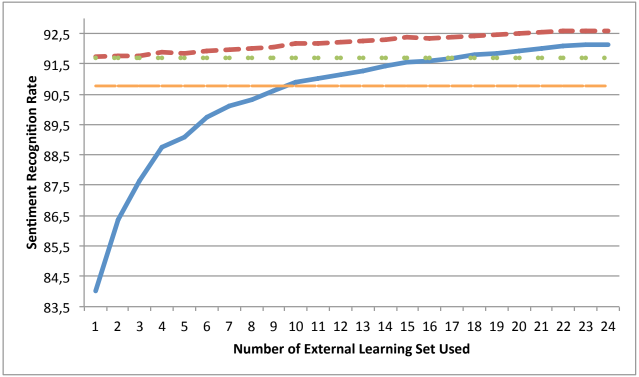


Rafrafi, Guigue, Gallinari CORIA, 2013 Multi-domain Sentiment Classification & Scaling

Multiple sources & scaling [Rafrafi 2013]

(Very) empirical study :
what append when we enlarge the learning set ?

In multi-domain context (25 Amazon's domains)



In-domain baseline
Best source (oracle)
Labeled target + all sources
All sources without any target data

Rafrafi, Guigue, Gallinari CORIA, 2013 Multi-domain Sentiment Classification & Scaling

Multiple sources & scaling [Rafrafi 2013]

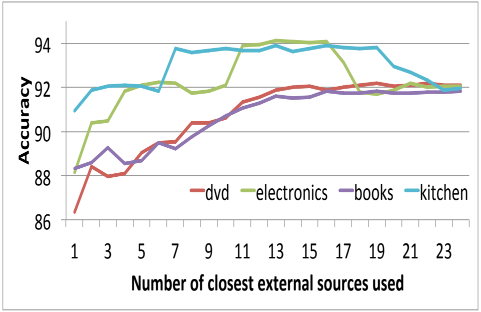
(Very) empirical study :
what append when we enlarge the learning set ?

Distance KL between the sources ⇒ multiple sources selection

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	0	8,55	10,05	8,35	9,82	10,08	11,24	12,39	11,27	13,3	13,12	14,84	16,42	16,53	18,15	15,08	16,88	19,06	20,87	19,63	19,26	20,87	19,09	18,68	33,49
2	8,55	0	13,35	7,83	11,07	11,2	12,31	13,96	12,84	14,88	13,62	15,57	10,11	16,85	19,49	15,73	18,88	20,24	21,64	20,31	19,93	21,57	16,6	15,6	33,63
3	10,05	13,35	0	12,1	13,98	10,5	15,74	16,74	15,18	16,72	18,45	19,06	19,6	19,89	19,38	17,17	14,37	16,83	18,44	18,16	19,7	19,13	20,6	20,66	37,69
4	8,35	7,83	12,1	0	11,04	10,96	9,86	14,01	12,78	14,29	13,8	15,85	15,32	16,86	19,57	16,41	19,44	20,82	22,31	20,99	20,31	22,36	15,78	15,4	33,58
5	9,82	11,07	13,98	11,04	0	14,45	13,02	9,22	14,86	15,76	14,22	10,16	19,37	17,78	16,95	11,24	16,29	20,01	21,9	20,78	20,51	20,42	21,61	21,35	35,25
6	10,08	11,2	10,5	10,96	14,45	0	14,55	17,28	14,17	17,93	16,63	18,55	18,61	20,23	22,19	19,84	21,3	22,25	23,72	22,9	23,25	23,85	21,74	20,9	36,63
7	11,24	12,31	15,74	9,86	13,02	14,55	0	16,27	16,17	16,48	15,65	17,74	19,94	19,99	22,2	18,18	21,85	23,37	24,85	23,3	22,72	25,04	20,95	21,05	35,38
8	12,39	13,96	16,74	14,01	9,22	17,28	16,27	0	16,96	17,66	17,46	14,34	21,43	20,57	20,88	15,47	19,02	21,46	23,03	22,5	21,96	23,87	23,68	23,47	37,14
9	11,27	12,84	15,18	12,78	14,86	14,17	16,17	16,96	0	13,72	18,16	18,09	17,44	19,18	21,32	20,74	22,25	22,73	23,98	23,67	22,32	24,34	20,7	20,43	36,8
10	13,3	14,88	16,72	14,29	15,76	17,93	16,48	17,66	13,72	0	19,48	19,04	20,43	21,49	22,29	21,27	23,46	23,71	24,9	24,79	23,78	25,28	22,9	22,88	37,74
11	13,12	13,62	18,45	13,8	14,22	16,63	15,65	17,46	18,16	19,48	0	18,18	20,18	21,67	23,41	19,07	22,89	25,32	26,69	25,43	25,02	26,47	24	23,8	34,92
12	14,84	15,57	19,06	15,85	10,16	18,55	17,74	14,34	18,09	19,04	18,18	0	22,64	21,88	22,19	16,38	21,41	25,06	26,8	25,62	24,46	25,64	24,9	25,16	38,34
13	16,42	10,11	19,6	15,32	19,37	18,61	19,94	21,43	17,44	20,43	20,18	22,64	0	22,46	25	21,89	23,07	23,79	24,74	24,04	22,88	24,02	18,39	18,76	38,03
14	16,53	16,85	19,89	16,86	17,78	20,23	19,99	20,57	19,18	21,49	21,67	21,88	22,46	0	24,99	21,84	24,96	25,79	27,08	24,24	24,4	26,7	24,29	24,25	38,71
15	18,15	19,49	19,38	19,57	16,95	22,19	22,2	20,88	21,32	22,29	23,41	22,19	25	24,99	0	22,79	24,01	26,07	27,89	26,66	26,89	24,34	26,51	26,3	39,88
16	15,08	15,73	17,17	16,41	11,24	19,84	18,18	15,47	20,74	21,27	19,07	16,38	21,89	21,84	22,79	0	16,09	20,47	22,13	18,6	19,11	21,65	23,72	23,56	38,49
17	16,88	18,88	14,37	19,44	16,29	21,3	21,85	19,02	22,25	23,46	22,89	21,41	23,07	24,96	24,01	16,09	0	14,66	16,13	17,26	19,78	17,46	23,81	24,44	40,67
18	19,06	20,24	16,83	20,82	20,01	22,25	23,37	21,46	22,73	23,71	25,32	25,06	23,79	25,79	26,07	20,47	14,66	0	1,83	10,6	17,34	11,36	23,36	24,64	42,3
19	20,87	21,64	18,44	22,31	21,9	23,72	24,85	23,03	23,98	24,9	26,69	26,8	24,74	27,08	27,89	22,13	16,13	1,83	0	10,77	18,53	12,98	24,48	25,89	43,25
20	19,63	20,31	18,16	20,99	20,78	22,9	23,3	22,5	23,67	24,79	25,43	25,62	24,04	24,24	26,66	18,6	17,26	10,6	10,77	0	13,05	16,1	23,62	24,99	41,89
21	19,26	19,93	19,7	20,31	20,51	23,25	22,72	21,96	22,32	23,78	25,02	24,46	22,88	24,4	26,89	19,11	19,78	17,34	18,53	13,05	0	19,43	22,66	22,82	41,46
22	20,87	21,57	19,13	22,36	20,42	23,85	25,04	23,87	24,34	25,28	26,47	25,64	24,02	26,7	24,34	21,65	17,46	11,36	12,98	16,1	19,43	0	24,36	25,18	42,69
23	19,09	16,6	20,6	15,78	21,61	21,74	20,95	23,68	20,7	22,9	24	24,9	18,39	24,29	26,51	23,72	23,81	23,36	24,48	23,62	22,66	24,36	0	10,15	40,28
24	18,68	15,6	20,66	15,4	21,35	20,9	21,05	23,47	20,43	22,88	23,8	25,16	18,76	24,25	26,3	23,56	24,44	24,64	25,89	24,99	22,82	25,18	10,15	0	39,5
25	33,49	33,63	37,69	33,58	35,25	36,63	35,38	37,14	36,8	37,74	34,92	38,34	38,03	38,71	39,88	38,49	40,67	42,3	43,25	41,89	41,46	42,69	40,28	39,5	0

1 Sports
2 Health
3 Games
4 Kitchen
5 Electronics
6 Baby
7 Outdoor
8 Camera
9 Apparel
10 Jewelry
11 Automotive
12 Cell
13 Beauty
14 Office
15 Instruments
16 Software
17 Videogames
18 Dvd
19 Video
20 Books
21 Magazines
22 Music
23 Food
24 Grocery
25 Tools

Performance using sources that are close to the target :



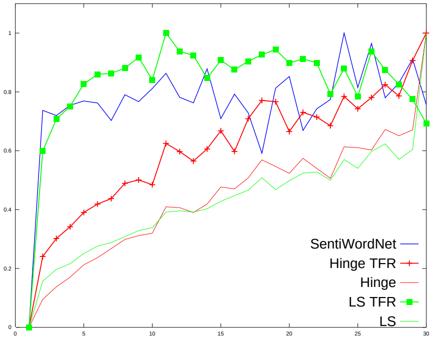
Rafrafi, Guigue, Gallinari CORIA, 2013 Multi-domain Sentiment Classification & Scaling

Subjectivity Appraisal [Rafrafi 2012]

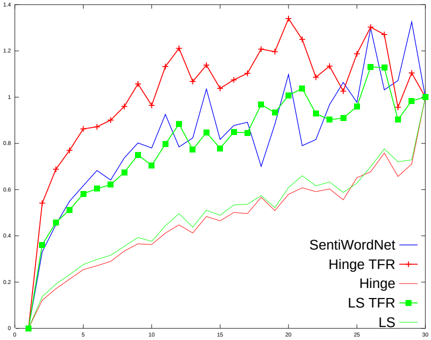
Scarce terms have more influence :

new subjectivity curves are closer to SentiWordNet

Unigrams :



Bigrams :

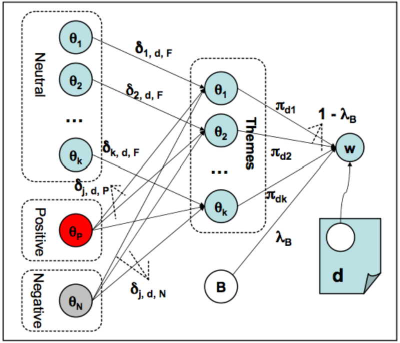


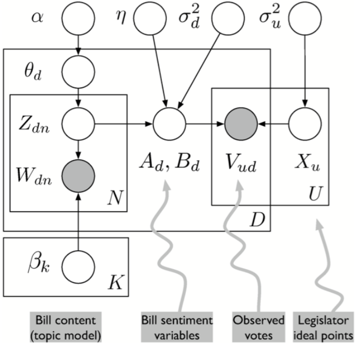
⇒ New opportunities for a better regularization

Semantics & graphical models

LSA for topic/sentiments

- Learning with EM (simple models)
- Learning with Gibbs Sampling (more complex model)





Liu, Huang, An, Yu SIGIR 2007 ARSA : a sentiment-aware model for predicting sales performance using blogs

Mei, Ling, Wondra, Su, Zhai WWW 2007 Topic Sentiment Mixture : Modeling Facets and Opinions in Weblogs

Gerrish, Blei ICML 2011 Predicting Legislative Roll Calls from Text

Multi-Aspect Review

- Using generative model (multi-grain LDA)
- + Dealing with multi-aspect reviews
- Summary ⇒ Extracting some group of words

Food: 5; Decor: 5; Service: 5; Value: 5
The chicken was great. On top of that our service was excellent and the price was right. Can't wait to go back!
Food: 2; Decor: 1; Service: 3; Value: 2
We went there for our anniversary. My soup was cold and expensive plus it felt like they hadn't painted since 1980.
Food: 3; Decor: 5; Service: 4; Value: 5
The food is only mediocre, but well worth the cost. Wait staff was friendly. Lot's of fun decorations.

→

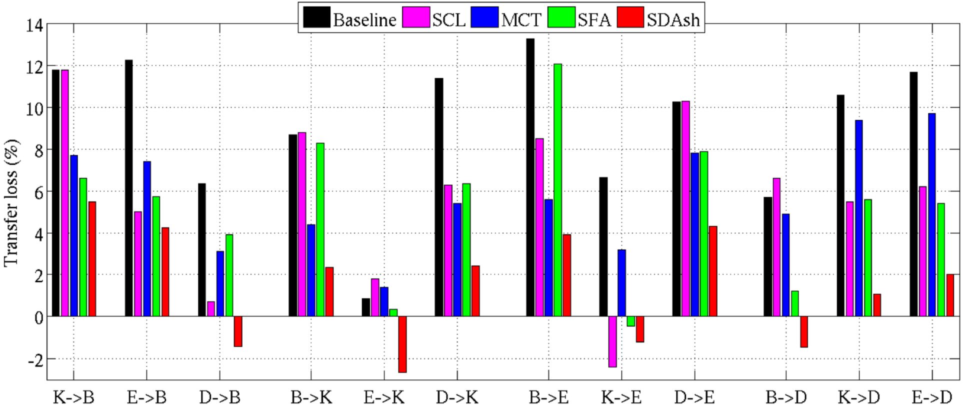
Food	"The chicken was great", "My soup was cold", "The food is only mediocre"
Decor	"it felt like they hadn't painted since 1980", "Lots of fun decorations"
Service	"service was excellent", "Wait staff was friendly"
Value	"the price was right", "My soup was cold and expensive", "well worth the cost"

I. Titov, R. McDonald, ACL 2008

A Joint Model of Text and Aspect Ratings for Sentiment Summarization

Semantic & autoencoders [Glorot et al. 2011]

- **Scaling** ⇒ neural networks & autoencoders (≈ stochastic PCA)
- Deep approach :
  1. Unsupervised learning : language modeling
  2. Fine Tuning : using the labels



Glorot, Bordes, Bengio, ICML 2011
Domain Adaptation for Large-Scale Sentiment Classification : A Deep Learning Approach

PLSA, PCA, Topic/sentiment

All most approaches operate at the document level...

- Extracting words or co-occurring words at the document level
  - topic modeling
  - sentiment for each topic
  - sentiment that are topic independent
- ... But no comprehensive model at the sentence level...
  - A real disadvantage for a *semantic* approach
- Also some scaling problem when bayesian graphical models become more & more complex

Conclusion

+ A general framework, critical in ML : **Transfer Learning**
+ Various tricks & algorithms useful for many NLP tasks
  - Yet another SVD use
  - New features
  - Graphical models

– Most approaches are deprecated due to recent advances in deep learning