More than Just Words: Discovering the Semantics of Text with a Minimum of Supervision

Marie-Francine Moens
Joint work with Erik Boiy,Jan De Belder, Koen Deschacht and Phi The Pham
Department of Computer Science
Katholieke Universiteit Leuven, Belgium

- We want to assign semantic meaning to content:
  - **Text**
  - Speech
  - Images
  - Video
  - Audio, ...

e.g., in the form of **semantic labels** => assist in search, mining, aggregation, summarization, ... of information

Text categorization

Google injects search savvy into display ad system

By MICHAEL LIEDTKE, AP Technology Writer - Fri Sep 18, 2009 4:28AM EDT

Add articles about technology to your My Yahoo! MY Y!

SAN FRANCISCO - Google Inc. is counting on the crown jewel of its online advertising empire to burnish a diamond in the rough.

Hoping to take an even bigger bite out of ad budgets, Google has melded the technology powering its lucrative search marketing network with a system that it bought 18 months ago to sell online billboards and other more visual commercials, including video.

Technology

The long-awaited combination poses another threat to Yahoo Inc., whose profits have been sliding the past three years. Yahoo is the Internet's largest seller of display advertising, a mantle that Google has set its sights on. Microsoft Corp. and Time Warner Inc.'s AOL also operate large exchanges that help manage display ads.

The upgrade announced Friday has been something Google has been working toward since it bought DoubleClick Inc. for $3.2 billion a year-and-a-half ago. Google prized DoubleClick largely for its tools for selling and serving display ads.

Sports

Canadian Press

Report: Gender tests on runner done in SAfrica

AP – 2 hrs 33 mins ago

JOHANNESBURG – A South African newspaper has published what it says are e-mails showing local track officials authorized gender tests done in the country on runner Caster Semenya. Full Story »

Sanofi buys Merck's half of animal health business

Associated Press

Buzz up! 0 votes | Send | Share | Print

By LINDA A. JOHNSON, AP Business Writer – 23 mins ago

RELATED QUOTES		
MRK	32.29	+0.28
PFE	16.64	+0.32
SAN	52.64	+0.68
SGP	28.31	+0.18
WYE	48.16	+0.31

TRENTON, N.J. – French drugmaker Sanofi-Aventis SA has completed its $4 billion purchase of Merck & Co.'s half interest in their veterinary medicine business, Merial Ltd., the companies said Friday.

Business

The move was required by regulators before Merck can close its $41 billion purchase of New Jersey neighbor Schering-Plough Corp., which also sells animal health products.

Merial, a joint venture founded in 1997, sells two widely used pet medicines, flea-and-tick blocker Frontline and chewable heartworm preventer Heartgard. It also sells Ivomec, which kills parasites in hogs and cattle, and other medicines and vaccines for livestock.

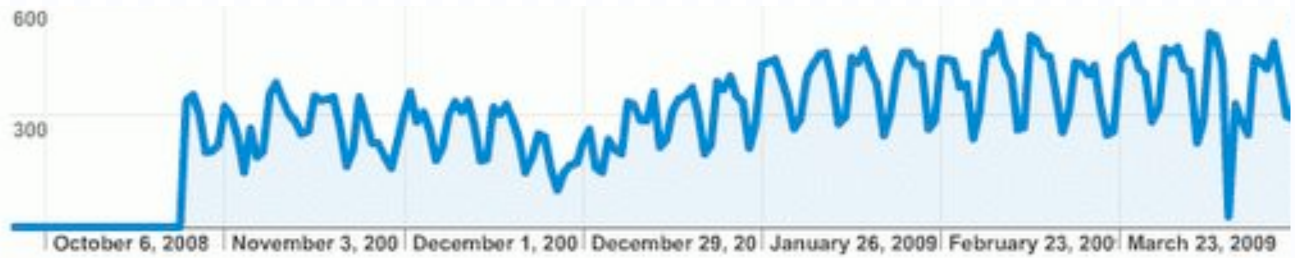
⇨ News filtering

Opinion recognition

Land Rover Range Rover Sport Massively improved, right up there with the nicer sporty-SUVs.

Just one thing: it's not a Range Rover.
The best Evo ever, Makkinen included.
But at 50,000 you've REALLY got to want one.

Mining: dashboard



Relation recognition

S_1: Search engine giant **Google** has bought videosharing website **YouTube** in a controversial $1.6 billion deal.

S_2: The companies will merge **Google**s search expertise with **You-Tube**s video expertise, pushing what executives believe is a hot emerging market of video offered over the Internet.

S_3: **Google** has acquired social media company, **YouTube** for $1.65 billion in a stock-for-stock transaction as announced by Google Inc. on October 9, 2006.

S_4: Drug giant **Pfizer Inc.** has reached an agreement to buy the priva-te biotechnology firm **Rinat Neuroscience** Corp., the companies announced Thursday.

S_5: He has also received consulting fees from Alpharma, Eli Lilly and Company, **Pfizer**, Wyeth Pharmaceuticals, **Rinat Neuroscience**, Elan Pharmaceuticals, and Forest Laboratories.

⇩

Search of a database

SELECT COMPANY

FROM ACQUISITION

WHERE ACQUIRER = ‘Google’

Semantic role labeling

Recognizing the basic event structure of a sentence
("**who**" "**does what**" "**to whom/what**" "**when**" "**where**", ...)

fall.01

Arg1: Logical subject, patient, thing falling

Arg2: Extent, amount fallen

Arg3: Start point

Arg4: End point, end state of Arg1

Ex1: [$_{Arg1}$ Sales] *fell* [$_{Arg4}$ to $251.2 million] [$_{Arg3}$ from $278.7 million].

Ex2: [$_{Ag1}$The average junk bond] *fell* [$_{Arg2}$ by 3.7%].

⇩

By how much has fallen the average junk bond?

Question answering search

Temporal relation recognition

```
John
<EVENT eid="e1" class="OCCURRENCE" tense="PAST"
aspect="PERFECTIVE">
left
</EVENT>
<MAKEINSTANCE eiid="ei1" eventID="e1"/>
<TIMEX3 tid="t1" type="DURATION" value="P2D"
temporalFunction="false">
2 days
</TIMEX3>
<SIGNAL sid="s1">
before
</SIGNAL>
the
<EVENT eid="e2" class="OCCURRENCE" tense="NONE"
aspect="NONE">
attack
</EVENT>
<MAKEINSTANCE eiid="ei2" eventID="e2"/>
```

Reconstruct the temporal sequence of events in a story

Classical approaches

- Parsing of text based on hand-coded **symbolic patterns**
- **Supervised learning of the patterns** based on annotated training data
- Typical features to describe the patterns:
  - Lexical features: unigram, bigram, ..
  - Syntactic features: Part-Of-Speech, dependency tree
  - Semantic features: obtained from knowledge resource or other extractions
  - Discourse and pragmatic features

What is the problem?

- Supervised techniques yield good results
- But, need a manual effort, which can be substantial given:
  - Many different semantic labels
  - Many languages, domains, etc.

- But results could be improved:
  - Large variation of patterns that have a similar meaning: low recall
  - Patterns are often ambiguous: low precision

Too much variation

Google owns YouTube

Google has acquired YouTube

YouTube is bought by Google

...

Too ambiguous

- The meaning of a word depends on the company it keeps

Take a right at the green plant which produces solar energy.

Tom Mitchell Center for Automated Learning and Discovery.

http://www.etoon.com/cartoon-store/Vlad-Kolarov/

- **Can we learn patterns for the semantic labeling with a minimum of supervision** taking into account the large variation of patterns and ambiguity?

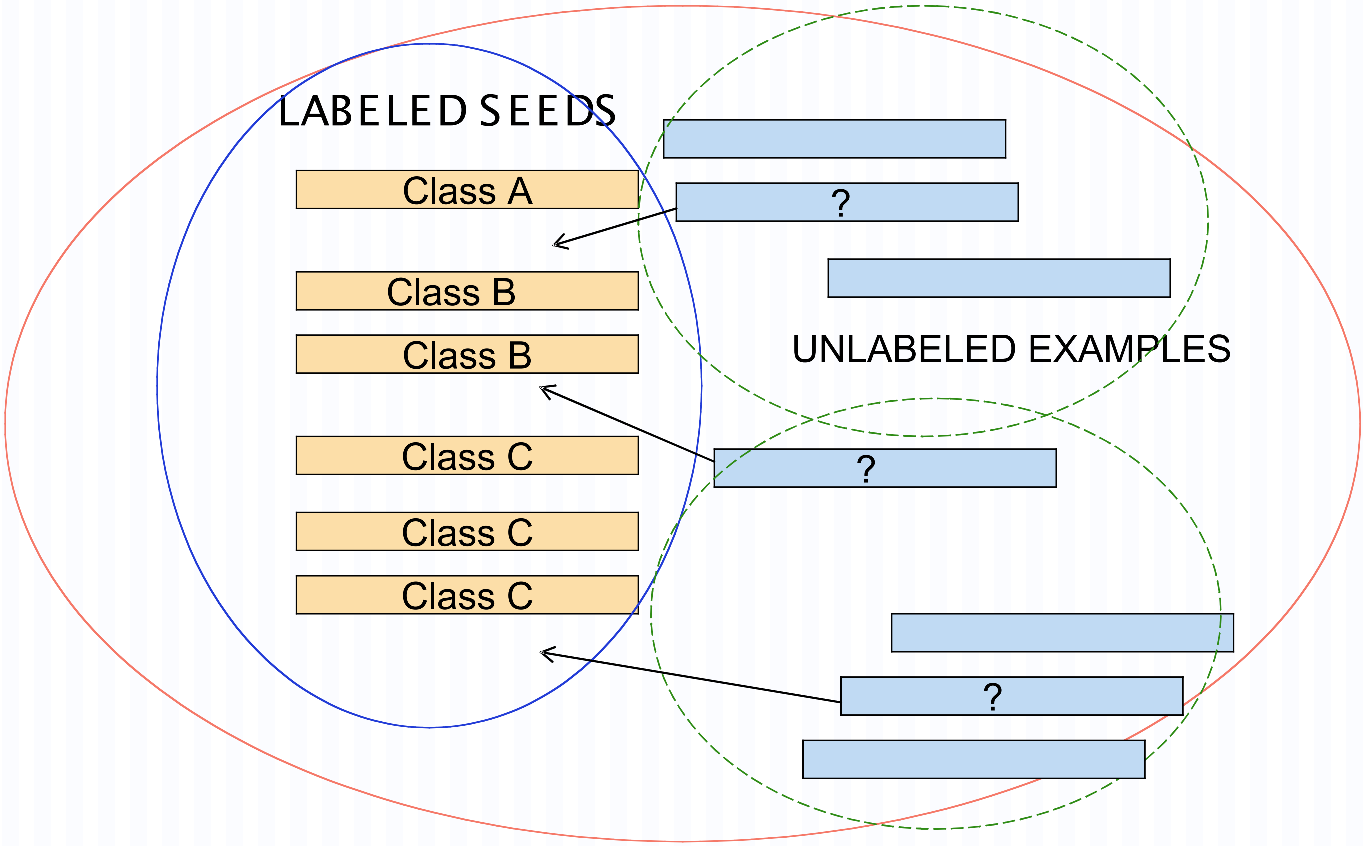
- In the following we focus on text, but many approaches can be ported to other media

Possible approaches (not exhaustive)

1. Examples are intelligently chosen by the machine and annotated by a human
2. Examples are automatically selected by the machine and annotated by the machine
3. A classifier is trained from labeled and unlabelled examples
4. We use few examples for training, but use additional constraints
5. We use few examples for training, but perform a constrained expansion
6. Multimodal processing

1. Examples are intelligently chosen by the machine and annotated by a human

Active learning



Opinion recognition

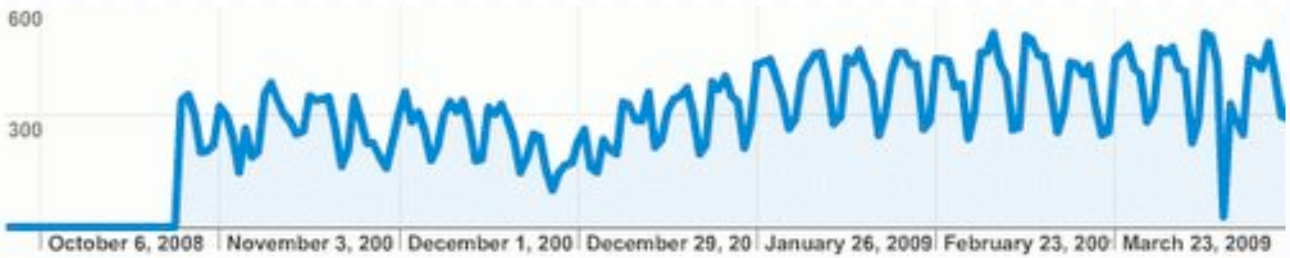
ATTENTIO™
Brand Dashboard
Market intelligence. Real-time.

\- *The movie really seems to be spilling the beans on a lot of stuff we didnt think we hand if this is their warm up, what is going to get us frothing in December*

-+ *de grote merken mogen er dan patserig uitzien en massa's pk hebben maarals de bomen wat dicht bij elkaar staan en de paadjes steil enbochtig,dan verkies ik mijn Jimny .*

\+ *L'é tro bel cet voitur Voici tt ce ki me pasione ds ma petite vi!!!é tt mé pote é pl1 dotre truk!!!Avou de Dcouvrir*

Mining: dashboard



Active learning for opinion recognition

- Selection of examples:
  - Uncertainty sampling: examples for which the current classifier assigns a class with low probability or confidence [Lewis & Catlett ICML 1994] [Tong & Koller JMLR 2001]
  - Redundancy and diversity sampling [Baram et al. ICML 2003]
  - Relevance sampling: to obtain more examples from a certain category
  - ...

Active learning for opinion recognition

- Best results for classifying sentences as positive, negative or neutral with regard to an input product name on English blogs (e.g., skyrock.com, live-journal.com, xanga.com, blogspot.com, forums.automotive.com):
  - Combination of:
    - **Random sampling** (to obtain some diversity of the patterns)
    - **Uncertainty sampling** (distance to the hyperplane found by Support Vector Machine): side effect: **reduction of redundancy**
    - **Relevance sampling**

    => less annotation and same accuracy

[Boiy & Moens IR 2009]

2. Examples are automatically selected by the machine and annotated by the machine

- World Wide Web = huge source of unlabeled examples !
- **Relation extraction**:
  - Search sentences on the Web:
    - containing 2 entities for which target relation holds => **positive bag**, but might contain some negative examples
    - containing 2 entities for which target relation does not hold = > **negative bag**, assumed to contain only negative examples

⇒ input for **Multiple Instance Learning** algorithm

[Bunescu & Mooney ACL 2007]

+/S_1: Search giant **Google** has bought video-sharing website **YouTube** in a controversial \$1.6 billion deal

-/S_2: The companies will merge **Google**"s search expertise with **YouTube**"s video expertise, pushing what executives believe is a hot emerging market of video offered over the internet

+/S_3: **Google** has acquired social media company, **YouTube** for \$1.65 billion in a stock-for-stock transaction as announces bu Google Inc. on October 9, 2006

+/S_4: Drug giant **Pfizer Inc.** has reached an agreement to buy the private biotechnology form **Rinat Neuroscience Corp.**, the companies announced Thursday.

-/S_5: Ha has also received consulting fees from Alpharma, Eli Lilly and Company, **Pfizer**, Wyeth Pharmaceuticals, **Rinat Neuroscience**, Elan Pharmaceuticals, and Forrect Laboratories

Subsequence kernel for relation extraction

[Bunescu & Mooney NIPS 2006] adapted with weighting scheme that takes into account frequency of a word (SSK-T2)

$$
\begin{aligned}
rK(s,t) &= fbK(s,t) + bK(s,t) + baK(s,t) \\
bK_i(s,t) &= K_i(s_b, t_b, 1) \cdot c(x_1, y_1) \cdot c(x_2, y_2) \cdot \lambda^{l(s_b^{'}) + l(t_b^{'})} \\
fbK(s,t) &= \sum_{i,j} bK_i(s,t) \cdot K_j^{'}(s_f, t_f), \quad 1 \leq \mathrm{i},\ 1 \leq \mathrm{j},\ \mathrm{i}+\mathrm{j} < \mathrm{fb}_{\mathrm{max}} \\
bK(s,t) &= \sum_{i} bK_i(s,t), \quad 1 \leq \mathrm{i} \leq \mathrm{b}_{\mathrm{max}} \\
baK(s,t) &= \sum_{i,j} bK_i(s,t) \cdot K_j^{'}(s_a^{-}, t_a^{-}), \quad 1 \leq \mathrm{i},\ 1 \leq \mathrm{j},\ \mathrm{i}+\mathrm{j} < \mathrm{ba}_{\mathrm{max}}
\end{aligned}
$$

Figure 3: Computation of relation kernel.

Adaptation of a Least Squares SVM [Suykens et al. 2002] to a MIL setting: Weighted Least Squares SVM (WLS-SVM)

[De Belder, De Smet, Mochales & Moens SIM 2009]

WLS-SVM: errors in positive and negative bags are weighted differently to comply with MIL setting

Minimize over *w,b* and *e:*

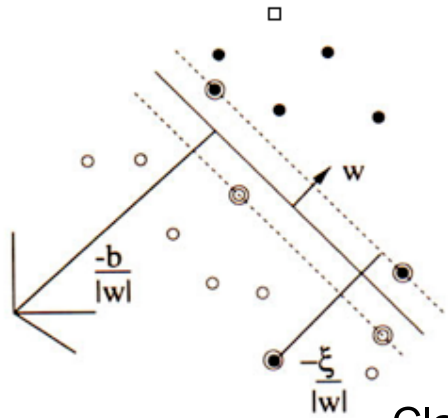
$$J(w,b,e) = \frac{1}{2}\|w\|^2 + \frac{C}{L}\left(v_p \Xi_p + \right.$$

$$\Xi_p = \sum_{B_i \in B^+} \sum_{x \in B_i} e_x^2$$

$$\Xi_n = \sum_{B_i \in B^-} \sum_{x \in B_i} e_x^2$$

$$v_p = \frac{1}{\gamma c_p}$$

$$v_n = \frac{1}{\gamma c_n}$$



Classical SVM

Subject to constraints:

$$w\phi(x) + b = +1 - e_x, \qquad \forall x \in B_i \in B^+,$$
$$w\phi(x) + b = -1 + e_x, \qquad \forall x \in B_i \in B^-,$$

[De Belder et al. 2009]

	SSK	BOW	SSK-T1	BOW-T2	SSK-T2
WLS:	80,41	33,46	85,80	55,08	**87,08**
QP:	77,62	25,50	83,93	44,77	86,09

Own dataset

	SSK	BOW	SSK-T1	BOW-T2	SSK-T2
WLS:	52,00	22,90	71,00	39,56	**86,28**
QP:	51,18	23,63	67,16	30,84	62,65

Dataset Raymond Mooney

Results “Acquisition” relation

	SSK	BOW	SSK-T1	BOW-T2	SSK-T2
WLS:	84,90	17,54	95,08	52,30	94,95
QP:	88,51	52,61	94,00	77,96	**96,07**

Own dataset

	SSK	BOW	SSK-T1	BOW-T2	SSK-T2
WLS:	68,07	8,76	87,54	18,92	89,38
QP:	84,77	18,69	88,40	20,00	**90,02**

Dataset Raymond Mooney

Results “Born in” relation

3. A classifier is trained from labeled and unlabelled examples

- = **Semi-supervised learning**: most known forms are:
  - Self-learning: iterative retraining after labeling of data points for which the current model is most confident
  - Transductive inference: no general decision rule is inferred, only the labels of the unannotated examples are predicted according to a most likely model

Self-training

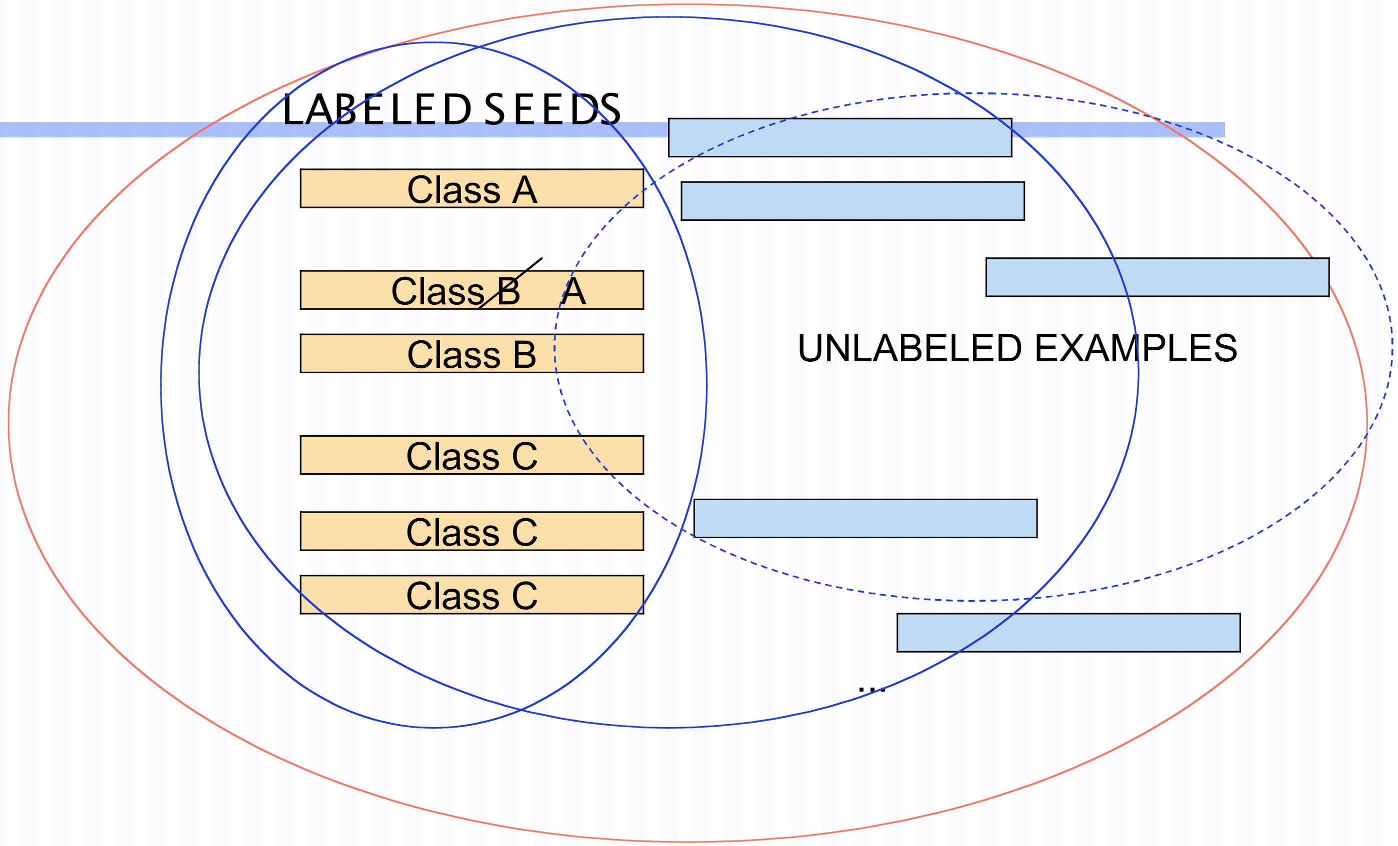


Fig. 6.3. Self-training: A classifier is incrementally trained (blue line), first based on the labeled seeds, and then based on the labeled seeds and a set of unlabeled examples that are labeled with the current classifier. The dotted blue line represents the set of all unlabeled examples that were considered for labeling in this step.

Benefit of semi-supervised learning

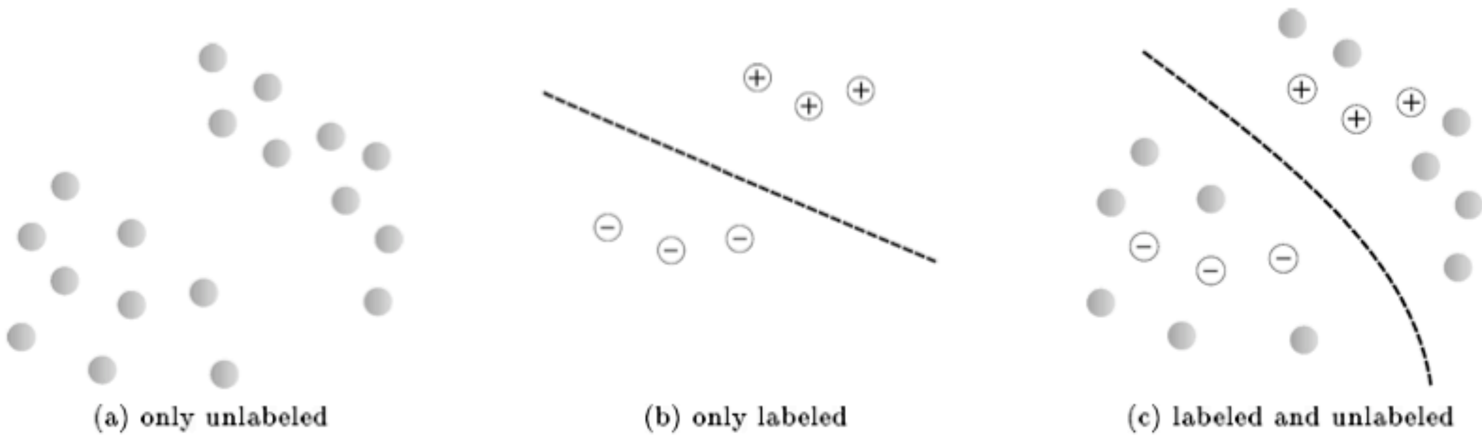


Figure 2: Schematic figure illustrating how unlabeled data might improve a supervised classifier. Grey dots are unlabeled data, white dots labeled data and the dotted line the classification boundary.

But does this hold for all types of text classification?

Google injects search savvy into display ad system

By MICHAEL LIEDTKE, AP Technology Writer - Fri Sep 18, 2009 4:28AM EDT

Add articles about technology to your My Yahoo! MY Y!

SAN FRANCISCO - Google Inc. is counting on the crown jewel of its online advertising empire to burnish a diamond in the rough.

Hoping to take an even bigger bite out of ad budgets, Google has melded the technology powering its lucrative search marketing network with a system that it bought 18 months ago to sell online billboards and other more visual commercials, including video.

Technology

The long-awaited combination poses another threat to Yahoo Inc., whose profits have been sliding the past three years. Yahoo is the Internet's largest seller of display advertising, a mantle that Google has set its sights on. Microsoft Corp. and Time Warner Inc.'s AOL also operate large exchanges that help manage display ads.

The upgrade announced Friday has been something Google has been working toward since it bought DoubleClick Inc. for $3.2 billion a year-and-a-half ago. Google prized DoubleClick largely for its tools for selling and serving display ads.

Sports

Canadian Press

Report: Gender tests on runner done in SAfrica

AP – 2 hrs 33 mins ago

JOHANNESBURG – A South African newspaper has published what it says are e-mails showing local track officials authorized gender tests done in the country on runner Caster Semenya. Full Story »

Sanofi buys Merck's half of animal health business

Associated Press

Buzz up! 0 votes | Send | Share | Print

By LINDA A. JOHNSON, AP Business Writer – 23 mins ago

TRENTON, N.J. – French drugmaker Sanofi-Aventis SA has completed its $4 billion purchase of Merck & Co.'s half interest in their veterinary medicine business, Merial Ltd., the companies said Friday.

Business

The move was required by regulators before Merck can close its $41 billion purchase of New Jersey neighbor Schering-Plough Corp., which also sells animal health products.

RELATED QUOTES

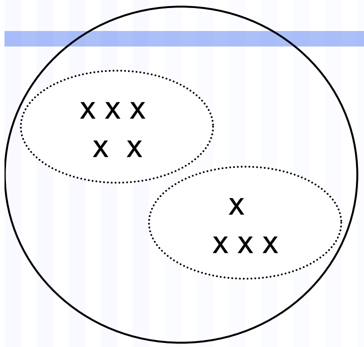
MRK	32.29	+0.28
PFE	16.64	+0.32
SAN	52.64	+0.68
SGP	28.31	+0.18
WYE	48.16	+0.31

Merial, a joint venture founded in 1997, sells two widely used pet medicines, flea-and-tick blocker Frontline and chewable heartworm preventer Heartgard. It also sells Ivomec, which kills parasites in hogs and cattle, and other medicines and vaccines for livestock.

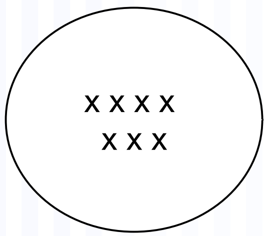
⇨ News filtering

Text categorization

- Good results:
  - Generative model: e.g., self learning with Naive Bayes and Expectation Maximization, e.g., ca. average 95% accuracy on standard Reuters text categories, but use of multiple mixture components per class [Nigam, McCallum & Mitchell SL 2006]
  - Discriminative model: transductive learning with SVM e.g., better results when few training data, approach results SVM with more training data (accuracy > 80 %) [Joachims SL 2006]



Semi-supervised smoothness assumption: if two points in a high density region are close, so should be the corresponding classes



Cluster assumption: the points of each class tend to form a cluster

=>These assumptions do not necessarily hold for fine-grained text classification tasks !

Manifold assumption: curse of dimensionality: many features → many training examples

[Chapelle, Schölkopf & Zien SL 2006]

When does semi-supervised learning work?

- For semi-supervised learning to work: it is an important prerequisite that the distribution of examples, which the unlabeled examples help elucidate, is relevant for the classification problem [Chapelle, Schölkopf & Zien SL 2006]

- Model learned from the labeled examples should be rather correct [Cozman & Cohen SL 2006]

- => evidenced by own research for semantic role labeling [Deschacht & Moens Technical report 2009]

4. We use few examples for training, but use additional constraints

- Knowledge of language or cognitive knowledge on how people understand text might help:
  - For selecting seed labeled examples
  - For adding additional constraints
- Illustrated with the recognition of temporal information in text

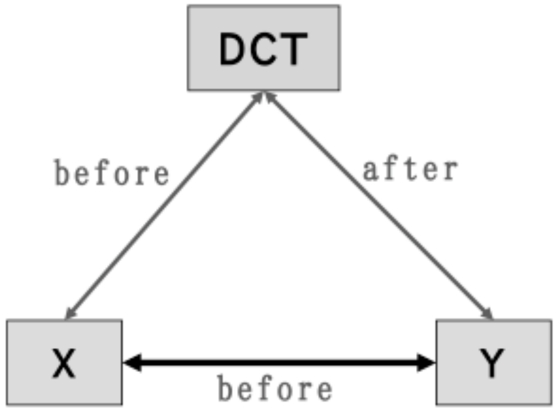
Recognition of temporal relations

- Between an event and time expressions that occur within the same sentence
- Between document creation time and an event
- Between the main events of adjacent sentences

⇒ Markov Logic model that jointly identifies these relations:

  ⇒ Explicit incorporation of (soft) temporal constraints

[Yoshikawa, Riedel, Asahura & Matsumoto ACL 2009]



Constraints help in **resolving ambiguity**: [Yoshikawa et al. 2009] reported 2% rise in accuracy compared to state of the art (average accuracy over above tasks = 68.9%)

=> **But still problem of the variety of patterns in language**

5. We use few examples for training, but perform a constrained expansion

Illustrated with semantic role labeling:

fall.01

Arg1: Logical subject, patient, thing falling

Arg2: Extent, amount fallen

Arg3: Start point

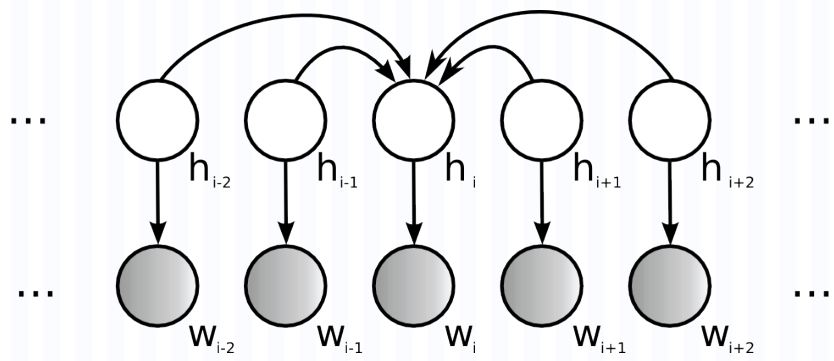
Arg4: End point, end state of Arg1

Ex1: [$_{Arg1}$ Sales] *fell* [$_{Arg4}$ to $251.2 million] [$_{Arg3}$ from $278.7 million].

Ex2: [$_{Ag1}$The average junk bond] *fell* [$_{Arg2}$ by 3.7%].

Latent Words Language Model

- Generative model of natural language
- Latent variable (**hidden word**) models *words* that have a *similar meaning* in a *specific* left and right *context*



[Deschacht & Moens EMNLP 2009]

Latent Words Language Model

- Model is trained on large corpus:
  - Initialization with trigram language model and Kneser-Ney smoothing
  - Updated with Gibbs sampling

$$P^{\tau}(h_j|w_j, \mathbf{h}_0^{j-1}, \mathbf{h}_{j+1}^{Z}) = \frac{P^{\tau}(w_j|h_j)P^{\tau}(h_j|\mathbf{h}_{j-2}^{j-1}\mathbf{h}_{j+1}^{j+2})}{\sum_{h_i} P^{\tau}(w_i|h_i)P^{\tau}(h_j|\mathbf{h}_{j-2}^{j-1}\mathbf{h}_{j+1}^{j+2})}$$

- Context-dependent distribution of hidden words can be inferred for a new text

Latent Words Language Model

Perplexity on unseen text

(when converted to sequential tri-gram model):

	Reuters	APNews	EnWiki
Kneser-Ney LM	113.15	132.99	160.83
Class based LM	108.38	125.65	149.21
LWLM	**99.12**	**116.65**	**148.12**

=> Outperforms other current language models for predicting the English language

[Deschacht & Moens Benelearn 2009]

Examples on unseen texts

Compuserve	**corp**	**said**	**Tuesday**	**it**	**anticipates**	**a**	**loss**
Microsoft	inc	told	Friday	they	expects	the	profit
Crysler	corp.	reported	Thursday	he	expected	some	gain
Oracle	ltd	added	Monday	she	assumes	an	deficit
Software	co	say	Wednesday	this	doubts	another	earnings

A	**Japanese**	**electronics**	**executive**	**was**	**kidnapped**	**in**	**Mexico**
the	u.s.	tobacco	director	is	abducted	on	usa
its	german	sales	manager	we	killed	at	uk
an	brish	consulting	economist	are	found	of	australia
one	russian	electric	spokesman	be	abduction	into	canada

Latent words for SRL

- Latent words are used as probabilistic features in MEMM for classification (results as F_1 - CoNLL dataset):

Amount of training data	5%	20%	50%	100%
Standard SRL	40.49%	67.23%	74.93%	78.65%
SRL + LW	60.29%	72.88%	76.42%	80.98%

→ **Word expansion** improves recall

→ **Word sense disambiguation** improves precision

→ **Easy to use** in many other NLP applications

[Deschacht & Moens EMNLP 2009]

6. Cross-modal processing

- Unsupervised alignment of names and faces in "Labeled faces in the wild" dataset: from Yahoo!news

  [Pham, Moens & Tuytelaars IEEE Trans. Multimedia in press]

Vice President **Dick Cheney** speaks at a luncheon for Republican U.S. Senate candidate **John Cornyn** Friday, July 19, 2002, in Houston. (AP Photo/Pat Sullivan)

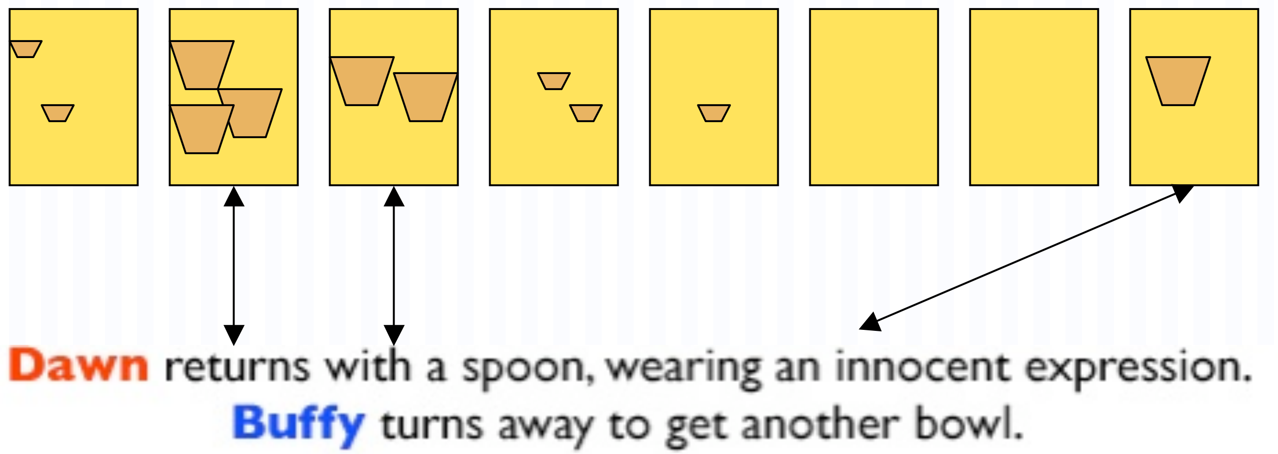
President-elect **Barack Obama** is inch-ing closer to naming former rival Sen. **Hillary Clinton** as his secretary of state, ABC News has learned. (Getty Images)

Danish director **Lars Von Trier** (C), Australian actress **Nicole Kidman** and Swedish actor **Stellan Skarsgard** (L) pose on a terrace of the Palais des festivals.(AFP/Boris Horvat)

Cross-modal processing

- Adaptation to alignment in **video** data

Time warping to align scripts and subtitles



Willow hugs Buffy.

Recognition of **event signaling verbs, nouns** though analysis of the video?

[CIAM IJCAI-09 workshop]

Conclusions

- Learning with a **minimum of supervision** when semantic labeling of text: very useful
- But, **understudied problem**
  - Active learning
  - Model building for semi-supervised learning
  - Frequent phenomena allow already extracting semantic knowledge: large data sets on the Web help !
  - Better knowledge on the distributions of language patterns in context will help: already shown with LWLM
  - Multimodal processing: AI again becoming a more integrated discipline?

References

Baram, Y., El-Yaniv, R. & Luz, K. (2003). Online choice of active learning algorithms. In *Proceedings of ICML-2003* (pp. 19-26).

Boiy, E. & Moens, M.-F. (2009). Sentiment Analysis in Multilingual Web Texts. *Information Retrieval*, 12 (5), 526-558.

Bunescu, R. & Mooney, R. (2006). Subsequence kernels for relation extraction. *Advances in Neural Information Processing Systems,* 18.

Bunescu, R. & Mooney, R. (2007). Learning to extract relations from the web using minimal supervision. In *Proceedings of the Annual Meeting for Computational Linguistics.*

Chapelle, O., Scholköpf, B. & Zien, A. (Eds.) (2006). *Semi-supervised Learning.* Cambridge, MA: MIT Press.

Cosman, F. & Cohen, I (2006). Risks of semi-supervised learning. In O. Chapelle, B. Scholköpf, B. & A. Zien (Eds.) (2006). *Semi-supervised Learning* (pp. 57-72). Cambridge, MA: MIT Press.

De Belder, J., De Smet, W., Mochales Palau, R. & Moens, M.-F. (2009). Google Owns YouTube? Entity Relationship Extraction with Minimal Supervision. In *Proceedings of the SIM Joint* Conferences: *ILP, SRL and MLG.*

Deschacht, K. & Moens, M.-F. (2009). The Latent Words Language Model. In *Proceedings of the 18th Annual Belgian-Dutch Conference on Machine Learning* (*Benelearn 09*).

Deschacht, K. & Moens, M.-F. (2009). Using the Latent Words Language Model for Semi-Supervised Semantic Role Labeling. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (*EMNLP 2009*). ACL.

Deschacht, K. & Moens, M.-F. (2009). Semantic role labeling in text. Learning from labeled and unlabeled examples. Technical Report.

Ferrari, V., Moens, M.-F.,Tuytelaars, T. & Van Gool (Eds.) (2009). *Proceedings of IJCAI-09 Workshop on Cross-media Information Access and Mining* (*CIAM 2009*). AAAI, (50 p.).

Joachims, T. (2006). Transductive support vector machines. In O. Chapelle, B. Scholköpf, B. & A. Zien (Eds.) (2006). *Semi-supervised Learning* (pp. 115-135). Cambridge, MA: MIT Press.

Lewis,D.D. & Catlett, J. (1994). Heterogeneous uncertaintly sampling for supervised learning. In *Proceedings of ICML.*

Nigam, K., McCallum, A. & Mitchell, T. (2006). Semi-supervised classification using EM. In O. Chapelle, B. Scholköpf, B. & A. Zien (Eds.) (2006). *Semi-supervised Learning* (pp. 33-55). Cambridge, MA: MIT Press.

Pham, P.T., Moens, M.-F., Tuytelaars, T. (in press). Cross-media alignment of names and faces. *IEEE Transactions on Multimedia.*

Suykens, J., De Brabanter, J., Lukas, L & Vandewalle, J. (2002). Weighted least squares support vector machines: Robustness and sparse approximation. *Neurocomputing,* 48, 85-105

Tong, S. & Koller, D. (2001). Support Vector Machine Active Learning with Applications to Text Classification. *Journal of Machine Learning Research*, 45-66.