

Learning Discriminative Projections for Text Similarity Measures

Scott Wen-tau Yih

Joint work with

Kristina Toutanova, John Platt, Chris Meek

Microsoft Research

Cross-language Document Retrieval

Spanish Document Set

English Query Doc

Article Discussion Read View source

Penguin

From Wikipedia, the free encyclopedia
(Redirected from Spheniscidae)

For other uses, see [Penguin \(disambiguation\)](#).

Penguins (order **Sphenisciformes**, family **Spheniscidae**) are a group of aquatic, flightless birds living almost exclusively in the southern hemisphere, especially in Antarctica. Highly adapted for life in the water, penguins have countershaded dark and white plumage, and their



Spheniscidae

«Pinguino» *redirige aquí*. Para otras acepciones, véase *Pinguino*.

La familia **Spheniscidae** abarca al conjunto de aves marítimas que se observan en las costas del Hemisferio Sur. Estas aves se caracterizan por su andar torpe y erguido y al ser un ave incapaz de volar. Los animales los llamaron *penguins* (del gaélico *penwyn*, *pe* = gigante del Atlántico norte (*Pinguinus impennis*). Sin embargo, por convergencia evolutiva, los pingüinos del Hemisferio Norte también se extinguieron al alca gigante a fines del siglo XIX, el nombre se aplicó a los pingüinos. Existen 18 diferentes especies de pingüinos.

Balaenidae

«Balaeno» *redirige aquí*. Para otras acepciones, véase *Balaeno*.

Los **balaenos**, **Balaenidae** son una familia de cetáceos (mamíferos) que incluye dos especies: el balaeno común (*Balaena mysticetus*) y el balaeno negro (*Balaena borealis*). Son los cetáceos más grandes que existen, con una longitud que puede aproximadamente 30 metros.

Web Search & Advertising

🔍

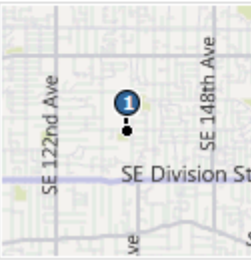
Web
Local
More ▾

ALL RESULTS 1-10 of 298,000 results · [Advanced](#)

[ACL HLT 2011](#)
 The 49th Annual Meeting of the **Association for Computational Linguistics: Human Language Technologies**. To be held at the **Portland Marriott Downtown Waterfront** in acl2011.org

[Acl PORTLAND](#)
 Results for **Acl in Portland**, **acl** reviews . **Acl Portland** and results for arthroscopy, joint, shoulder, orthopaedic, knee, hip, total hip replacement, orthopedic care ...
www.cylex-usa.com/portland/acl.html

[ACL Construction LLC \(Portland\)](#)
www.superpages.com/bp/Portland-OR/ACL-Construction-LLC-L2145388956.htm
ACL Construction LLC in Portland, OR -- Map, Phone Number, Reviews, Photos and Video Profile for **Portland ACL Construction LLC**. **ACL Construction LLC** appears in: Building ...



ACL Construction Llc
 13333 SE Mill St · Portland · (503) 505-7969
[Directions](#)

Neighborhood: Hazelwood


Ads

[Austin City Limits Tix](#)
 Cheap Austin City Limits Tickets. Check Our Prices. Save 15% or More.
www.TicketLiquidator.com

[Expert Knee Surgeons](#)
 Get the best knee doctor for your torn **ACL** surgery.
EverettBoneAndJoint.com/Knee

[See your message here](#)

Web Search & Advertising

ACL in Portland 

Web Local More ▾

ALL RESULTS 1-10 of 298,000 results · [Advanced](#)

Query: ACL in Portland

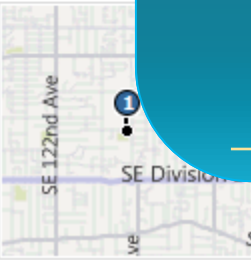
[ACL Construction LLC \(Portland\)](#)
ACL Construction LLC in Portland, OR --
Map, Phone Number, Reviews, ...
www.superpages.com

[ACL HLT 2011](#)
The 49th Annual Meeting of the Association
for Computational Linguistics...
acl2011.org

ACL HLT 2011
The 49th Annual Meeting of the Association for Computational Linguistics...
acl2011.org

ACL PORTLAND
Results for ACL Construction LLC in Portland, OR --
www.cylex.com

ACL Construction LLC
www.superpages.com
ACL Construction LLC Profile for Portland, OR



Ads

[Austin City Limits Tix](#)
Austin City Limits Tickets.
Our Prices. Save 15% or More.
TicketLiquidator.com

[Knee Surgeons](#)
The best knee doctor for your torn
knee surgery.
BoneAndJoint.com/Knee
[Send us your message here](#)

Web Search & Advertising

ACL in Portland

Web Local More

ALL RESULTS

ACL HLT 2011
The 49th Annual
Language Techn
acl2011.org

AcI PORTLA
Results for AcI
shoulder, orthop
www.cylex-usa.com

ACL Constr
www.superpage.com
ACL Construction
Profile for Portla

SE 122nd Ave
SE 148th Ave
SE Division St

Neighborhood: Hazelwood

Query: ACL in Portland

Don't Have ACL Surgery
Used By Top Athletes Worldwide
Don't Let... See Us First
www.arpw.com

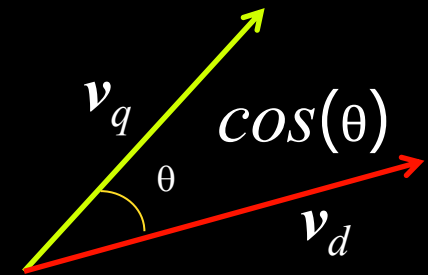
Expert Knowledge
Get the best... for your torn
ACL surgery.
EverettBoneAndJoint.com/Knee

City Limits Tix
...in City Limits Tickets.
Prices. Save 15% or More.
Liquidator.com

Knee Surgeons
...st knee doctor for your torn
y.
EverettBoneAndJoint.com/Knee
message here

ACL: Anterior Cruciate Ligament injuries

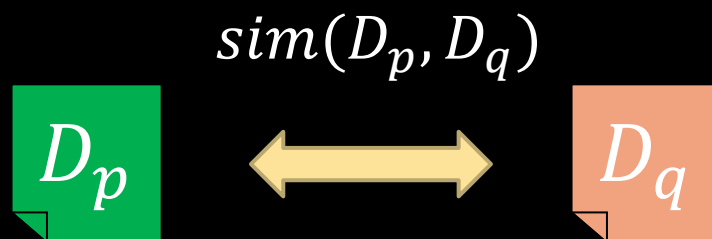
Vector Space Model



- Represent text objects as vectors
 - Word/Phrase: term co-occurrences
 - Document: term vectors with TFIDF/BM25 weighting
 - Similarity is determined using functions like cosine of the corresponding vectors
- Weaknesses
 - Different but related terms cannot be matched
 - e.g., (buy, used, car) vs. (purchase, pre-owned, vehicle)
 - Not suitable for cross-lingual settings

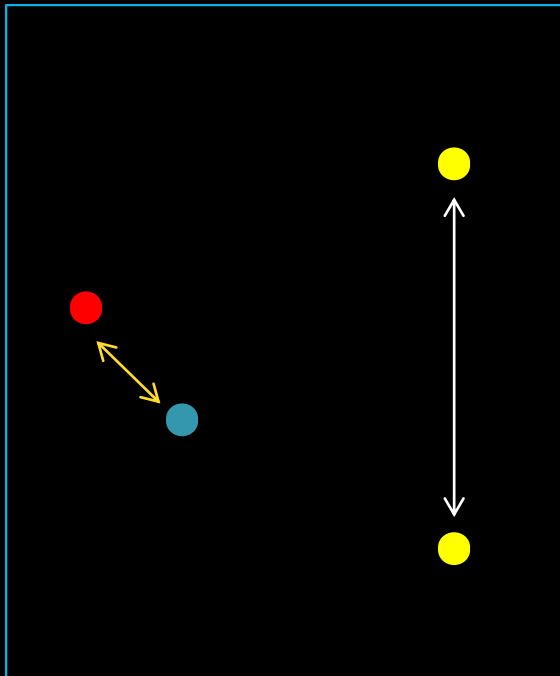
Learning Concept Vector Representation

- Are D_p and D_q relevant or semantically similar?

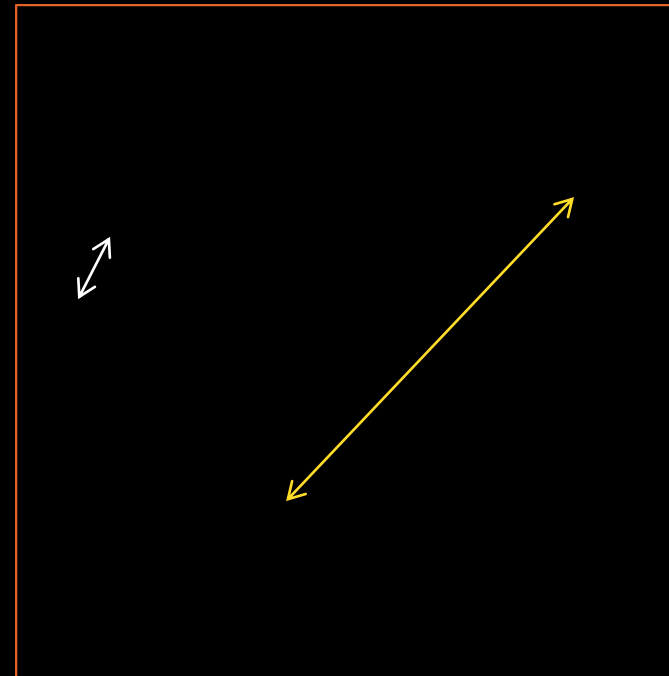


- Input: high-dimensional, sparse term vectors
- Output: low-dimensional, dense concept vectors
- Model requirements
 - Transformation is easy to compute
 - Provide good similarity measures

Ideal Mapping



High-dimensional space



Low-dimensional space

Dimensionality Reduction Methods

Supervised

CPLSA

OPCA

S2Net

PLTM

HDLR

JPLSA

CCA

CL-LSI

Unsupervised

PLSA

PCA

LDA

LSA

Probabilistic

Projection

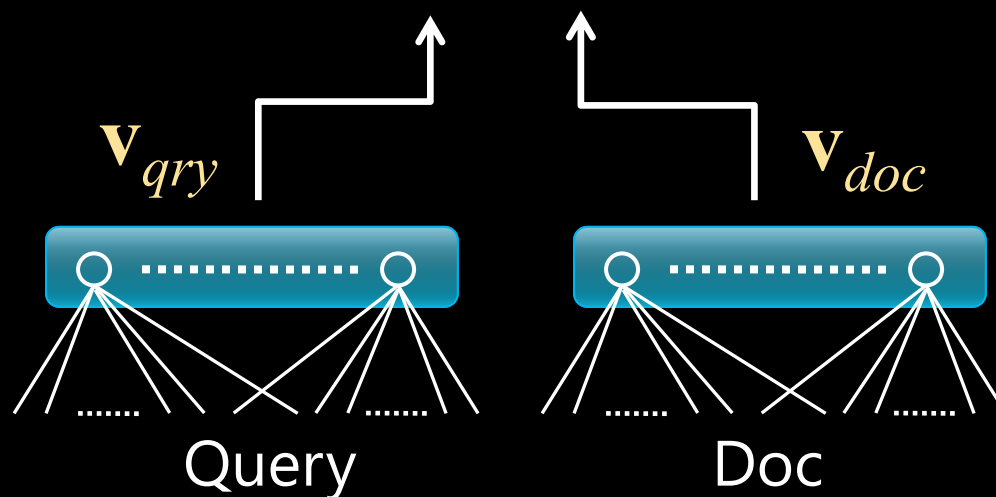
Outline

- Introduction
- **Problem & Approach**
- Experiments
 - Cross-language document retrieval
 - Ad relevance measures
 - Web search ranking
- Discussion & Conclusions

Goal – Learn Vector Representation

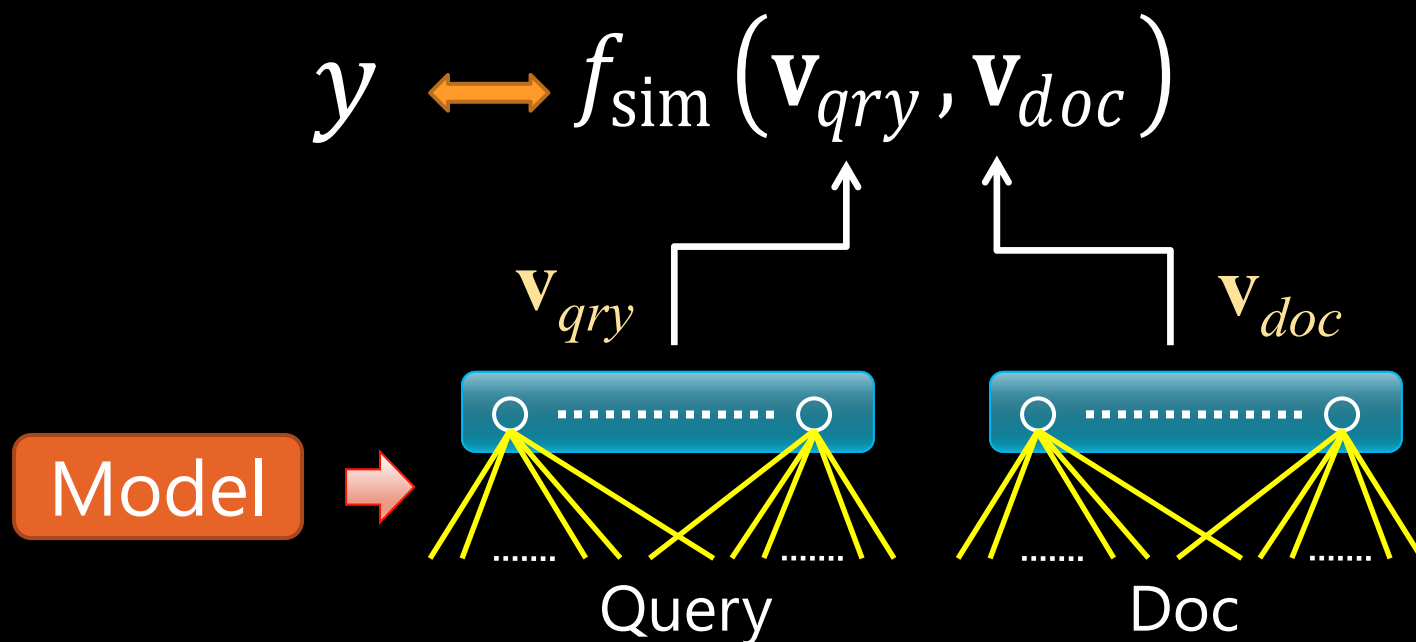
- Approach: Siamese neural network architecture
 - Train the model using labeled (query, doc)
 - Optimize for pre-selected similarity function (cosine)

$$y \iff f_{\text{sim}}(\mathbf{v}_{\text{qry}}, \mathbf{v}_{\text{doc}})$$



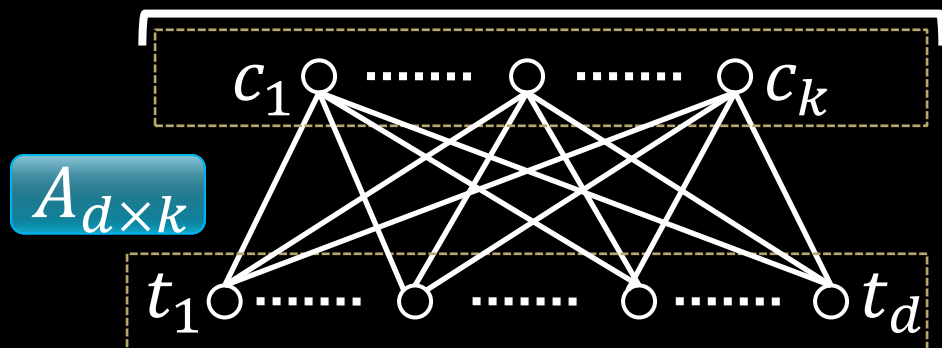
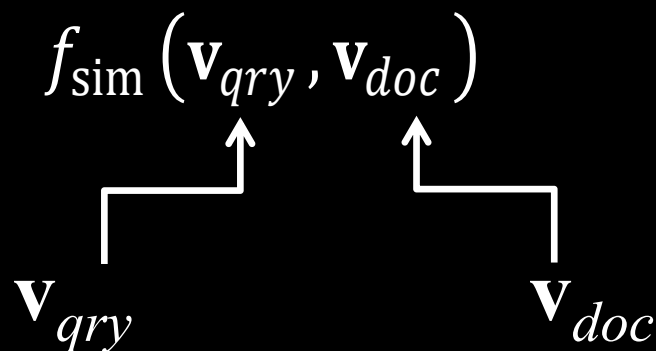
Goal – Learn Vector Representation

- Approach: Siamese neural network architecture
 - Train the model using labeled (query, doc)
 - Optimize for pre-selected similarity function (cosine)



S2Net – Similarity via Siamese NN

- Model form is the same as LSA/PCA
- Learning the projection matrix discriminatively



$$V_{\text{qry}} = A^T F_{\text{qry}}$$

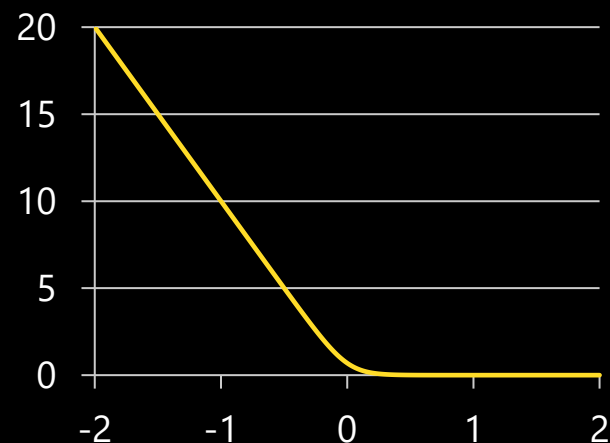
Pairwise Loss – Motivation

- In principle, we can use a simple loss function like mean-squared error: $\frac{1}{2} \left(y - f_{sim}(v_{qry}, v_{doc}) \right)^2$. But...

	<i>Doc</i> ₁	–	0.95
	<i>Doc</i> ₂	+	0.89
<i>Query</i>	<i>Doc</i> ₃	–	0.54
	<i>Doc</i> ₄	+	0.43
	<i>Doc</i> ₅	–	0.21

Pairwise Loss

- Consider a query q and two documents d_1 and d_2
 - Assume d_1 is more related to q , compared to d_2
- F_q, F_{d_1}, F_{d_2} : original term vectors of q, d_1 and d_2
- $\Delta = f_{sim}(A^T F_q, A^T F_{d_1}) - f_{sim}(A^T F_q, A^T F_{d_2})$
- $Loss(\Delta; A) = \log(1 + \exp(-\gamma\Delta))$
 - γ : scaling factor, as $\Delta \in [-2, 2]$
 - $\gamma = 10$ in the experiments



Model Training

- Minimizing the loss function can be done using standard gradient-based methods
 - Derive batch gradient and apply L-BFGS
- Non-convex loss
 - Starting from a good initial matrix helps reduce training time and converge to a better local minimum
- Regularization
 - Model parameters can be regularized by adding a smoothing term $\frac{\beta}{2} \|A - A_0\|^2$ in the loss function
 - Early stopping can be effective in practice

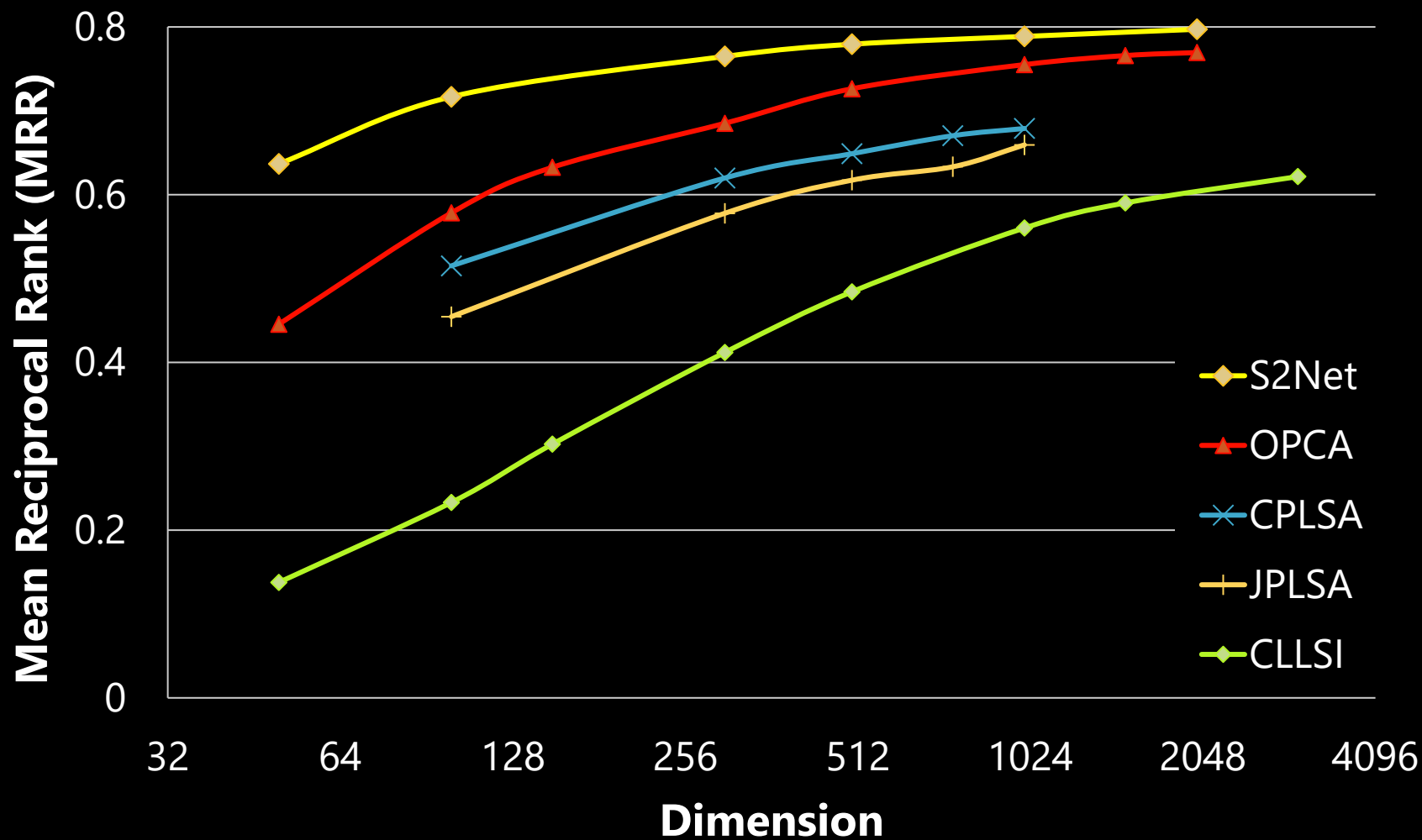
Outline

- Introduction
- Problem & Approach
- Experiments
 - Cross-language document retrieval
 - Ad relevance measures
 - Web search ranking
- Discussion & Conclusions

Cross-language Document Retrieval

- Dataset: pairs of Wiki documents in EN and ES
 - Same setting as in [Platt et al. EMNLP-10]
 - #document in each language
 - Training: 43,380, Validation: 8,675, Test: 8,675
 - Effectively, $43380^2 = 1.88$ billion training examples
 - Positive: EN-ES documents in the same pair
 - Negative: All other pairs
- Evaluation: find the comparable document in the different language for each query document

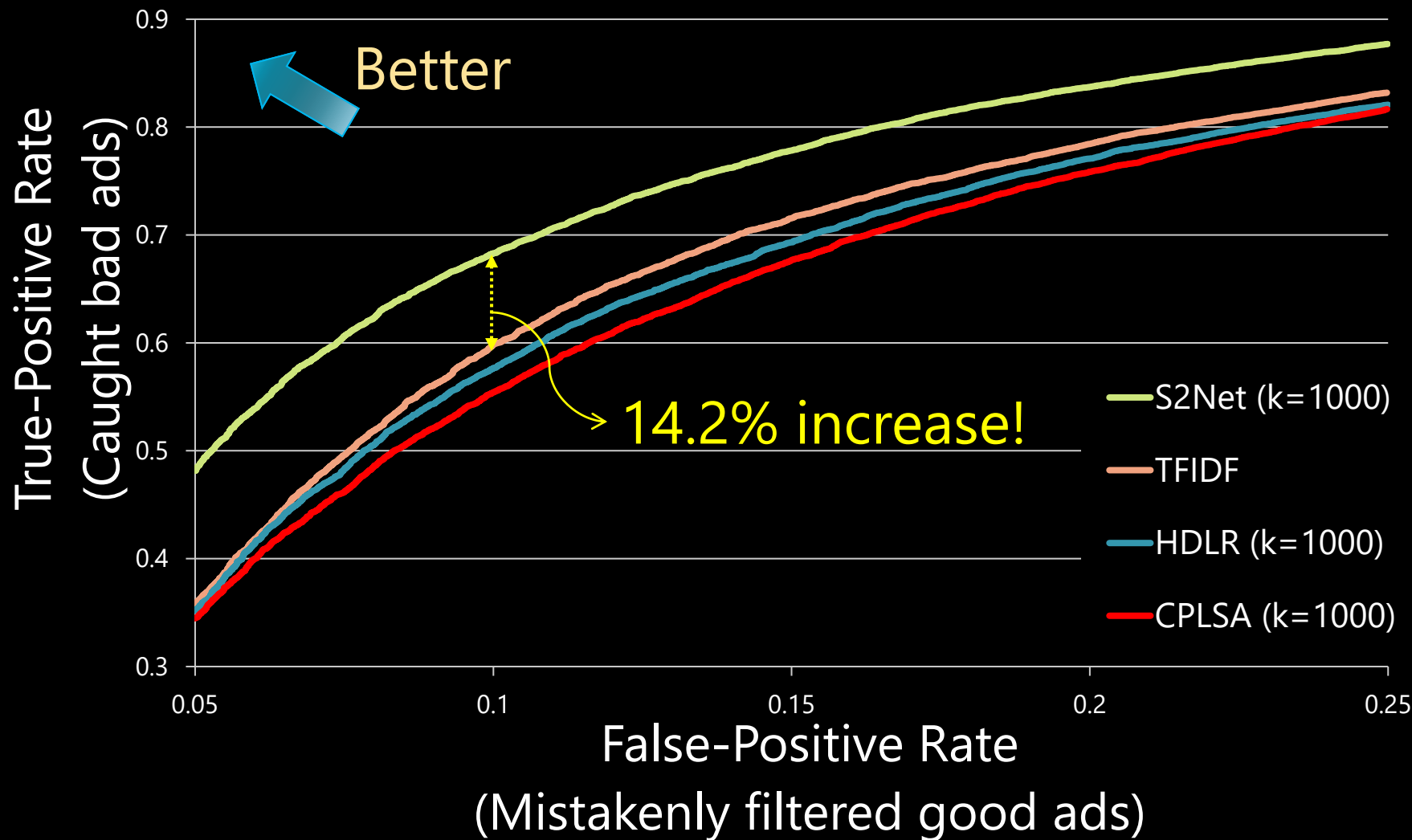
Results on Wikipedia Documents



Ad Relevance Measures

- Task: Decide whether a paid-search ad is relevant to the query
 - Filter irrelevant ads to ensure positive search experience
 - F_{qry} : pseudo-document from Web relevance feedback
 - F_{doc} : ad landing page
- Data: query-ad human relevance judgment
 - Training: 226k pairs
 - Validation: 169k pairs
 - Testing: 169k pairs

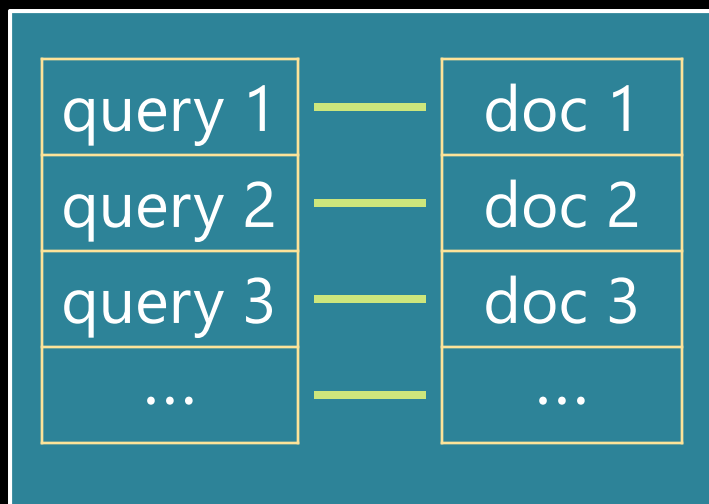
The ROC Curves of the Ad Filters



Web Search Ranking [Gao et al., SIGIR-11]

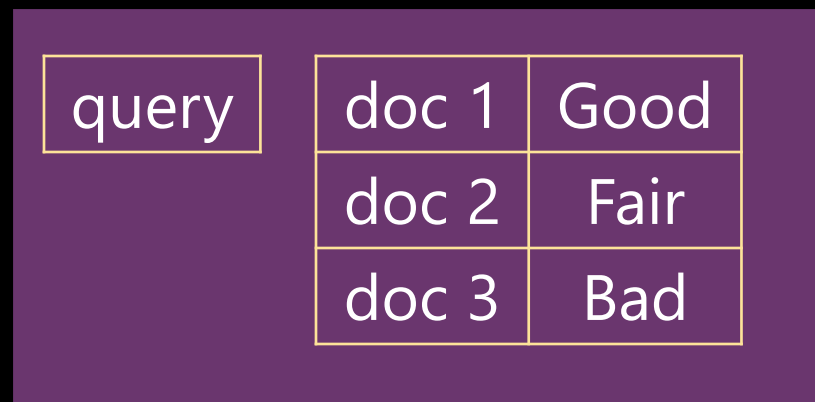
Train latent semantic models

- *Parallel corpus* from clicks
- 82,834,648 query-doc pairs

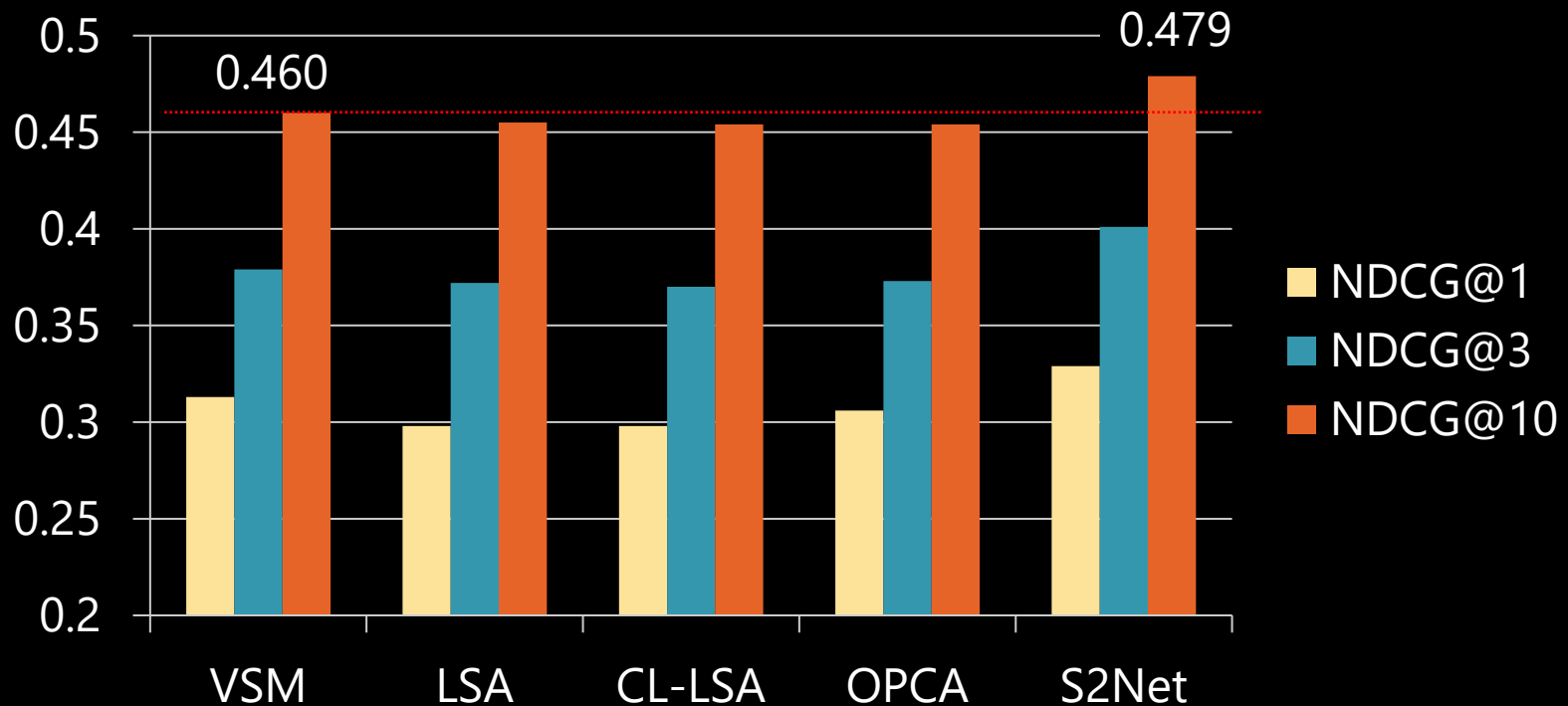


Evaluate using labeled data

- Human relevance judgment
- 16,510 queries
- 15 doc per query in average

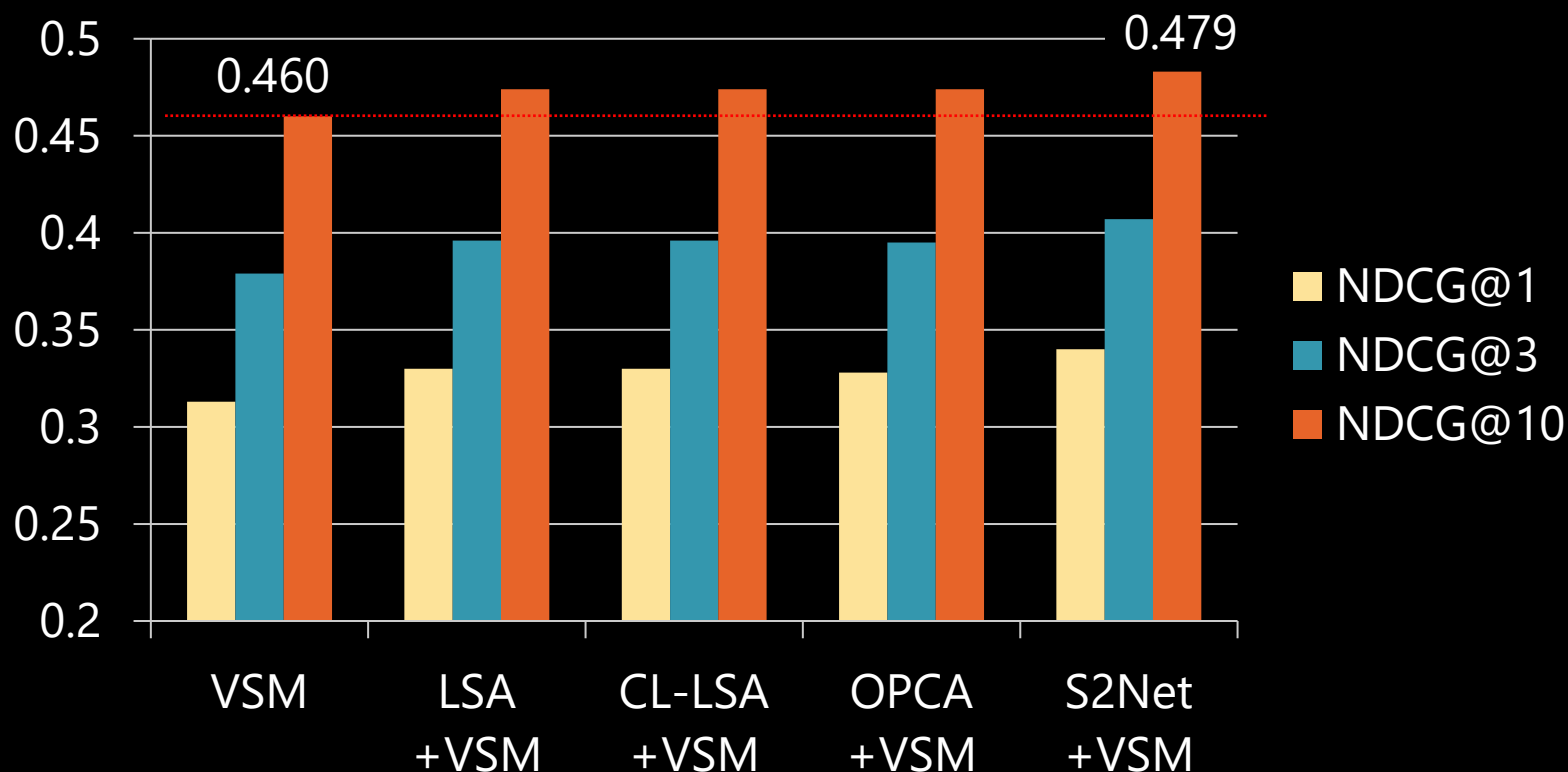


Results on Web Search Ranking



- Only S2Net outperforms VSM compared to other projection models

Results on Web Search Ranking



- After combined with VSM, results are all improved
- More details and interesting results of generative topic models can be found in [SIGIR-11]

Outline

- Introduction
- Problem & Approach
- Experiments
 - Cross-language document retrieval
 - Ad relevance measures
 - Web search ranking
- **Discussion & Conclusions**

Model Comparisons

- S2Net vs. generative topic models
 - Can handle explicit negative examples
 - No special constraints on input vectors
- S2Net vs. linear projection methods
 - Loss function designed to closely match the true objective
 - Computationally more expensive
- S2Net vs. metric learning
 - Target high-dimensional input space
 - Scale well as the number of examples increases

Why Does S2Net Outperform Other Methods?

- Loss function
 - Closer to the true evaluation objective
- Slight nonlinearity
 - Cosine instead of inner-product
- Leverage a large amount of training data
 - Easily parallelizable: distributed gradient computation

Conclusions

- S2Net: Discriminative learning framework for dimensionality reduction
 - Learns a good projection matrix that leads to robust text similarity measures
 - Strong empirical results on different tasks
- Future work
 - Model improvement
 - Handle Web-scale parallel corpus more efficiently
 - Convex loss function
 - Explore more applications
 - e.g., word/phrase similarity