Modeling Interestingness with Deep Neural Networks

Jianfeng Gao, Patrick Pantel, Michael Gamon, Xiaodong He, Li Deng, Yelong Shen

Presented by Scott Wen-tau Yih

Microsoft Research (Redmond, USA)

Computing Semantic Similarity

- Fundamental to almost all NLP tasks, e.g.,
  - Machine translation: similarity between sentences in different languages
  - Web search: similarity between queries and documents

- Problems of the existing approaches
  - Lexical matching cannot handle language discrepancy.
  - Unsupervised word embedding or topic models are not optimal for the task of interest.

Deep Semantic Similarity Model (DSSM)

- *Semantic*: map texts to real-valued vectors in a latent *semantic* space that is language independent
- *Deep*: the mapping is performed via *deep* neural network models that are optimized using a task-specific objective
- *State-of-the-art* results in many NLP tasks (e.g., Shen et al. 2014; Gao et al. 2014, Yih et al. 2014)
- This paper: DSSM to model interestingness for recommendation –

  What interests a user when she is reading a doc?

Outline

- Introduction
- Tasks of modeling Interestingness
  - Automatic highlighting
  - Contextual entity search
- A Deep Semantic Similarity Model (DSSM)
- Experiments
- Conclusions

Two Tasks of Modeling Interestingness

- Automatic highlighting
  - Highlight the key phrases which represent the entities (person/loc/org) that interest a user when reading a document
  - Doc semantics influences what is perceived as interesting to the user
  - e.g., article about movie → articles about an actor/character

- Contextual entity search
  - Given the highlighted key phrases, recommend new, interesting documents by searching the Web for supplementary information about the entities
  - A key phrase may refer to different entities; need to use the contextual information to disambiguate

The Einstein Theory of Relativity

(1) The perihelion of Mercury shows a discrepancy which has long puzzled astronomers. This discrepancy is fully accounted for by Einstein. At the time when he published his theory, this was its only experimental verification.

(2) Modern physicists were willing to suppose that light might be subject to gravitation—i.e., that a ray of light passing near a great mass like the sun might be deflected to the extent to which a particle moving with the same velocity would be deflected according to the orthodox theory of gravitation. But Einstein's theory required that the light should be deflected just twice as much as this. The matter could only be tested during an eclipse among a number of bright stars. Fortunately a peculiarly favourable eclipse occurred last year. The results of the observations

The Einstein Theory of Relativity

(1) The perihelion of Mercury shows a discrepancy which has long puzzled astronomers. This discrepancy is fully accounted for by Einstein. At the time when he published his theory, this was its only experimental verification.

(2) Modern physicists were willing to suppose that light might be subject to gravitation—i.e., that a ray of light passing near a great mass like the sun might be deflected to the extent to which a particle moving with the same velocity would be deflected according to the orthodox theory of gravitation. But Einstein's theory required that the light should be deflected just twice as much as this. The matter could only be tested during an eclipse among a number of bright stars. Fortunately a peculiarly favourable eclipse occurred last year. The results of the observations

The Einstein Theory of Relativity

(1) The perihelion of Mercury shows a discrepancy which has long puzzled astronomers. This discrepancy is fully accounted for by Einstein. At the time when he published his theory, this was its only experimental verification.

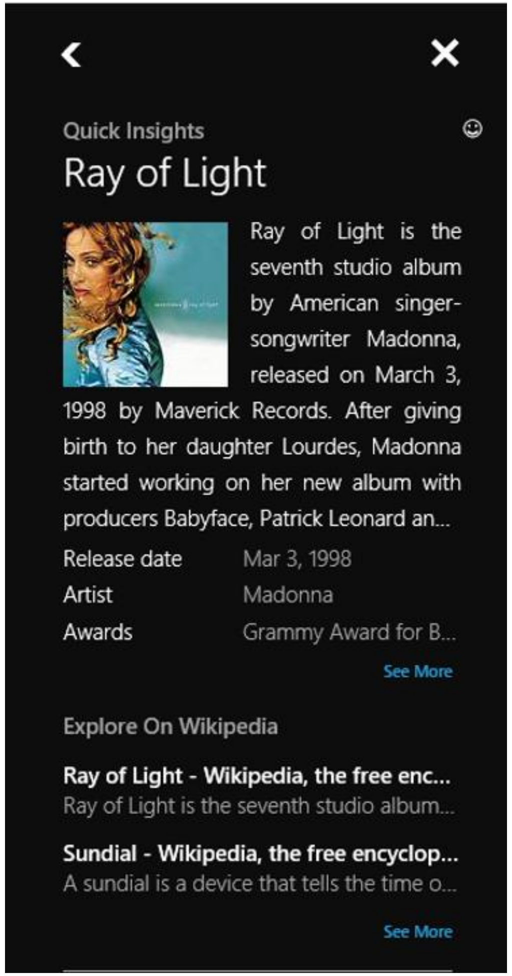
(2) Modern physicists were willing to suppose that light might be subject to gravitation—i.e., that a ray of light passing near a great mass like the sun might be deflected to the extent to which a particle moving with the same velocity would be deflected according to the orthodox theory of gravitation. But Einstein's theory required that the light should be deflected just twice as much as this. The matter could only be tested during an eclipse among a number of bright stars. Fortunately a peculiarly favourable eclipse occurred last year. The results of the observations

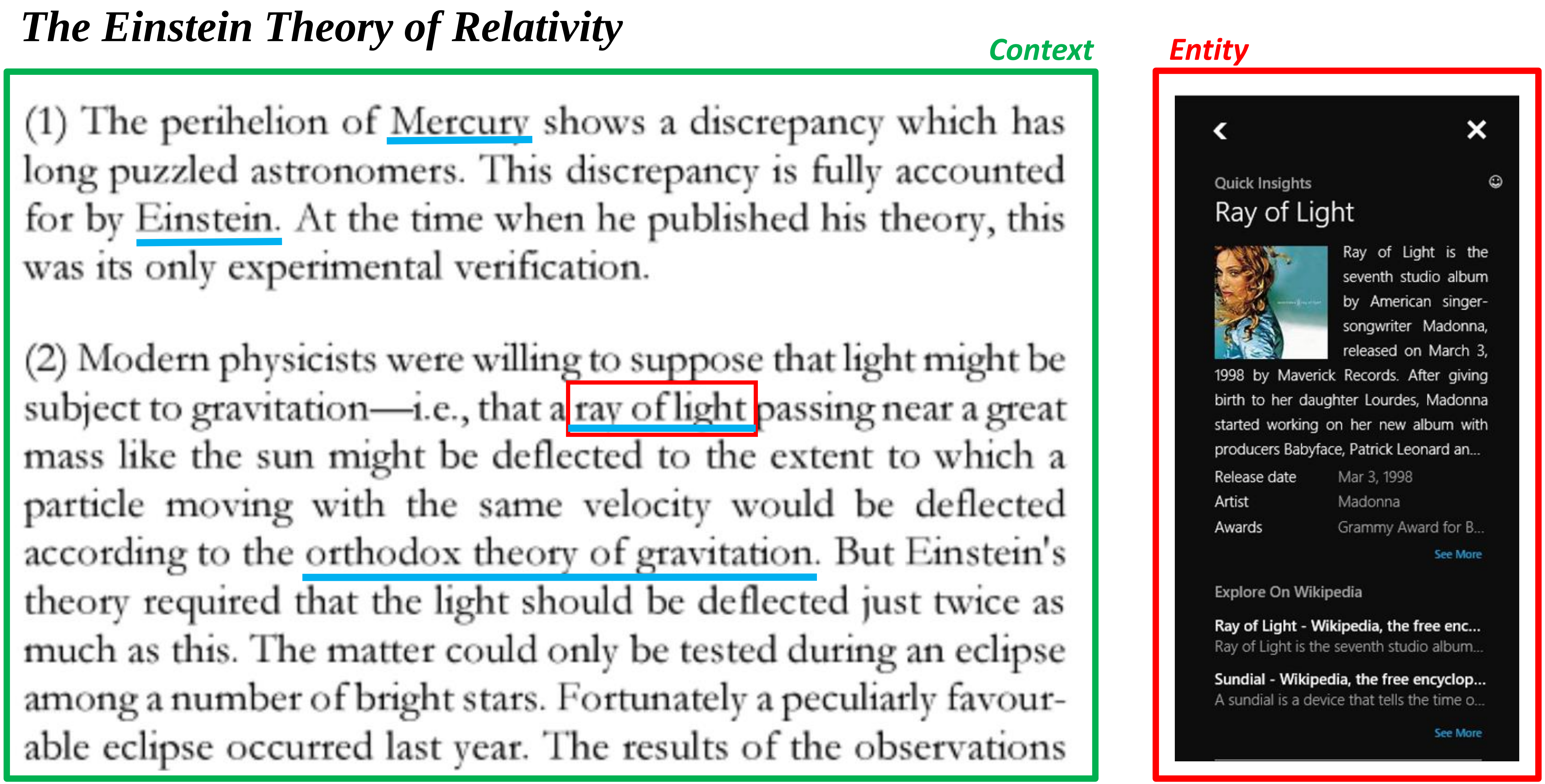
The Einstein Theory of Relativity

(1) The perihelion of Mercury shows a discrepancy which has long puzzled astronomers. This discrepancy is fully accounted for by Einstein. At the time when he published his theory, this was its only experimental verification.

(2) Modern physicists were willing to suppose that light might be subject to gravitation—i.e., that a ray of light passing near a great mass like the sun might be deflected to the extent to which a particle moving with the same velocity would be deflected according to the orthodox theory of gravitation. But Einstein's theory required that the light should be deflected just twice as much as this. The matter could only be tested during an eclipse among a number of bright stars. Fortunately a peculiarly favourable eclipse occurred last year. The results of the observations

Entity




The Einstein Theory of Relativity
Context
Entity
(1) The perihelion of Mercury shows a discrepancy which has long puzzled astronomers. This discrepancy is fully accounted for by Einstein. At the time when he published his theory, this was its only experimental verification.
(2) Modern physicists were willing to suppose that light might be subject to gravitation—i.e., that a ray of light passing near a great mass like the sun might be deflected to the extent to which a particle moving with the same velocity would be deflected according to the orthodox theory of gravitation. But Einstein's theory required that the light should be deflected just twice as much as this. The matter could only be tested during an eclipse among a number of bright stars. Fortunately a peculiarly favourable eclipse occurred last year. The results of the observations
Quick Insights
Ray of Light
Ray of Light is the seventh studio album by American singer-songwriter Madonna, released on March 3, 1998 by Maverick Records. After giving birth to her daughter Lourdes, Madonna started working on her new album with producers Babyface, Patrick Leonard an...
Release date
Mar 3, 1998
Artist
Madonna
Awards
Grammy Award for B...
See More
Explore On Wikipedia
Ray of Light - Wikipedia, the free enc...
Ray of Light is the seventh studio album...
Sundial - Wikipedia, the free encyclop...
A sundial is a device that tells the time o...
See More

The Einstein Theory of Relativity

Context

(1) The perihelion of Mercury shows a discrepancy which has long puzzled astronomers. This discrepancy is fully accounted for by Einstein. At the time when he published his theory, this was its only experimental verification.

(2) Modern physicists were willing to suppose that light might be subject to gravitation—i.e., that a ray of light passing near a great mass like the sun might be deflected to the extent to which a particle moving with the same velocity would be deflected according to the orthodox theory of gravitation. But Einstein's theory required that the light should be deflected just twice as much as this. The matter could only be tested during an eclipse among a number of bright stars. Fortunately a peculiarly favourable eclipse occurred last year. The results of the observations

Entity

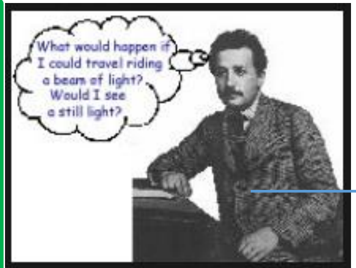
DSSM for Modeling Interestingness

Context

Key phrase

(1) The perihelion of Mercury shows a discrepancy which has long puzzled astronomers. This discrepancy is fully accounted for by Einstein. At the time when he published his theory, this was its only experimental verification.

(2) Modern physicists were willing to suppose that light might be subject to gravitation—i.e., that a ray of light passing near a great mass like the sun might be deflected to the extent to which a particle moving with the same velocity would be deflected according to the orthodox theory of gravitation. But Einstein's theory required that the light should be deflected just twice as much as this. The matter could only be tested during an eclipse among a number of bright stars. Fortunately a peculiarly favourable eclipse occurred last year. The results of the observations



Entity page (reference doc)

Tasks	*X (source text)*	*Y (target text)*
Automatic highlighting	***Doc in reading***	***Key phrases to be highlighted***
Contextual entity search	*Key phrase and context*	*Entity and its corresponding (wiki) page*

DSSM for Modeling Interestingness

Context

Key phrase

(1) The perihelion of Mercury shows a discrepancy which has long puzzled astronomers. This discrepancy is fully accounted for by Einstein. At the time when he published his theory, this was its only experimental verification.

(2) Modern physicists were willing to suppose that light might be subject to gravitation—i.e., that a ray of light passing near a great mass like the sun might be deflected to the extent to which a particle moving with the same velocity would be deflected according to the orthodox theory of gravitation. But Einstein's theory required that the light should be deflected just twice as much as this. The matter could only be tested during an eclipse among a number of bright stars. Fortunately a peculiarly favourable eclipse occurred last year. The results of the observations

Entity page (reference doc)

Tasks	*X (source text)*	*Y (target text)*
Automatic highlighting	*Doc in reading*	*Key phrases to be highlighted*
Contextual entity search	***Key phrase and context***	***Entity and its corresponding (wiki) page***

Outline

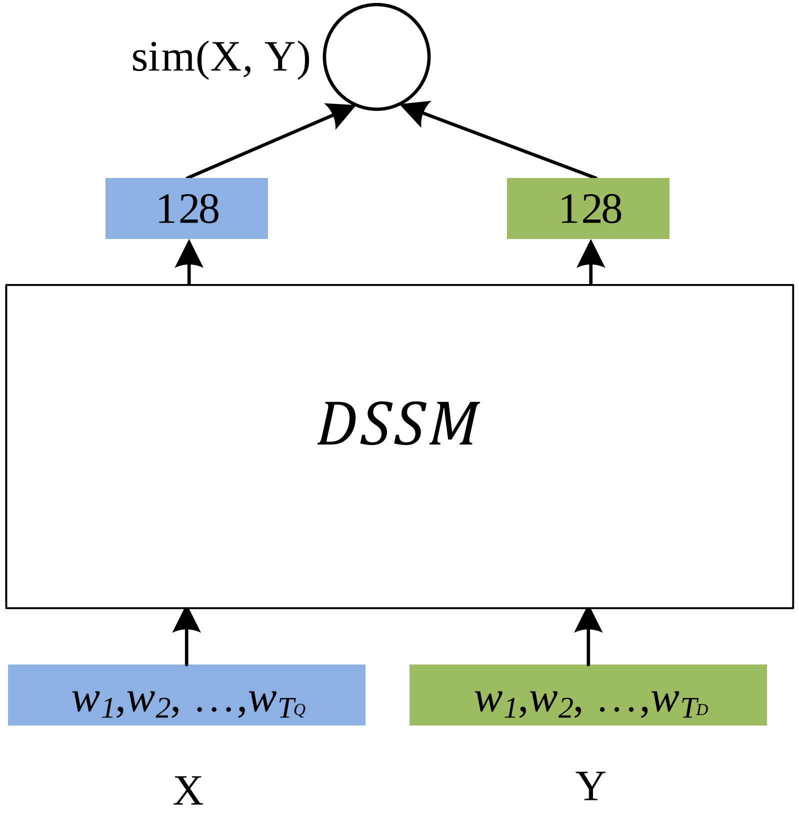
- Introduction
- Tasks of modeling Interestingness
- A Deep Semantic Similarity Model (DSSM)
- Experiments
- Conclusions

DSSM: Compute Similarity in Semantic Space

Relevance measured by cosine similarity

Learning: maximize the similarity between X (source) and Y (target)

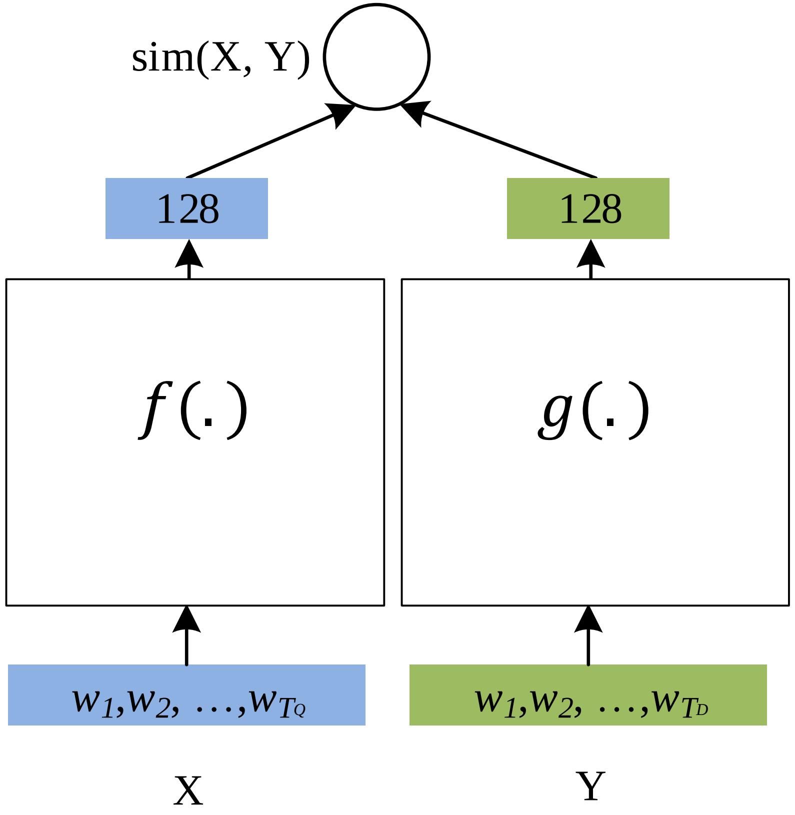
Word sequence x_t



DSSM: Compute Similarity in Semantic Space

Relevance measured by cosine similarity

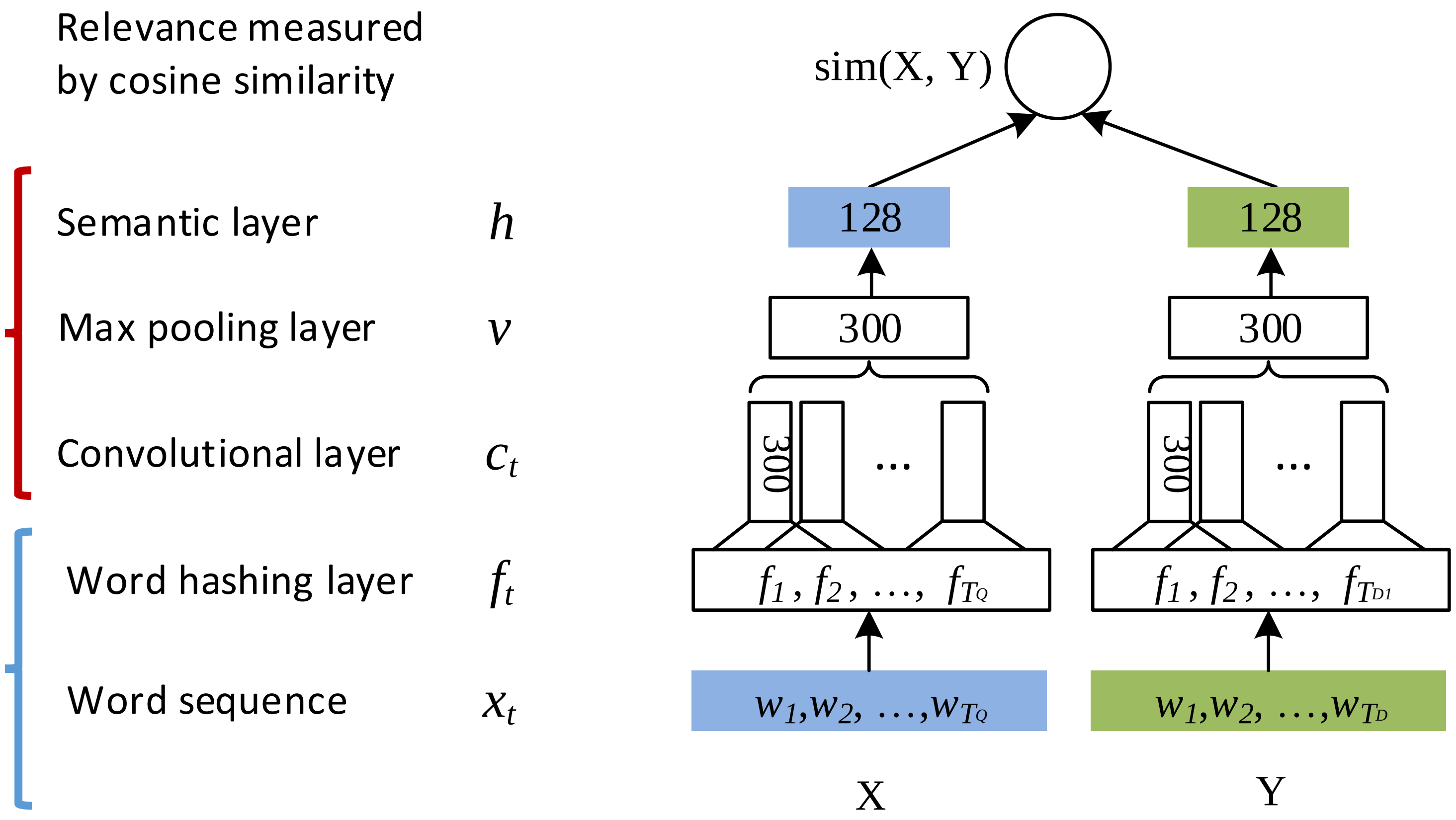
Word sequence x_t



Learning: maximize the similarity between X (source) and Y (target)

Representation: use DNN to extract abstract semantic representations

DSSM: Compute Similarity in Semantic Space



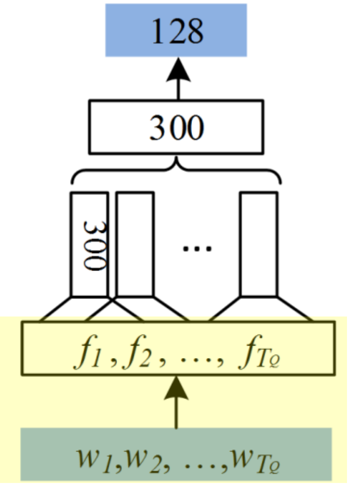
Learning: maximize the similarity between X (source) and Y (target)

Representation: use DNN to extract abstract semantic representations

Convolutional and Max-pooling layer: identify key words/concepts in X and Y

Word hashing: use sub-word unit (e.g., letter n-gram) as raw input to handle very large vocabulary

Letter-trigram Representation

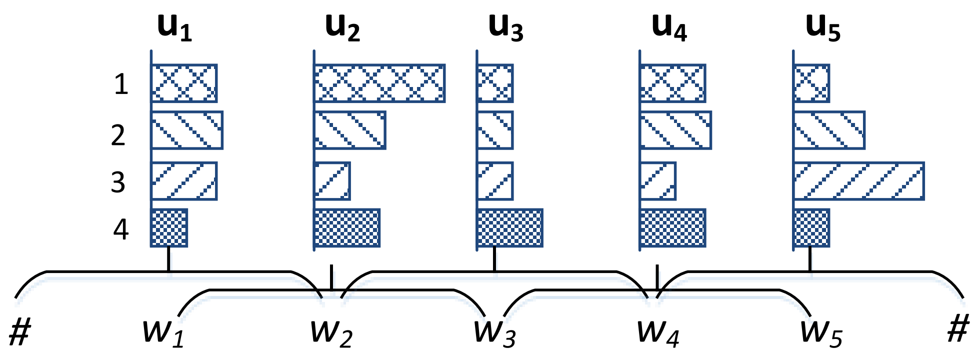


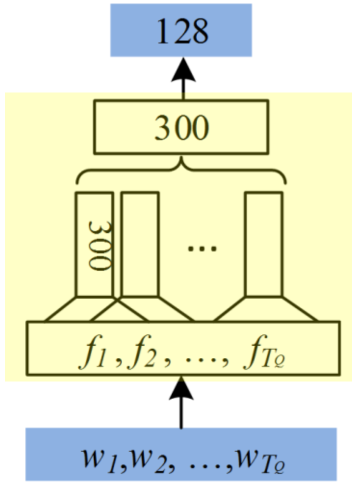
- Control the dimensionality of the input space
  - e.g., cat → #cat# → #-c-a, c-a-t, a-t-#
  - Only ~50K letter-trigrams in English; no OOV issue
- Capture sub-word semantics (e.g., prefix & suffix)
- Words with small typos have similar raw representations

- Collision: different words with same letter-trigram representation?

Vocabulary size	# of unique letter-trigrams	# of Collisions	Collision rate
40K	10,306	2	0.0050%
500K	30,621	22	0.0044%
5M	49,292	179	0.0036%

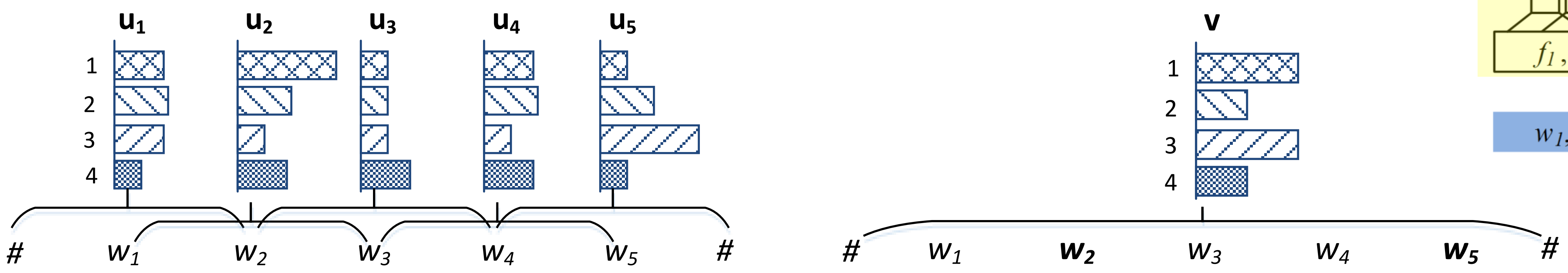
Convolutional Layer

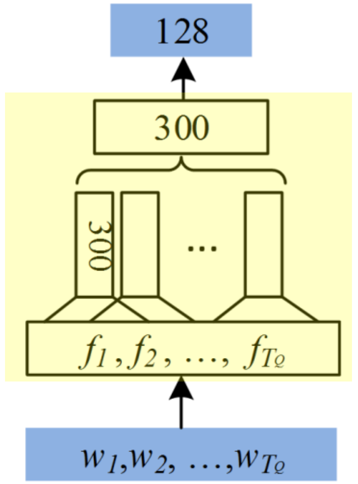




- Extract local features using convolutional layer
  - {w1, w2, w3} → topic 1
  - {w2, w3, w4} → topic 4

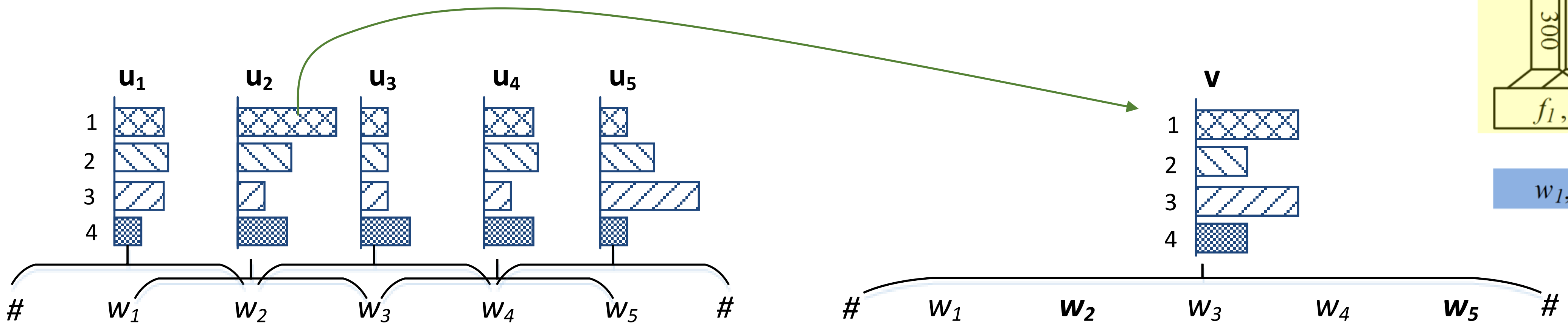
Max-pooling Layer

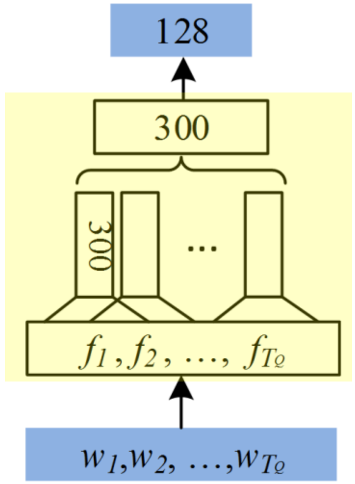




- Extract local features using convolutional layer
  - {w1, w2, w3} → topic 1
  - {w2, w3, w4} → topic 4
- Generate global features using max-pooling
  - Key topics of the text → topics 1 and 3
  - keywords of the text: w2 and w5

Max-pooling Layer





- Extract local features using convolutional layer
  - {w1, w2, w3} → topic 1
  - {w2, w3, w4} → topic 4
- Generate global features using max-pooling
  - Key topics of the text → topics 1 and 3
  - keywords of the text: w2 and w5

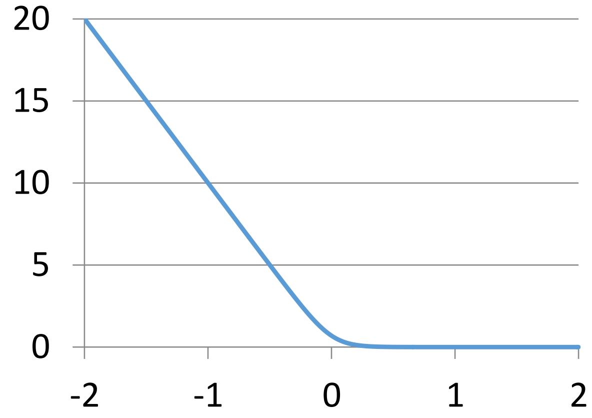
Learning DSSM from Labeled X-Y Pairs

- Consider a doc X and two key phrases Y^+ and Y^-
  - Assume Y^+ is more interesting than Y^- to a user when reading X
- $\mathrm{sim}_{\boldsymbol{\theta}}(X, Y)$ is the cosine similarity of X and Y in semantic space, mapped by DSSM parameterized by $\boldsymbol{\theta}$

Learning DSSM from Labeled X-Y Pairs

- Consider a doc X and two key phrases Y^+ and Y^-
  - Assume Y^+ is more interesting than Y^- to a user when reading X
- $\mathrm{sim}_{\boldsymbol{\theta}}(X, Y)$ is the cosine similarity of X and Y in semantic space, mapped by DSSM parameterized by $\boldsymbol{\theta}$

- $\Delta = \mathrm{sim}_{\boldsymbol{\theta}}(X, Y^+) - \mathrm{sim}_{\boldsymbol{\theta}}(X, Y^-)$
  - We want to maximize Δ
- $Loss(\Delta; \boldsymbol{\theta}) = \log(1 + \exp(-\gamma\Delta))$
- Optimize $\boldsymbol{\theta}$ using mini-batch SGD on GPU



Outline

- Introduction
- Tasks of modeling Interestingness
- A Deep Semantic Similarity Model (DSSM)
- Experiments – Two Tasks of Modeling Interestingness
    - Data & Evaluation
    - Results
- Conclusions

Extract Labeled Pairs from Web Browsing Logs
Automatic Highlighting

- When reading a page P, the user *clicks* a hyperlink H

- (text in P, anchor text of H)

Extract Labeled Pairs from Web Browsing Logs
Contextual Entity Search

- When a hyperlink H points to a Wikipedia P'

http://runningmoron.blogspot.in/

…

I spent a lot of time finding music that was motivating and that I'd also want to listen to through my phone. I could find none. None! I wound up downloading three Metallica songs, a Judas Priest song and one from Bush.

…

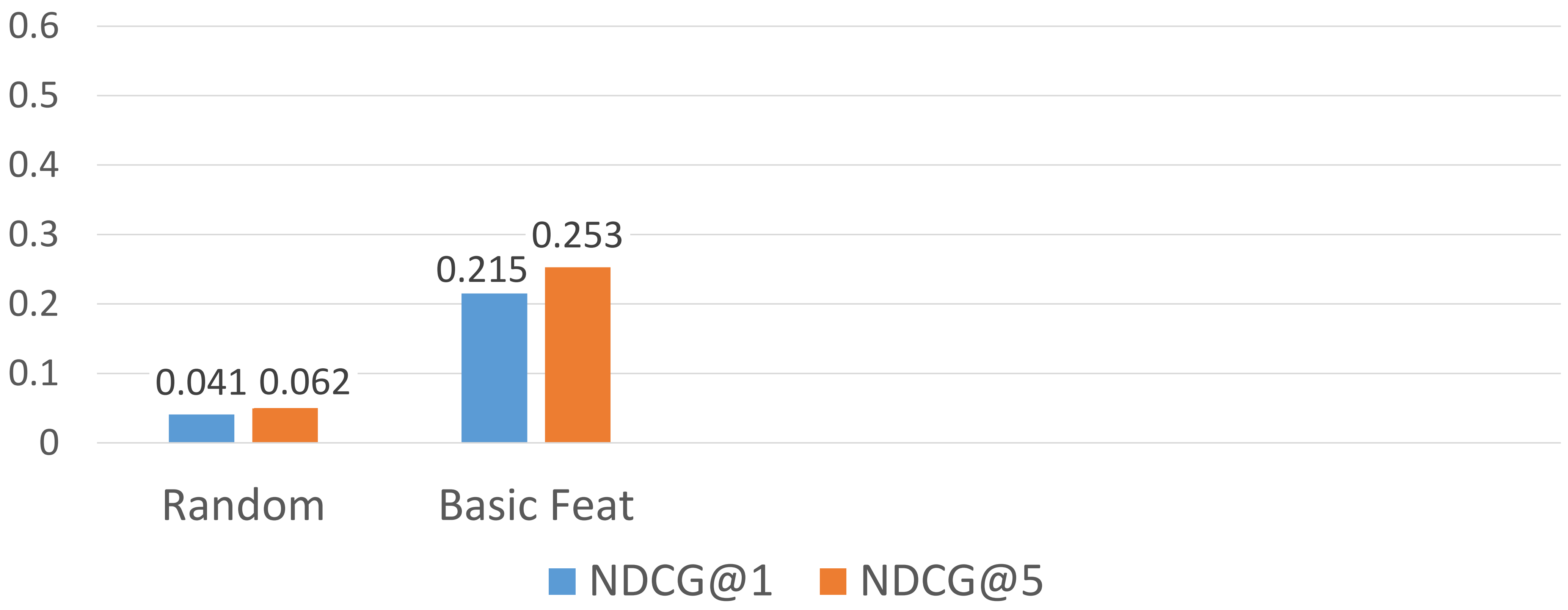
http://en.wikipedia.org/wiki/Bush_(band)

- (anchor text of H & surrounding words, text in P')

Automatic Highlighting: Settings

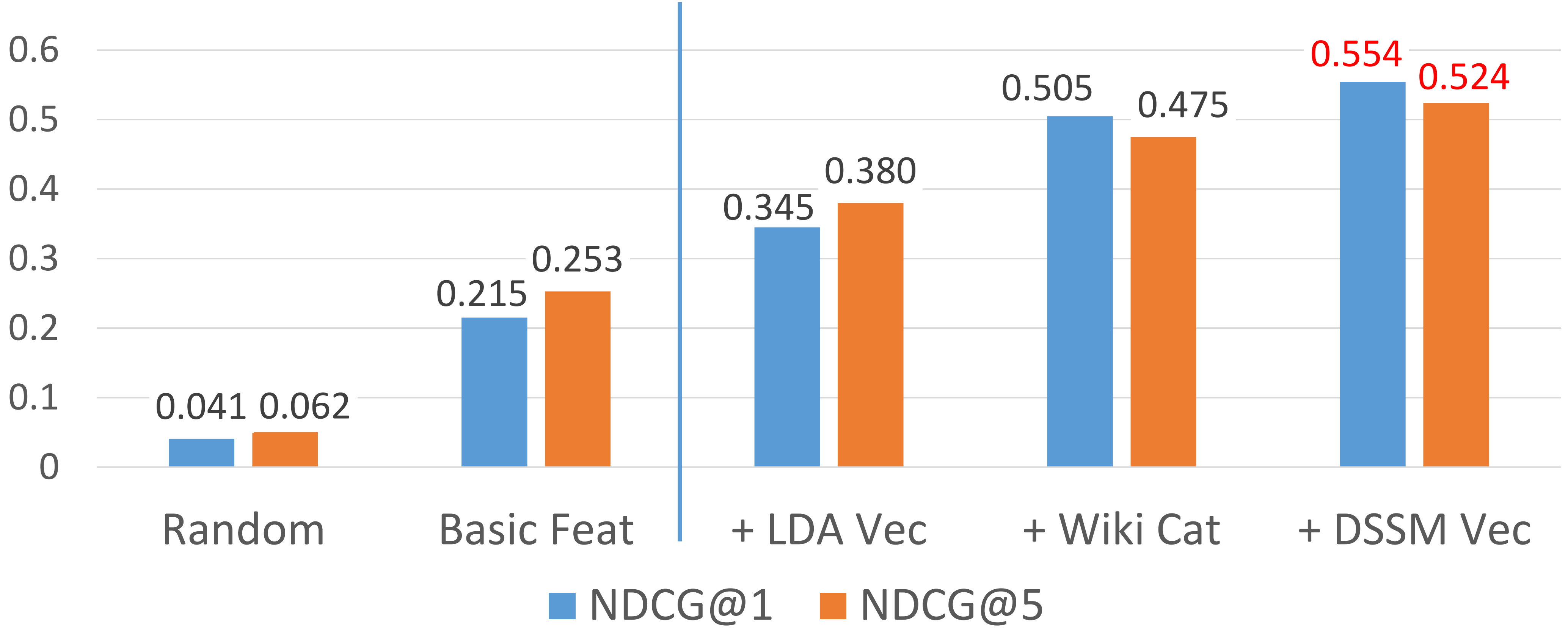
- Simulation
  - Use a set of anchors as candidate key phrases to be highlighted
  - Gold standard rank of key phrases – determined by # user clicks
  - Model picks top-k keywords from the candidates
  - Evaluation metric: NDCG
- Data
  - 18 million occurrences of user clicks from a Wiki page to another, collected from 1-year Web browsing logs
  - 60/20/20 split for training/validation/evaluation

Automatic Highlighting Results: Baselines



- Random: Random baseline
- Basic Feat: Boosted decision tree learner with document features, such as anchor position, freq. of anchor, anchor density, etc.

Automatic Highlighting Results: Semantic Features

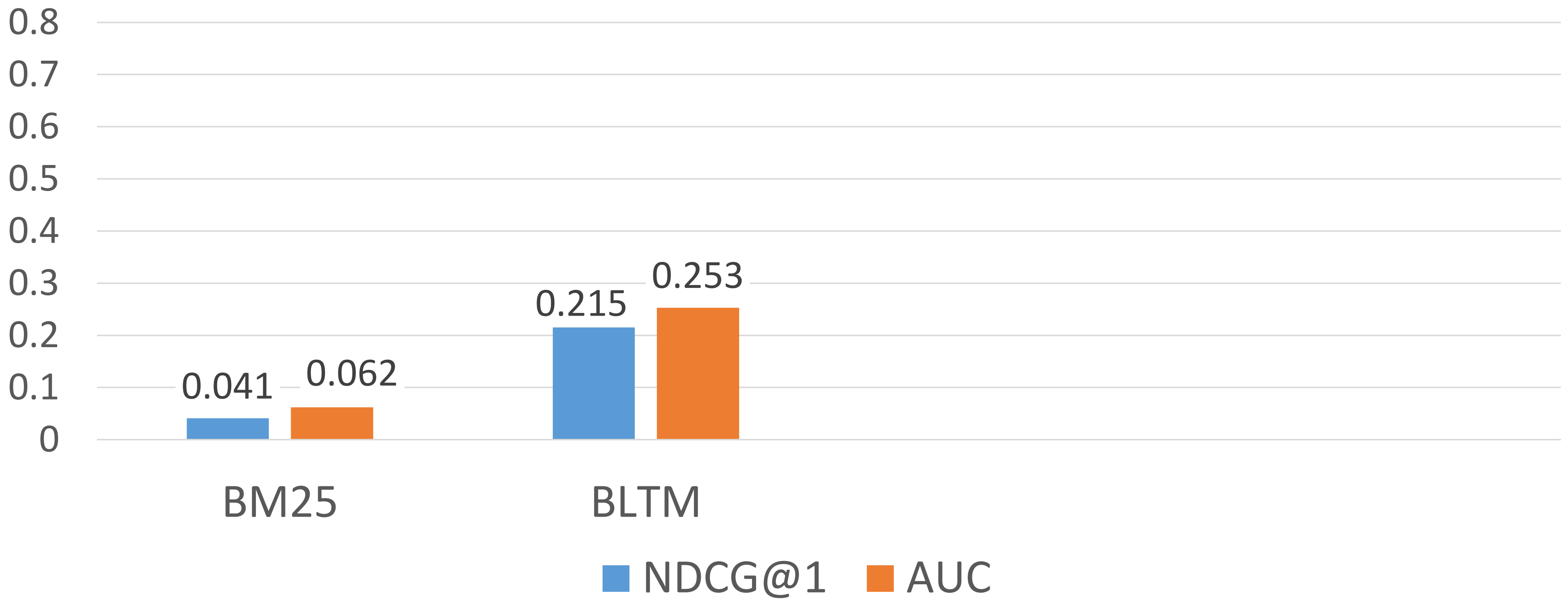


- + LDA Vec: Basic + Topic model (LDA) vectors [Gamon+ 2013]
- + Wiki Cat: Basic + Wikipedia categories (do not apply to general documents)
- + DSSM Vec: Basic + DSSM vectors

Contextual Entity Search: Settings

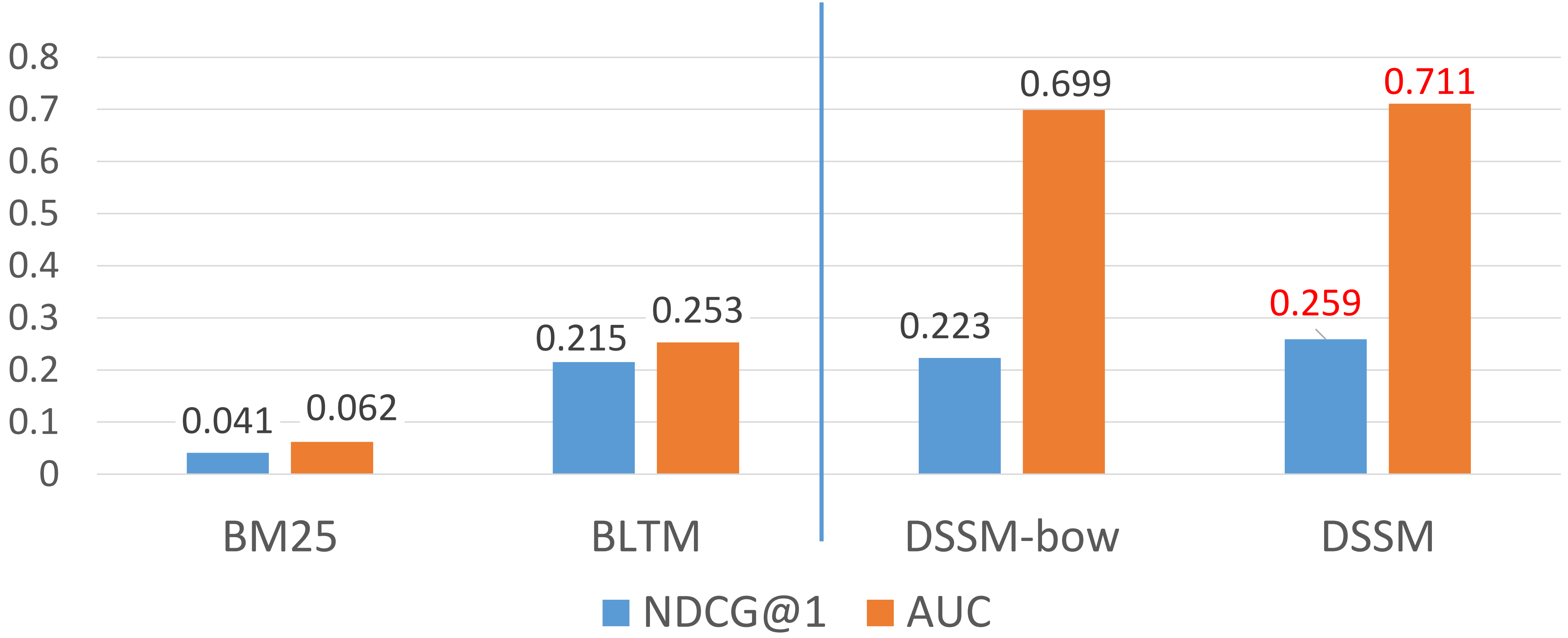
- Training/validation data: same as in *automatic highlighting*
- Evaluation data
  - Sample 10k Web documents as the source documents
  - Use named entities in the doc as query; retain up to 100 returned documents as target documents
  - Manually label whether each target document is a good page describing the entity
  - 870k labeled pairs in total
- Evaluation metric: NDCG and AUC

Contextual Entity Search Results: Baselines



- BM25: The classical document model in IR [Robertson+ 1994]
- BLTM: Bilingual Topic Model [Gao+ 2011]

Contextual Entity Search Results: DSSM

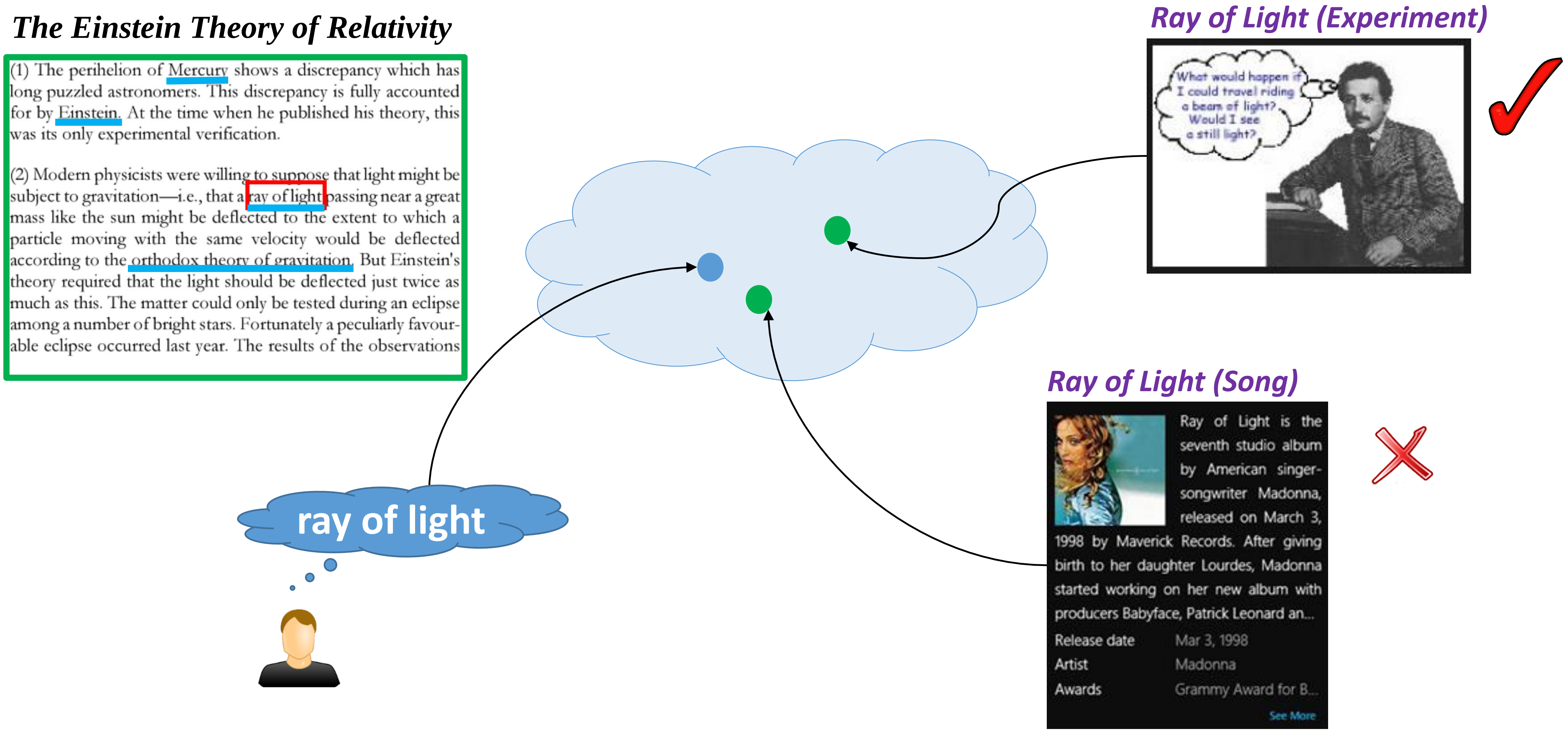


- DSSM-bow: DSSM without convolutional layer and max-pooling structure
- DSSM outperforms classic doc model and state-of-the-art topic model

Conclusions

- Modeling interestingness for recommendation –
  What interests a user when she is reading a doc?
- Deep Semantic Similarity Model (DSSM)
  - *Semantic*: map texts to feature vectors in a latent *semantic* space that is language independent
  - *Deep*: the mapping is performed via *deep* neural network models that are optimized using a task-specific objective
  - *Best* results in modeling interestingness (and other NLP tasks)
- Future work
  - Improve DSSM by incorporating more structure information
  - Apply DSSM to more applications

Learning DSSM from Labeled X-Y Pairs



Learning DSSM from Labeled X-Y Pairs