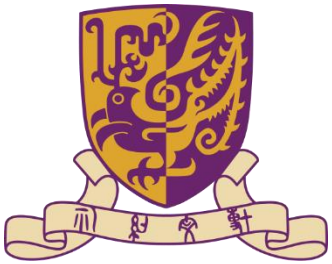
Deep Summarizer

Piji Li

pjli@se.cuhk.edu.hk



Outline

- Introduction
- Related Work
- Inspiration

Outline

- **Introduction**
- Related Work
- Inspiration

Introduction

- **Automatic Summarization**
  - The process of reducing a text document with a computer program in order to create a summary that retains the most important points of the original document.
  - Summaries pervade internet search results, restaurant and consumer product review sites and the front pages of news outlets.
  - Single/Multi – Document Summarization

- **Methods**
  - Extraction, Compression, Abstraction

- **Deep Summarizer**
  - Summarization using Deep Learning

Introduction

- **Two Standard MDS Data Sets**
  - DUC 2004 ~ 2007
  - TAC 2008 ~ 2011
  - 50 topics, 10 news, 4 model summaries
  - DUC<=250 words, TAC<=100 words

- **Metrics**
  - ROUGE (R-1, R-2, R-SU4)
    - http://www.berouge.com/
  - Pyramid
    - *Evaluating Content Selection in Summarization: the Pyramid Method* NAACL-HLT 2004

Outline

- Introduction
- **Related Work**
- Inspiration

MDS – Deep Belief Network

- Liu, Yan, Sheng-hua Zhong, and Wenjie Li. "**Query-Oriented Multi-Document Summarization via Unsupervised Deep Learning**." In *AAAI*. 2012.

  - DBN
  - Query related weight

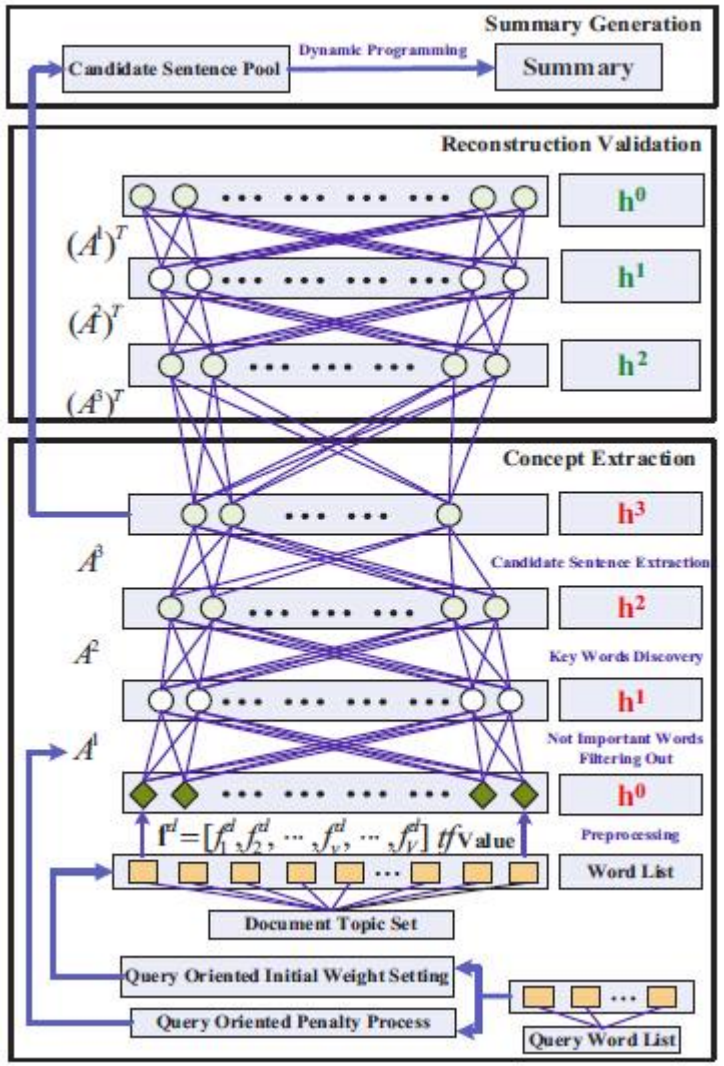
$$A_{ij}^1 = \max(A^1) \quad \text{if } v_i \in \mathbf{q}$$

$$\Delta A_{ij}^1 = \gamma \Delta A_{ij}^1 \quad \text{if } v_i \in \mathbf{q}$$

  - Important words extraction

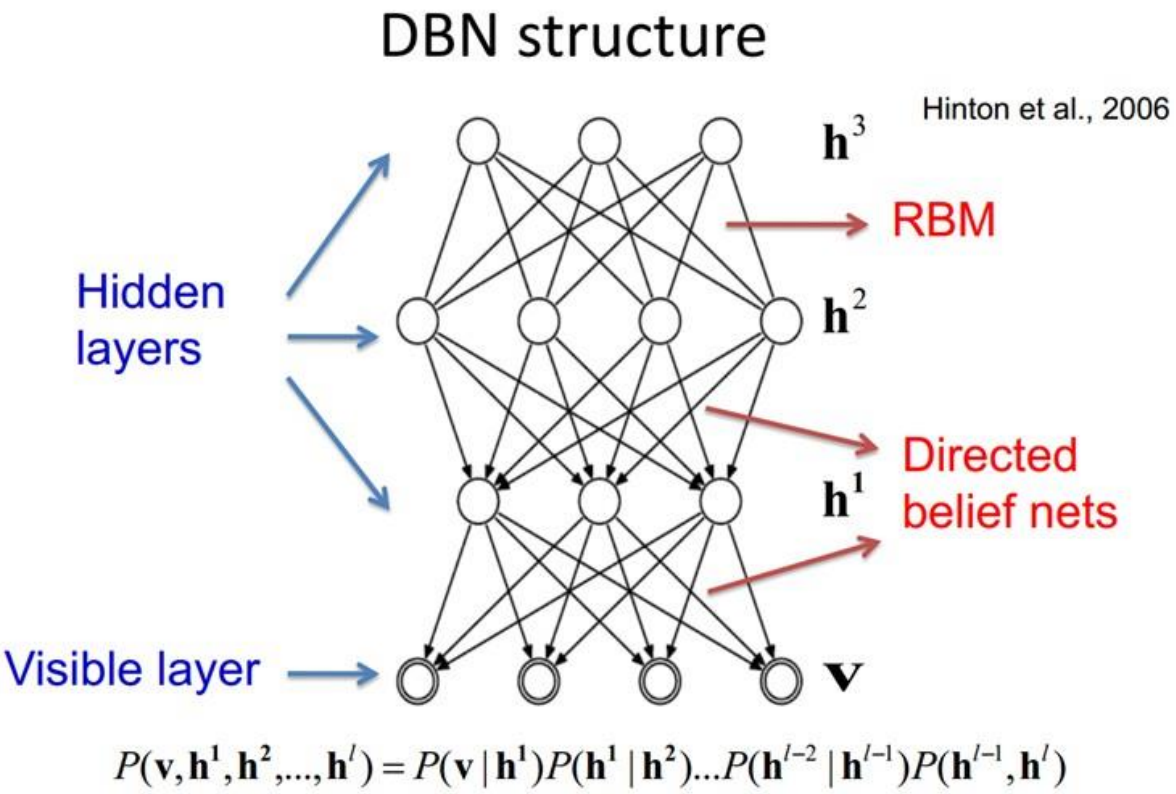
$$AF = [\underbrace{(\mathbf{f}^{D_1})^T, (\mathbf{f}^{D_2})^T, \cdots, (\mathbf{f}^{D_j})^T, \cdots, \mathbf{f}^{D})^T}_{K_3}](A^1 A^2 A^3)$$

  - Diverse Sentence Selection
    Intersection with important words



DBN

Hinton, Geoffrey, Simon Osindero, and Yee-Whye Teh. "**A fast learning algorithm for deep belief nets.**" *Neural computation* 18, no. 7 (2006): 1527-1554.



MDS – Deep Belief Network

- Liu, Yan, Sheng-hua Zhong, and Wenjie Li. "**Query-Oriented Multi-Document Summarization via Unsupervised Deep Learning**." In *AAAI*. 2012.
- **Results**

Table 3. Comparison to representative algorithms on the DUC 2006

System	ROUGE-1	ROUGE-2	ROUGE-SU4
QODE	0.4015	**0.0928***	0.1479
Manifold-ranking	**0.4101***	0.0886	0.1420
Multiple-modality	0.4031	0.0851	0.1400
Document-sensitive		0.0899	0.1427
SVM Classification		0.0834	0.1387
Ranking SVM		0.0890	0.1443
Regression		0.0926	**0.1485***
NIST Baseline		0.0491	0.0962

Table 4. Comparison to representative algorithms on the DUC 2007

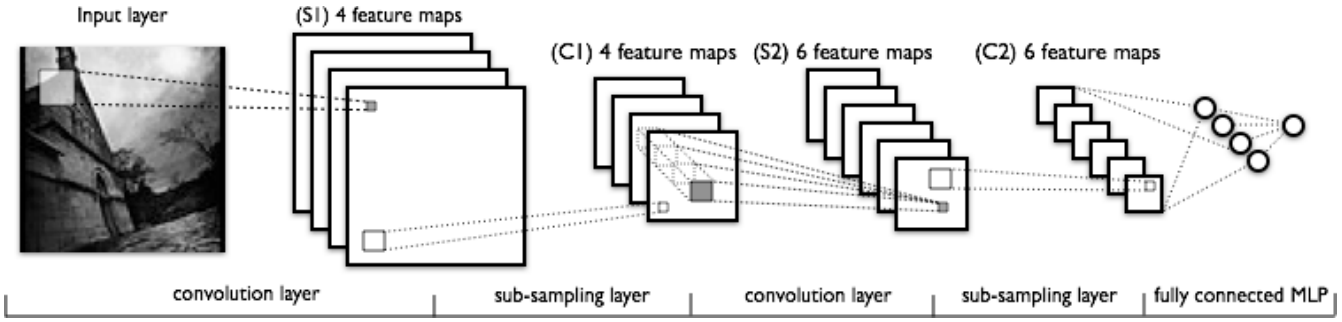
System	ROUGE-1	ROUGE-2	ROUGE-SU4
QODE	0.4295	0.1163	**0.1685***
Manifold-ranking	0.4204	0.1030	0.1460
Multiple-modality		0.1123	0.1682
Document-sensitive	0.4211	0.1103	0.1628
SVM Classification		0.1075	0.1616
Ranking SVM	**0.4301***	**0.1175***	0.1682
NIST Baseline	0.3091	0.0599	0.1036

MDS – Deep Belief Network

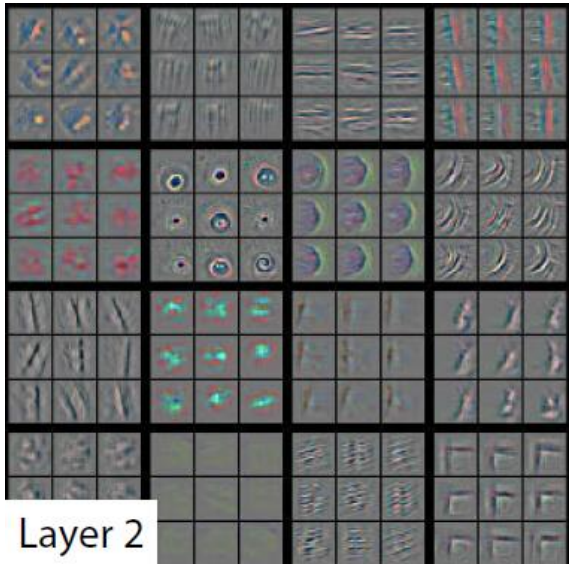
- Liu, Yan, Sheng-hua Zhong, and Wenjie Li. "**Query-Oriented Multi-Document Summarization via Unsupervised Deep Learning**." In *AAAI*. 2012.
- **Strengthens**
  - DBN can extract the abstract concept layer by layer
  - Significant ROUGE results
- **Weakness**
  - Only focus on word level
  - Need more works on sentences

Salience Sentences Detection using CNN

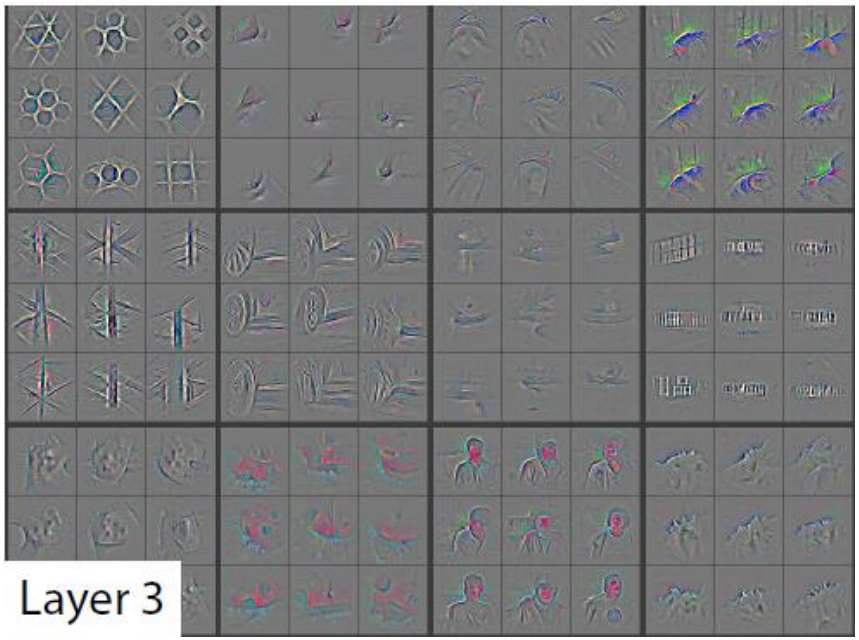
- Denil, Misha, Alban Demiraj, and Nando de Freitas. "**Extraction of Salient Sentences from Labelled Documents.**" ICLR (2014).



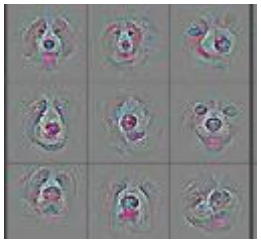
Layer1



Layer 2



Layer 3

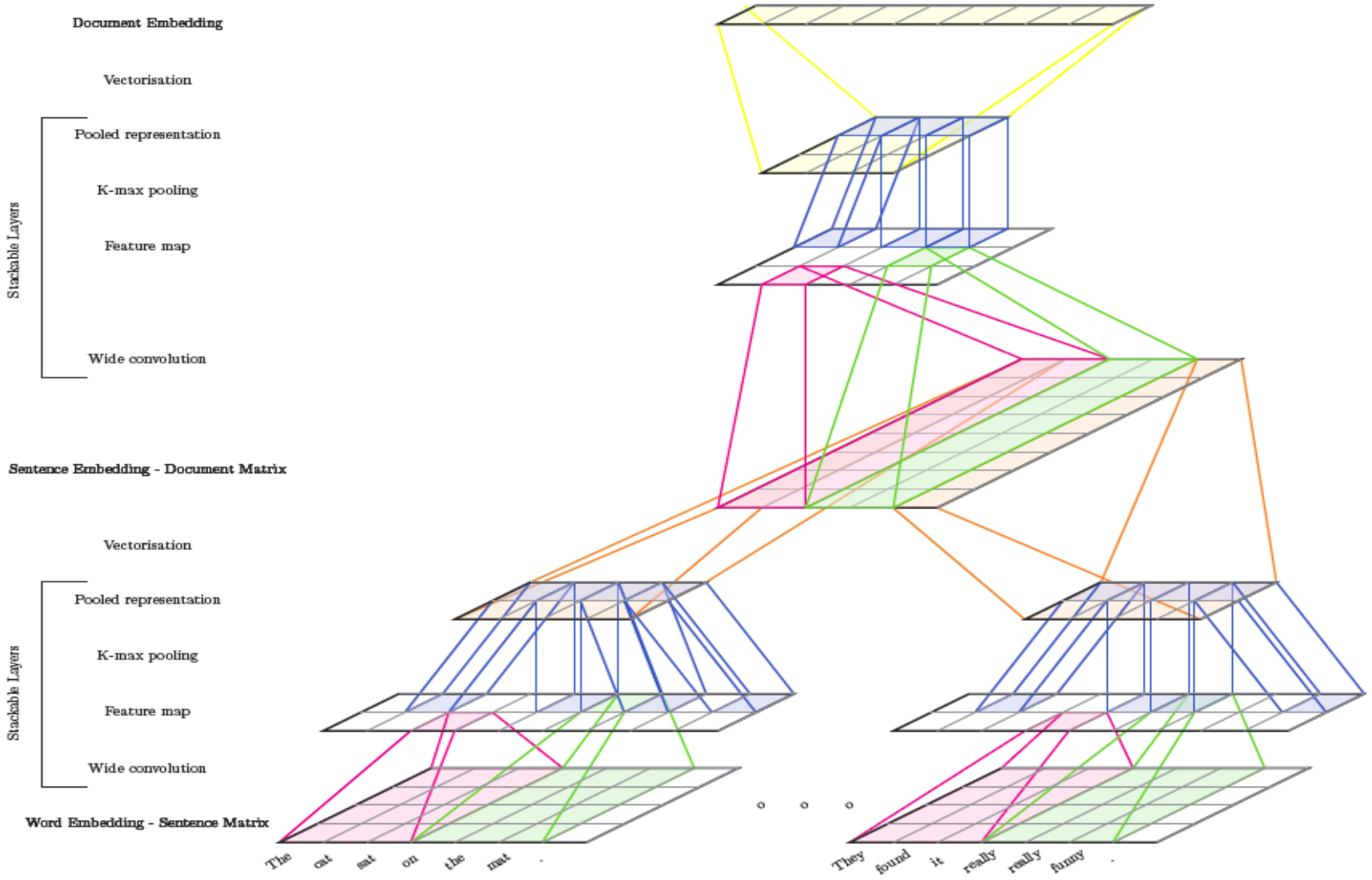


Layer4

Layer5

Zeiler, Matthew D., and Rob Fergus. "**Visualizing and understanding convolutional networks**." ECCV 2014

Salience Detection using CNN



Kalchbrenner, Nal, Edward Grefenstette, and Phil Blunsom. "**A convolutional neural network for modelling sentences.**" arXiv:1404.2188 (2014)

Salience Detection using CNN

- Denil, Misha, Alban Demiraj, and Nando de Freitas. "**Extraction of Salient Sentences from Labelled Documents**." ICLR (2014).
- **Data set**
  - IMDB movie review sentiment data set
- **Evaluation**
  - Train a model an the full document data set
  - Test each sample only using the summaries.

Proportion	ConvNet	Word2Vec	Rand.	Fixed	ConvNet	Word2Vec	Rand.
50%	**82.74**	81.98	79.79	Pick 5	**83.12**	82.26	80.02
33%	**82.72**	80.39	76.72	Pick 4	**82.91**	81.92	79.05
25%	**82.94**	80.18	74.87	Pick 3	**82.59**	81.48	77.15
20%	**82.84**	79.70	73.20	Pick 2	**81.71**	80.39	74.48
Full	83.04						
First+Last	68.62						

Salience Detection using CNN

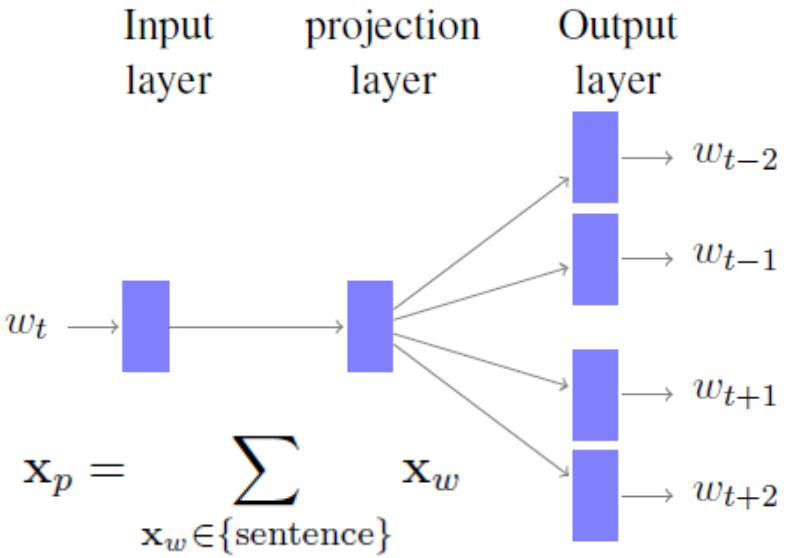
- Denil, Misha, Alban Demiraj, and Nando de Freitas. "**Extraction of Salient Sentences from Labelled Documents**." ICLR (2014).
- **Strengthens**
  - Salience detection using supervised CNN is effective
- **Weakness**
  - Supervised
  - Need fine-grained processing for short-text

Sentences Modelling + Submodular

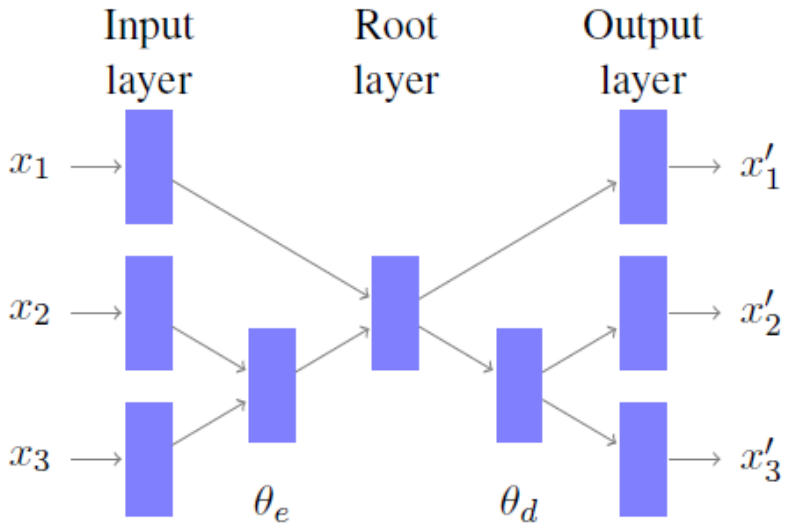
- Kågebäck, Mikael, et al. "**Extractive summarization using continuous vector space models**." EACL 2014.
- **Submodular Optimization**

$$\mathcal{F}(S) = \mathcal{L}(S) + \lambda \mathcal{R}(S) \qquad \mathrm{Sim}(i, j)$$

- **Distributed Sentence Representation**



Word2Vec



Recursive Auto-Encoder

Sentences Modelling + Submodular

- Kågebäck, Mikael, et al. "**Extractive summarization using continuous vector space models**." EACL 2014.
- **Data set**
  - The Opinosis dataset consists of short user reviews in 51 different topics. http://kavita-ganesan.com/opinosis-opinion-dataset
- **Results**

ROUGE-1	R	P	F
OPT_R	**57.86**	21.96	30.28
OPT_F	45.93	48.84	**46.57**
CW_RAE_{Cos}	27.37	19.89	22.00
CW_RAE_{Euc}	29.25	19.77	22.62
CW_Add_{Cos}	**34.72**	11.75	17.16
CW_Add_{Euc}	29.12	**22.75**	**24.88**
$W2V\_Add_{Cos}$	30.86	16.81	20.93
$W2V\_Add_{Euc}$	28.71	16.67	20.75
Original	25.82	19.58	20.57

ROUGE-2	R	P	F
OPT_R	**22.96**	12.31	15.33
OPT_F	20.42	19.94	**19.49**
CW_RAE_{Cos}	4.68	3.18	3.58
CW_RAE_{Euc}	4.82	3.24	3.67
CW_Add_{Cos}	**5.89**	1.81	2.71
CW_Add_{Euc}	5.12	**3.60**	**4.10**
$W2V\_Add_{Cos}$	5.71	3.08	3.82
$W2V\_Add_{Euc}$	3.86	1.95	2.54
Original	3.92	2.50	2.87

Sentences Modelling + Submodular

- Kågebäck, Mikael, et al. "**Extractive summarization using continuous vector space models**." EACL 2014.
- **Strengthens**
  - Experiments using different sentence modelling methods show the performance of the different distributed representation methods
- **Weakness**
  - Lack of novelty

Sentence Modelling + Pagerank

- Wenpeng Yin， Yulong Pei. "**Optimizing Sentence Modeling and Selection for Document Summarization**." IJCAI 2015
- **Method**
  - CNN+Word2Vec to generate distributed representations for sentences
  - Greedy Sentence Selection
- **Strength**
  - Unsupervised CNN via Word2Vec
- **Weakness**
  - The basic MDS framework is still PageRank.
  - Just replace the sentence representation part using a distributed representation method.
  - The novelty and contribution are limit for MDS research

Outline

- Introduction
- Related Work
- **Inspiration**

Inspiration

- CNN or Tree-LSTM to learn the hierarchical summarizing relation between sentences
- Fine-grained: word, phrase
- Sparse Latent Semantic Reconstruction
- Besides extraction, generate compressive or abstractive summaries to cover more important information

Thanks a lot!