Question Answering Research Progress in ITNLP

Chengjie Sun, Baoxun Wang, Bingquan Liu and Ming Liu

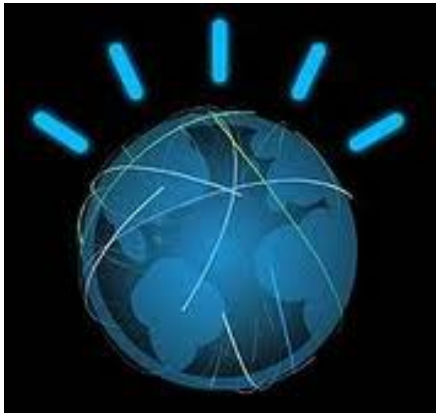
Intelligent Technology and Natural Language Processing Lab (ITNLP)
Harbin Institute of Technology
8/28/2012

Outline

- New era of QA
- Answer Detection
  - Deep belief network based answer detection
  - Thread segmentation based answer detection in Chinese online forums
- Question Generation
- Future works

New era of QA (1)

Single Applications



Watson 2011-2-14 Jeopardy! game

IBM Supercomputer Watson to Help Diagnose Medical Problems

Published September 12, 2011 / Associated Press

IBM's Watson Moves From 'Jeopardy' To Real Game

INVESTOR'S BUSINESS DAILY Investor's Business Daily – Fri, Mar 9, 2012 4:00 PM EST

April 28, 2010, acquired by Apple

New era of QA (2)

Auxiliary Applications

百度知道 > 搜索结果

全部回答 | 待完善问题

看看以下回答是否解决了您的疑问

问 深圳的别名是什么

答 鹏城

来自百度百科词条 深圳

深圳的别名是什么

鹏城...

393059 - 2011-04-01 12:07 - 最佳回答者: wuyu8

Yahoo CEO

About 123,000,000 results (0.20 seconds)

Marissa Mayer

Yahoo!, CEO

Hide details

Investors Show Displeasure ... - forbes.com

... new Yahoo CEO **Marissa Mayer**'s ...

Yahoo! - Wikipedia, the free ... - wikipedia.org

Marissa Mayer (CEO). Products, See Yahoo!

深圳的别名是什么

約有 733,000 項結果 (搜尋時間：0.20 秒)

將 "深圳 的 别名" 從中文(簡體)翻譯為目標語言

translate.google.com.hk

深圳 的 别名 - 深圳 的 别名

深圳的别名是什么 - 已回答- 搜搜问问

wenwen.soso.com › ... › 地区问题 › 广东 - 頁庫存檔 - 轉為繁體網頁

6 個答案 - 2010年4月24日

鹏城鹏城，希望采纳鹏城鹏城的说法又从何而来呢？在去大亚湾的途中，有一座城门，模样很象北京的德胜门，只是规模小了很多。城墙经风雨侵蚀，...

深圳的别名是什么？ 搜房问答

www.soufun.com/ask/.../c.../b_3843733.html - 頁庫存檔 - 轉為繁體網頁

深圳的别名是什么？ 2012-03-22 18:08. 提问者：寒风啸弑 | 浏览次数：50 次 | 问题来自：深圳. 寒风啸弑 | 一年级. 发送短消息. 我来帮他解答. 输入内容已经达到长度 ...

深圳的别名是什么_百度知道

zhidao.baidu.com › 地区 › 广东 - 頁庫存檔 - 轉為繁體網頁

1 個答案 - 2011年4月1日

最佳解答： 鹏城

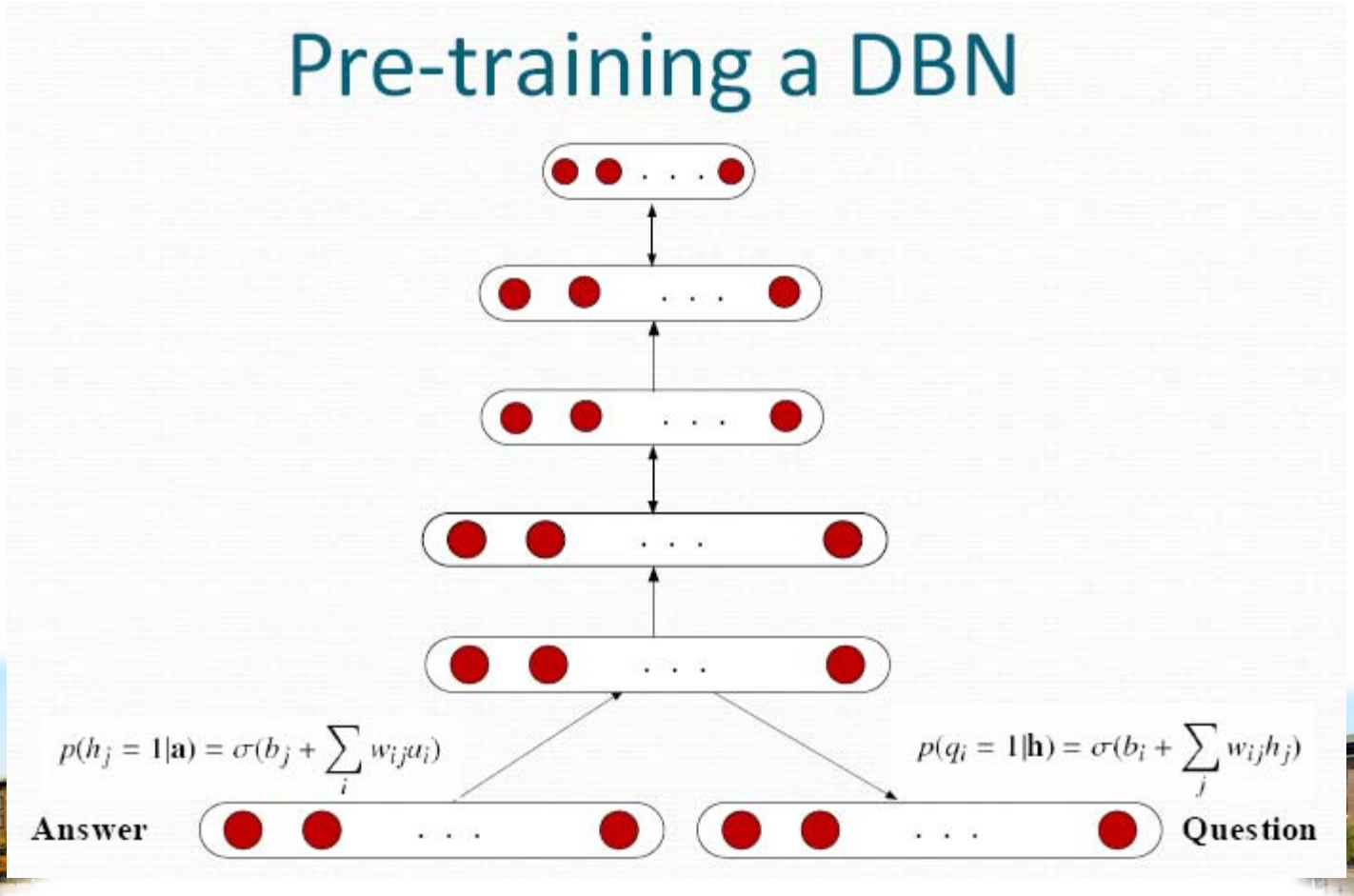
Learning from the past:
Answering new questions with past answer

Deep belief network based answer detection

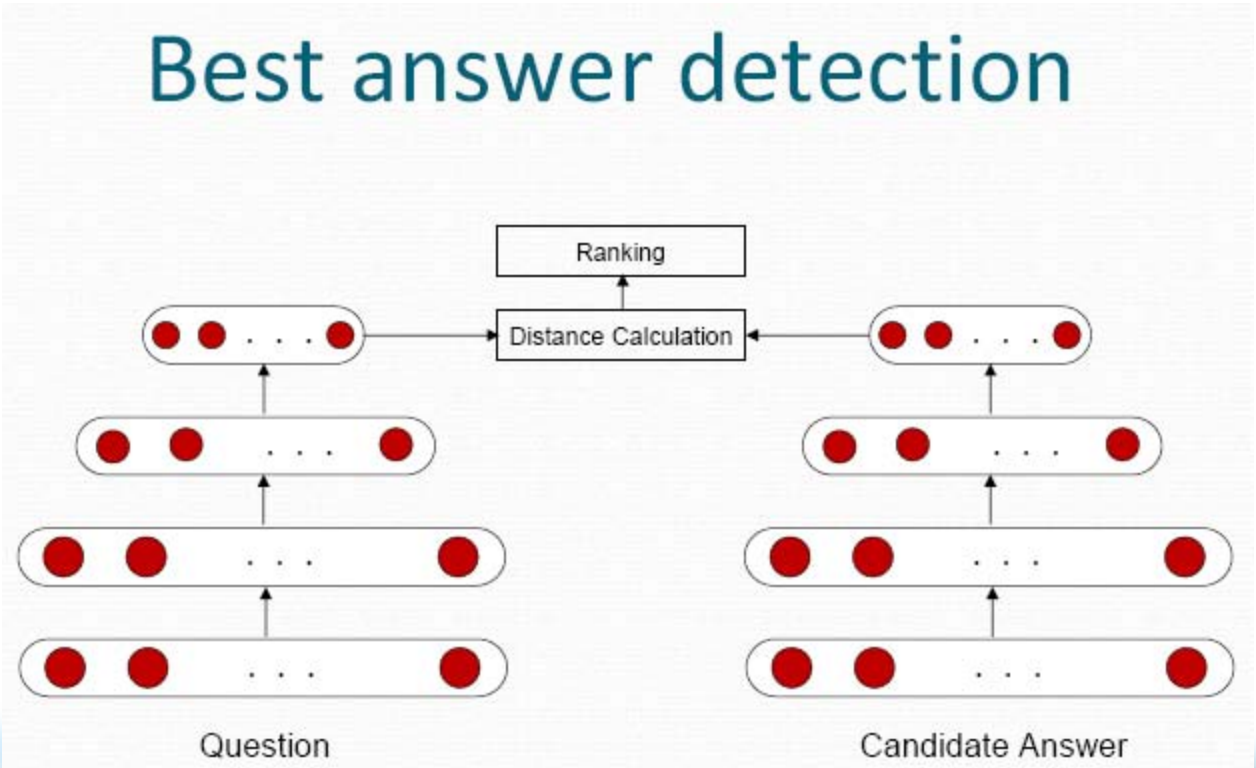
Motivation

- Automatically detect answer for questions in cQA and online forum
  - Train one model to work on both cQA and forum datasets
  - Avoid hand-annotating work
- Social media corpus
  - Feature sparsity
  - Low word frequency

Method(1)


Pre-training a DBN
$p(h_j = 1|\mathbf{a}) = \sigma(b_j + \sum_i w_{ij} a_i)$
$p(q_i = 1|\mathbf{h}) = \sigma(b_i + \sum_j w_{ij} h_j)$
Answer
Question

Method(2)


Best answer detection
Ranking
Distance Calculation
Question
Candidate Answer

Features

- 1,500 dimensional feature vector
  - 1,300 most frequent content words based on professional dictionary
  - 200 function words
- All the features are binary
  - Denote whether a word appears or not

Experiment(1)

- Experiment setup
  - Architecture of the Network
  - 1500-1500-1000-600 architecture
- Dataset
  - Training: 12,600 human generated QA pairs (From Baidu Zhidao)
  - Testing: 2,000 cQA pages (from: Baidu Zhidao) & 2,000 forum threads (from: http://bbs.cfanclub.net)
- Baseline
  - Cosine Similarity
  - HowNet based Similarity
  - KL-divergence Language Model
- Metrics
  - P@1
  - MRR

Experiment(2)

- Results and Analysis on forum test set

Method	P@1 (%)	MRR (%)
Nearest Answer	21.25	38.72
Cosine Similarity	23.15	43.50
HowNet	22.55	41.63
KL divergence	25.30	51.40
DBN (without FT)	**41.45**	**59.64**
DBN (with FT)	**45.00**	**62.03**

- DBN based approach outperforms the baseline methods
- The model is able to learn semantic relationship between QA pairs
- Training data from cQA provide rich information

Experiment(3)

- Results and Analysis on cQA test set

Method	P@1 (%)	MRR (%)
Nearest Answer	36.05	56.33
Cosine Similarity	44.05	62.84
HowNet	41.10	58.75
KL divergence	43.75	63.10
DBN (without FT)	**56.20**	**70.56**
DBN (with FT)	**58.15**	**72.74**

- Our approach obtains a 32.0% improvement in P@1 and a 15.3% improvement in MRR at least
- Compared to results on forum data, all the approaches perform much better

Experiment(4)

- Reasons for the unsatisfying performance of the baseline approaches
  - The sparsity of the features
  - Morphologically different words with the same meaning used (especially for Chinese)
  - Large amount of words not defined by the electronic world knowledge system
  - The baseline results indicate that the online forum is a complex environment for answer detection.

Contributions

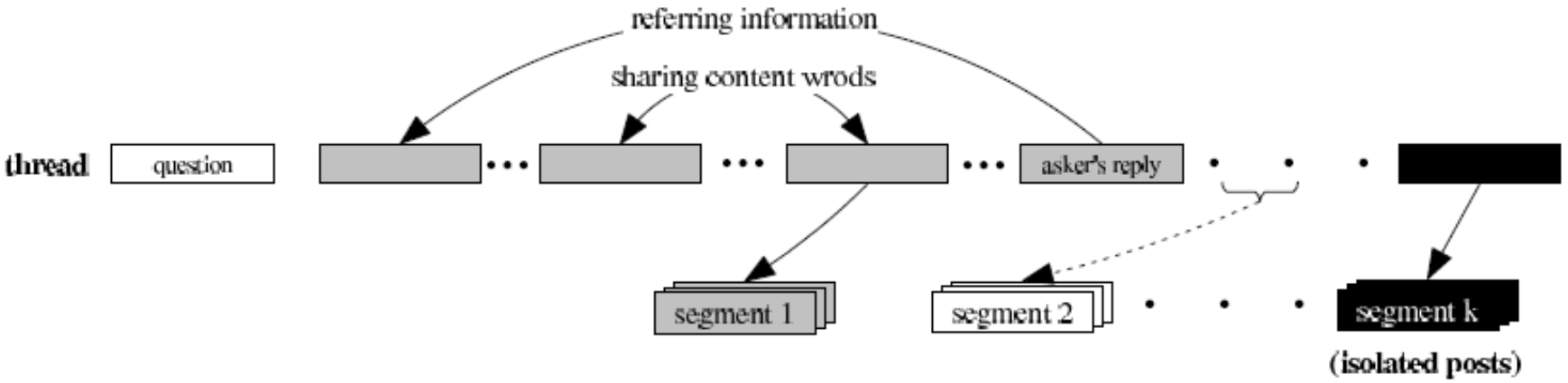
- The deep belief network proposed shows good performance on answer detection.
- Introduce a novel learning strategy to show good performance on both cQA and forum datasets.

Thread Segmentation based Answer Detection in Chinese Online Forums

Motivation

- Extract large amount of QA pairs from forum
- Using structure information of forum for answer detection

Illustration of the thread segmentation strategy



Thread Segmentation Example

Question	user1: 笔记本无法装系统，每次装的时候都提示如图，想知道到底是什么问题。[pic] user1: I can not install OS on my notebook. The computer gives the following information while installing. I wonder what the problem is. [pic]
Negative	user2: 估计又是端口问题! 先格式化C盘，再找个带SATA驱动的安装盘安装。 user2: I guess it is the port problem! You may format C:, then install the system using an installation disk with SATA driver. user3: 用U盘启动，通过PE进去后，把这个驱动装了试下[attachment] user3: Boot from USB disk and login via PE, then try this driver. [attachment] user1: 回 1楼(user2) 的帖子 不要说格式化C盘，我整个硬盘都格式化过。试过了所有可用的办法，不行。到底是哪里出问题了啊? user1: Re: 1# (user2) I have formatted the whole hard disk, let alone C:. Every possible method has been tried and does not work. What's the reason?
Negative	user2: 有必要重新分区看看! user2: It is necessary to repartition! user1: 回 4楼(user2) 的帖子 重新分区过了。重新分区分成4个盘过，不行，分过5个盘过，也不行。 user1: Re: 4# (user2) I have tried that. It changes nothing after repartitioning the disk into 4 or 5 parts.

Negative	user4: 由于提示是不清楚的硬件错误，因此，先要确定一下出现问题前是否动过硬件。 user4: Because the computer reports an uncertain error, you have to make clear whether you have changed your hardwares. user1: 回 6楼(user4) 的帖子 硬件没有怎么动过啊。换了根内存，可是内存我又换回原来的，还是这样的问题。 user1: Re: #6 (user4) I have rarely touched the hardwars, but changed a memory. The problem remains after plugging the old one back.
Negative	user5: 使用Ghost版系统用安装版系统试试 user5: Try the Ghost installation. user1: 回 8楼(user5) 的帖子 连买笔记本送的安装版光盘都找出来装的。还是不行。 user1: Re: 8# (user5) Even the OEM installation can not give any help.
Positive	user6: 开机自检是否正常通过? 若出错时，已经开始启动系统，请先从光盘启动或换一个硬盘试试; 若DOS自检也通不过，说明硬件有问题了 user6: Notice the start checking process. If it give an error but the system starts, you can reboot from CD-ROM or change a hard disk. Otherwise, it indicates the hardware problems. user1: 回 10楼(user6) 的帖子 问题已经解决，谢谢。我拿去换了一个硬盘就可以了。 user1: Re: 10# (user6) Problem solved, thanks. I have changed a new hard disk and everything goes well.

Workflow

Threads

Thread segmentation

Thread segments

Sentiment analysis

Positive
Neutral
Negative

position of segment
size of segment
segment type
number of the segments with each label
total number of the segments in the thread
Cosine similarity between question and candidate answer
KL-divergence of question and candidate answer
length of candidate answer
distance from the question post
answerer's activity
quote times of candidate answer
times of candidate answer being acknowledged

Answers

SVM training

on models

assifier

Experiment(1)

- Data set
  - 1,293 threads (randomly choose from 108,000 threads in Chinese forum ComputerFansClub)
  - Manually annotation

	Number of Segments	Number of Answers
positive	221	188
negative	302	60
neutral	1038	278
isolated	1076	767

Experiment(2)

- Effect of thread segmentation

Method	Without Segment Prediction			With Segment Prediction		
	P@1	MRR	P@3	P@1	MRR	P@3
KL-divergence	14.86%	40.69%	46.18%	29.00%	58.54%	78.70%
Cosine Similarity	32.41%	55.87%	60.94%	45.17%	64.81%	84.42%
Baseline-1	35.96%	53.09%	62.37%	37.38%	59.59%	77.92%
Baseline-2	49.06%	72.89%	84.56%	53.82%	72.58%	91.34%

Baseline-1: classification based method taking only the post position and the authorship as the features.
Baseline-2: without segment features

Experiment(2)

- Performance comparison with previous works

Method	P@1	MRR	P@3
Baseline-1	35.96%	53.09%	62.37%
Baseline-2	49.06%	72.89%	84.56%
TSAD	60.03%	77.98%	92.35%

Contribution

- Presented a thread segmentation strategy to reorganize the posts in a forum thread
- Based on the generated segments, a group of new features are taken into the feature collection for answer detection
- Applying the segment features, a classification based answer detecting approach is proposed. The experimental results have shown that our approach outperforms the baseline methods

Question Generation

Motivation

- Given answers, generate corresponding questions
  - Collection large QA pairs as knowledge source
  - Broaden the usage of automatically question generation
  - Use machine learning method to do question generation
    - Pattern or rule based strategies are current mainstream

Solution

- Focus on non-factoid answers
- Three sub-tasks [Nielsen, 2008]
  - Question type determination
  - Concept selection
  - Question construction

Experiment(1)

- Data Set
  - The “Travel" category of Yahoo! Answers.
    - 4 different topics: trip plans, cutting down expense, necessary items, and VISA application.
    - “resolved questions" with their best answers are crawled.
    - Remove answers less than 10 words or containing the URLs only
  - From each topic
    - 4,500 QA pairs for training
    - 100 randomly selected QA pairs for testing.

Experiment(2)

- Results

Method	Precision of question word generation			
	trip plans	cutting down expense	necessary items	VISA application
NB-qwg	0.19	0.27	0.23	0.28
KNN-qwg	0.22	0.36	0.25	0.33
DBN-qwg	0.36	0.62	0.57	0.49

Method	Precision of content word generation			
	trip plans	cutting down expense	necessary items	VISA application
NB-cwg	0.28	0.38	0.32	0.35
KNN-cwg	0.26	0.45	0.36	0.42
DBN-cwg	0.51	0.72	0.69	0.58

Future works

- Web-based QA
  - Extract answers using the results of existing search engine
    - Question → queries
    - Answer extraction & ranking
    - Confidence estimation
- Multi-media QA
  - Provide more answer source: image, audio and video
- Social QA
  - Find suitable experts to solve users' questions

Reference

- Baoxun Wang, Bingquan Liu, Xiaolong Wang, Chengjie Sun. ***Deep Learning Approaches to Semantic Relevance Modeling for Chinese Question-Answer Pairs***. Transactions on Asian Language Information Processing. 2011
- Baoxun Wang, Bingquan Liu, Chengjie Sun, Xiaolong Wang***. Thread Segmentation based Answer Detection in Chinese Online Forums.*** Acta Automatica Sinica, 2012