Wikipedia Question Generating and Answering System

Zhiyong Yang Chen-Chien Chung

Introduction

- This question generating system take a wikipedia article and a number as input and output that number of questions. The answering system take wikipedia article and a question file corresponding this wikipedia article and output the answers according to the given questions.

- The reason we choose wikipedia articles is that their contents are general and random so that we can have dynamic and totally random test cases.

High Level Description of Approach

1.Article Simplifying—Extracting Sentences

2.Generate “WH” Questions—Replacing NP

3.Generate “Yes/No” Questions—Tsurgeon Pattern

4.Select Top Questions—Language_Check Tools

Article Simplifying

1.NLTK-Tokenizer

2.Stanford NLP Parser

3.Dependency Structure

compound(English-2, Middle-1)
nsubj(began-3, English-2)
root(ROOT-0, began-3)
case(century-8, in-4)
det(century-8, the-5)
amod(century-8, late-6)
amod(century-8, 11th-7)
nmod:in(began-3, century-8)
case(conquest-12, with-9)
det(conquest-12, the-10)
compound(conquest-12, Norman-11)
nmod:with(began-3, conquest-12)
case(England-14, of-13)
nmod:of(conquest-12, England-14)

......

'English has developed over the course of more than 1,400 years.', \' The "inner circle" countries with many native speakers of English share an international standard of written English and jointly influence speech norms of English around the world.', 'English does not belong to just one country, and it does not belong solely to descendants of English settlers.', 'English is an official language of countries populated by few descendants of native speakers of English.', 'It has also become by far the most important language of international communication when people who share no native language meet anywhere in the world.']

Figure2. The Sentences that are selected

Question Generating

Dependency Parse

```
1   The        _   DT    DT    _    2    det        _          _
2   countries  _   NNS   NNS   _    10   nsubjpass  _          _
3   in         _   IN    IN    _    4    case       _          _
4   which      _   WDT   WDT   _    7    nmod       _          _
5   English    _   NNP   NNP   _    7    nsubjpass  _          _
6   is         _   VBZ   VBZ   _    7    auxpass    _          _
7   spoken     _   VBN   VBN   _    2    acl:relcl  _          _
8   can        _   MD    MD    _    10   aux        _          _
9   be         _   VB    VB    _    10   auxpass    _          _
10  grouped    _   VBN   VBN   _    0    root       _          _
11  into       _   IN    IN    _    13   case       _          _
12  different  _   JJ    JJ    _    13   amod       _          _
13  categories _   NNS   NNS   _    10   nmod       _          _
14  by         _   IN    IN    _    18   mark       _          _
15  how        _   WRB   WRB   _    16   advmod     _          _
16  English    _   JJ    JJ    _    18   advmod     _          _
17  is         _   VBZ   VBZ   _    18   auxpass    _          _
18  used       _   VBN   VBN   _    10   advcl      _          _
19  in         _   IN    IN    _    21   case       _          _
20  each       _   DT    DT    _    21   det        _          _
21  country    _   NN    NN    _    18   nmod       _          _

1   As         _   IN    IN    _    3    case       _          _
2   of         _   IN    IN    _    3    case       _          _
3   2010       _   CD    CD    _    8    nmod       _          _
5   359        _   CD    CD    _    6    compound   _          _
6   million    _   CD    CD    _    7    nummod     _          _
7   people     _   NNS   NNS   _    8    nsubj      _          _
8   spoke      _   VBD   VBD   _    0    root       _          _
9   English    _   NNP   NNP   _    8    dobj       _          _
10  as         _   IN    IN    _    13   case       _          _
11  their      _   PRP$  PRP$  _    13   nmod:poss  _          _
12  first      _   JJ    JJ    _    13   amod       _          _
13  language   _   NN    NN    _    8    nmod       _          _
```

Constituency Parse

```
(ROOT
  (S
    (NP
      (NP (DT The) (NNS countries))
      (SBAR
        (WHPP (IN in)
          (WHNP (WDT which)))
        (S
          (NP (NNP English))
          (VP (VBZ is)
            (VP (VBN spoken))))))
    (VP (MD can)
      (VP (VB be)
        (VP (VBN grouped)
          (PP (IN into)
            (NP (JJ different) (NNS categories)))
          (PP (IN by)
            (SBAR
              (WHNP (WRB how) (JJ English))
              (S
                (VP (VBZ is)
                  (VP (VBN used)
                    (PP (IN in)
                      (NP (DT each) (NN country))))))))))
    (. .)))

(ROOT
  (S
    (PP (IN As)
      (PP (IN of)
        (NP (CD 2010))))
    (, ,)
    (NP
      (QP (CD 359) (CD million))
      (NNS people))
    (VP (VBD spoke)
      (NP (NNP English))
      (PP (IN as)
        (NP (PRP$ their) (JJ first) (NN language))))
    (. .)))
```

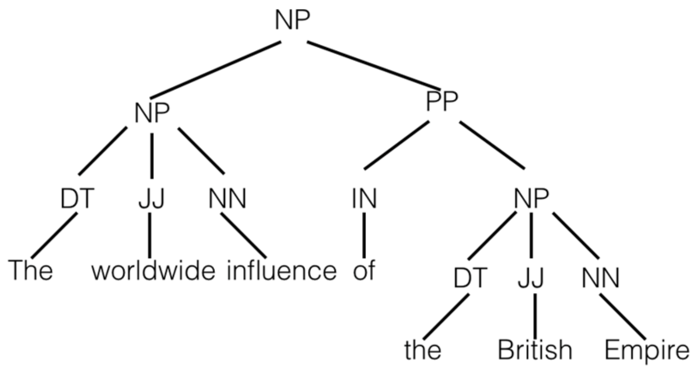
"WH" Questions

1.Extracting Subjects—Dependency → Three-Tuple

[((u'began', u'VBD'), u'nsubj', (u'English', u'NNP')), ((u'English', u'NNP'), u'compound', (u'Middle', u'NNP')), ((u'began', u'VBD'), u'nmod', (u'century', u'NN')), ((u'century', u'NN'), u'case', (u'in', u'IN')), ((u'century', u'NN'), u'det', (u'the', u'DT')), ((u'century', u'NN'), u'amod', (u'late', u' JJ')), ((u'century', u'NN'), u'amod', (u'11th', u' JJ'))]

Figure 5 The three-element tuples

2.Replace NP with "WH" —Tracing back, NLTK.Lesk()



What is an official language of countries populated by few descendants of native speakers of What ?
However , when combining native and non-native speakers What is probably the most commonly spoken language in the world ?
What can be grouped into different categories by how English is used in each country ?
What has also become by far the most important language of international communication when people who share no native language meet anywhere in the world ?

Figure 7 The "WH" questions generated

"Yes/No" Questions

1.Declarative →General Question—Tsurgeon Syntax Patterns

```
(ROOT
  (S
    (NP (NNP Middle) (NNP English))
    (VP (VBD began)
      (PP (IN in)
        (NP (DT the) (JJ late) (JJ 11th) (NN
century)))
      (PP (IN with)
        (NP
          (NP (DT the) (NNP Norman) (NN
conquest))
          (PP (IN of)
            (NP (NNP England))))))
    (. .)))
```

```
(ROOT
  (S (DO Did)
    (NP (NNP Middle) (NNP English))
    (VP (VERBBASE began)
      (PP (IN in)
        (NP (DT the) (JJ late) (JJ 11th) (NN
century)))
      (PP (IN with)
        (NP
          (NP (DT the) (NNP Norman) (NN
conquest))
          (PP (IN of)
            (NP (NNP England))))))
    (. .)))
```

```
NP=subj [. VBD=verb & !,DO]

insert (DO Did) $+ subj
relabel verb VERBBASE
```

2.Tense Change— "en" from NodeBox

Score Questions

1. Pronouns Elimination— “he” , “him” ,” They” ,” Their”

2. The lower, the better—language_check.LanguageTool

Is English not spoken by communities on every continent and on oceanic islands in all the major oceans ?
Is English probably the third largest language by number of native speakers , after Mandarin and Spanish ?
Are The earliest forms of English , a set of Anglo-Frisian dialects brought to Great Britain by Anglo-Saxon settlers in the fifth century , called Old English ?
Has English not developed over the course of more than 1,400 years ?

Figure 10. The “Yes/No” questions generated

Results and Evaluation

1. No Quantitative way—Subjective Judgement

2. Best way so far—Mutual Evaluate with Answering System

```
Zhiyongs-MacBook-Pro:ProjectTry zhiyongyang$ python ProjectTry.py ShortEnglish.txt 4
What has also become by far the most important language of international communication when p
eople who share no native language meet anywhere in the world ?
Is English not  spoken by communities on every continent and on oceanic islands in all the ma
jor oceans ?
The countries in which is English  spoken can be grouped into different categories by how Eng
lish is used in each country ?
What is an official language of countries populated by few descendants of native speakers of
What ?
Zhiyongs-MacBook-Pro:ProjectTry zhiyongyang$
```

```
Zhiyongs-MacBook-Pro:ProjectTry zhiyongyang$ python ProjectTry.py China.txt 5
Who -RSB- The state is governed by the Communist Party of China and its capital is Beijing ?
-LSB- 15 -RSB- The state is governed by the Communist Party of China and What is Beijing ?
Where is a great power and a major regional power within Asia , and has been characterized as
 a potential superpower ?
Did -LSB- 15 -RSB- The state is governed by the Communist Party of China and its capital is B
eijing ?
What mountain ranges separate China from much of South and Central Asia ?
```

```
Zhiyongs-MacBook-Pro:ProjectTry zhiyongyang$ python ProjectTry.py Chinese.txt 8
Who -LRB- 1987 -RRB- , The Languages of China , Princeton University Press , ISBN 978-0-691-0
1468-5 ?
As such , have most of these words not  been replaced -LRB- in speech , if not in writing -RR
B- with a longer , less-ambiguous compound ?
Does Ramsey , S. Robert -LRB- 1987 -RRB- , The Languages of China , Princeton University Pres
s , ISBN 978-0-691-01468-5 ?
Has Sound change over time not  steadily reduced the number of possible syllables ?
Would A better term for a Chinese character be morpheme , as characters represent the smalles
t grammatical units , individual meanings , and/or syllables in the Chinese language ?
Did Jerry Norman estimate that there are hundreds of mutually unintelligible varieties of Chi
nese ?
Who have then been borrowed freely between languages ?
Wade -- was Giles not  found in academic use in the United States , particularly before the 1
980s , and until recently was widely used in Taiwan ?
```

Answering Procedure

- 1. Find the sentence matching between question and sentences in the article to get the target sentence
- 2. Question type recognition
- 3. Process question and target sentence using NLTK, Stanford NLP tools
- 4. Applying rule to get the answers

Binary Question

- Simply use Fuzzywuzzy matching score to get the target sentence

```
Donald John Trump (born June 14, 1946) is the 45th and current President of
 the United States.
[Q: is donald trump the president of united states of america? ]
[A:-----answer from QAsystem----->] Yes
```

Who Question

- Stanford Name Entity Recognizer

```
[NERtag]= [(u'Donald', u'PERSON'), (u'John', u'PERSON'), (u'Trump', u'PERSO
N'), (u'(', u'O'), (u'born', u'O'), (u'June', u'O'), (u'14', u'O'), (u',',
u'O'), (u'1946', u'O'), (u')', u'O'), (u'is', u'O'), (u'the', u'O'), (u'45t
h', u'O'), (u'and', u'O'), (u'current', u'O'), (u'President', u'O'), (u'of'
, u'O'), (u'the', u'O'), (u'United', u'LOCATION'), (u'States', u'LOCATION')
, (u'.', u'O')]
Donald John Trump
[Q: who is current president of the united states? ]
[A:-----answer from QAsystem----->] Donald John Trump
```

Where Question

- Stanford Name Entity Recognizer

```
[NERtag]= [(u'Trump', u'PERSON'), (u'was', u'O'), (u'born', u'O'), (u'and',
 u'O'), (u'raised', u'O'), (u'in', u'O'), (u'Queens', u'LOCATION'), (u',',
u'O'), (u'New', u'LOCATION'), (u'York', u'LOCATION'), (u'City', u'LOCATION'
), (u',', u'O'), (u'and', u'O'), (u'earned', u'O'), (u'an', u'O'), (u'econo
mics', u'O'), (u'degree', u'O'), (u'from', u'O'), (u'the', u'O'), (u'Wharto
n', u'ORGANIZATION'), (u'School', u'ORGANIZATION'), (u'of', u'ORGANIZATION'
), (u'the', u'ORGANIZATION'), (u'University', u'ORGANIZATION'), (u'of', u'O
RGANIZATION'), (u'Pennsylvania', u'ORGANIZATION'), (u'.', u'O')]
Here here
[Q: where was president trump born and raised? ]
[A:-----answer from QAsystem----->] Queens
```

How Many Question

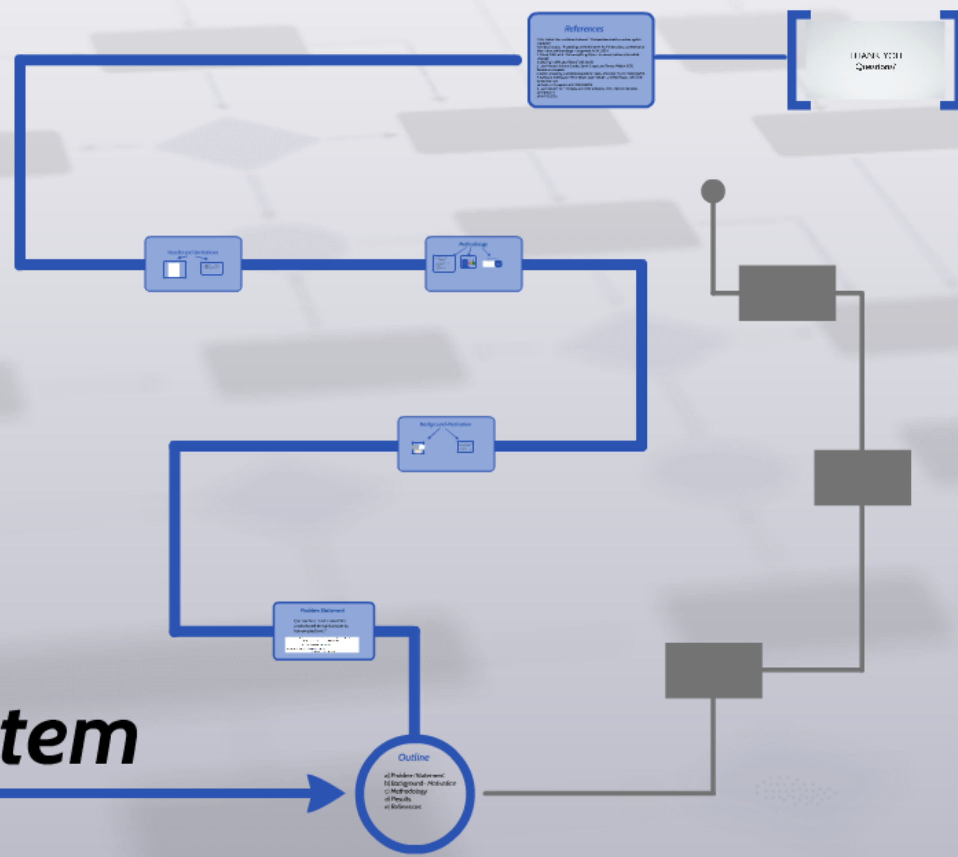
- Stanford POS tagger, Stanford PCFG Parser
- Search the pattern [CD, NNS]

```
[question sentence pos tag]= [(u'how', u'WRB'), (u'many', u'JJ'), (u'grandc
hildren', u'NNS'), (u'does', u'VBZ'), (u'president', u'NN'), (u'trump', u'N
N'), (u'have', u'VB'), (u'?', u'.')]
[target sentenct pcfg parse]= (ROOT
  (NP
    (NP (NNP Main) (NN article))
    (: :)
    (S
      (NP (NNP Trump))
      (VP
        (VBZ has)
        (NP (CD five) (NNS children))
        (PP
          (IN by)
          (NP
            (NP (CD three) (NNS marriages))
            (, ,)
            (CC and)
            (NP (CD eight) (NNS grandchildren))))))
    (. .)))
[Q: how many grandchildren does president trump have? ]
[A:-----answer from QAsystem---->] eight
```

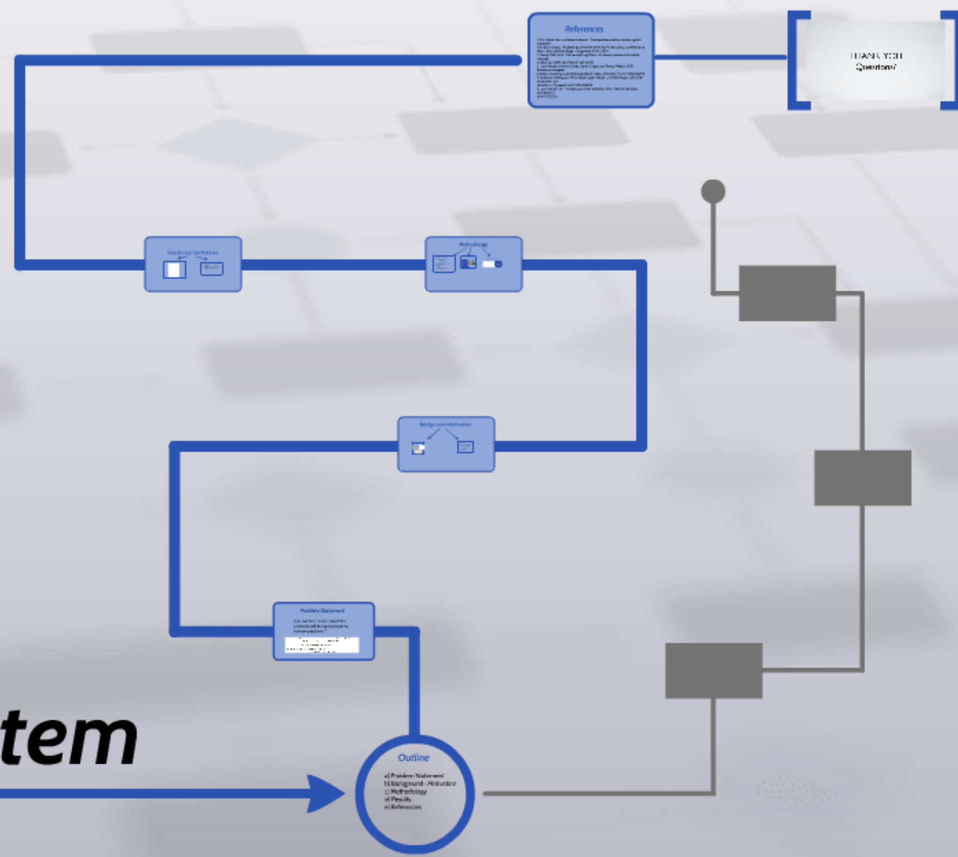
Results and Evaluation

- For the test file Trump.txt/Trump_question.txt, it can answer correctly four out of six question.
- When target sentenct does not contain answer, it will output wrong answer or just give the target sentence
- Adding more rules/patterns can enhance the ability of this answering system

Question Answering System



Question Answering System



Outline

a) Problem Statement
b) Background- Motivation
c) Methodology
d) Results
e) References

Problem Statement

Can machine "understand" the unstructured text and answer to human questions ?

Data comprises of information and question. Information can contain a number of sentences-

$$I = \{s_1, s_2, \ldots\ldots, s_n\} \qquad ; s_i = i^{th} sentence$$

$$q = question\ based\ on\ information\ I$$

The goal of the QA system is to compute an answer

$$a = answer\ to\ the\ question\ q$$

Background-Motivation

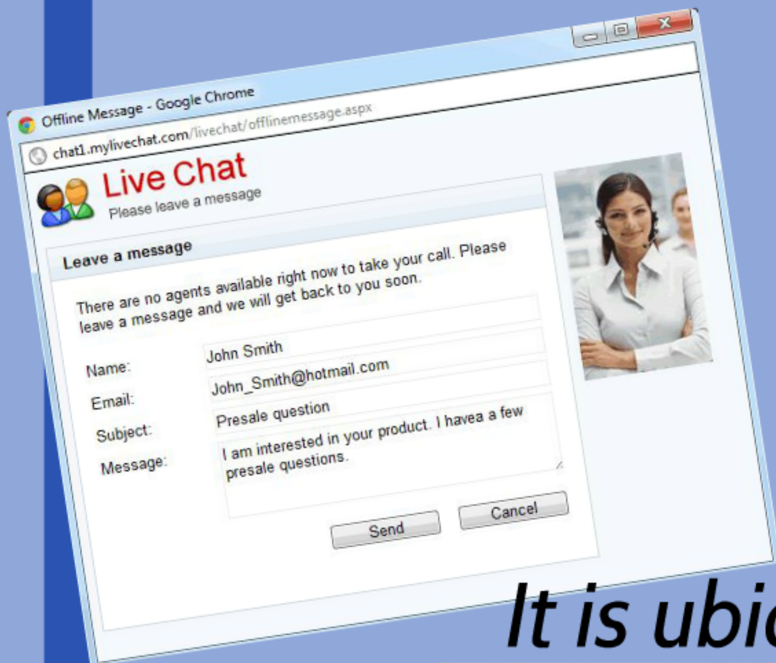
Applications

Approaches

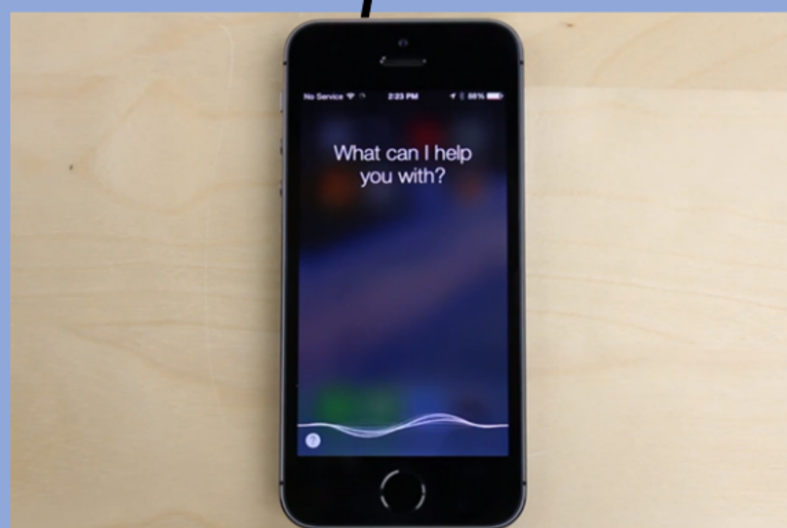
1. Information Retrieval/Extraction (Pattern) based
   - Identify question type and target words
   - Faster but doesn't work for complex tasks

2. Memory Net based
   - Neural networks to "remember" sequence
   - Computationally expensive but better scalability

Applications



It is ubiquitous !!



Approaches

1. Information Retrieval/Extraction (Pattern) based

- Identify question type and target words
- Faster but doesn't work for complex tasks

2. Memory Net based

- Neural networks to "remember" sequence
- Computationally expensive but better scalability

Methodology

Dataset- fb babi tasks

Format: Text- Question- Answer

1 Mary moved to the bathroom.
2 John went to the hallway.
3 Where is Mary?   bathroom   1

20 tasks- 1000 training and testing in each

- Single/two/three supporting facts
- Yes/ No questions
- Simple Negation
- Counting
- Compound coreference
- Basic deduction/ induction
- Path Finding
- ----

Implementations

Idea- use information extraction method

Identify POS, NER, SRL in input sentence
Extract information using predicate grouping
Identify entities using NER
Store extracted information in ***knowledge bank***.
Process question in same way
Derive answer using knowledge bank

For example,

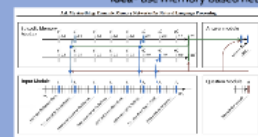
John went to hallway

Person | motion verb | Location

Knowledge Bank:
Location: {'John': Hallway}

Implementations

Idea- use memory based neural network (LSTM)



Encode input story into hidden states using LSTM
Encode ques into hidden states using same embedding
Episodic Memory construction
Generate Answer

Dataset- fb babi tasks

Format: Text- Question- Answer

1 Mary moved to the bathroom.
2 John went to the hallway.
3 Where is Mary? bathroom 1

20 tasks- 1000 training and testing in each

- Single/two/three supporting facts
- Yes/ No questions
- Simple Negation
- Counting
- Compound coreference
- Basic deduction/ induction
- Path Finding
- ----

Implementations

Idea- use information extraction method

Identify POS, NER, SRL in input sentence

Extract information using predicate grouping

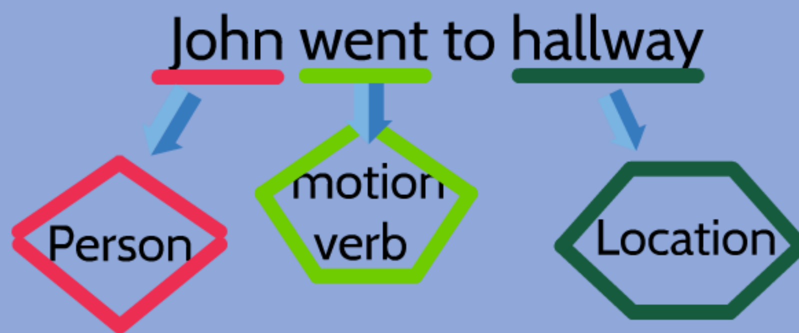
Identify entities using NER

Store extracted information in ***knowledge bank***.

Process question in same way

Derive answer using knowledge bank

For example,

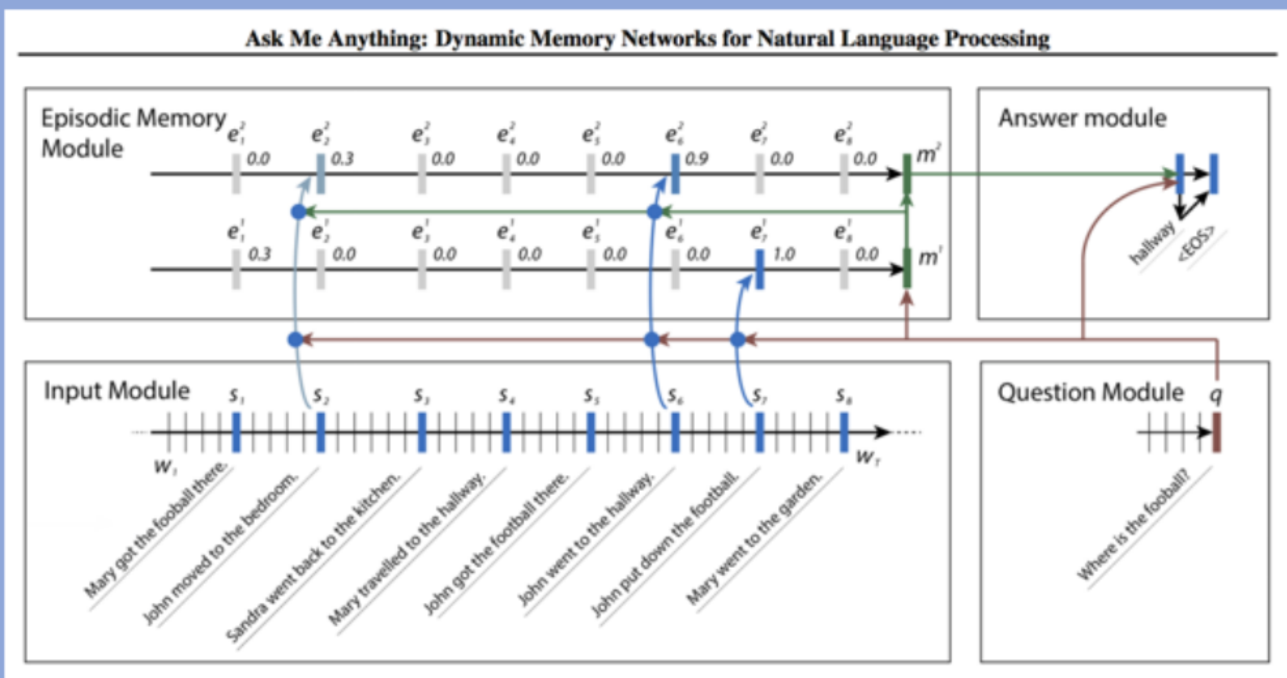


Knowledge Bank:

Location: {'John': Hallway}

Implementations

Idea- use memory based neural network (LSTM)



Encode input story into hidden states using LSTM

Encode ques into hidden states using same embedding

Episodic Memory construction

Generate Answer

Results and Limitations

Results

Limitations

- Information Extraction based QA is modeled for 4 bAbI tasks, can extend it more
- Modify algorithm for advanced tasks like news summarization and event extraction
  - Use hybrid algorithms- Information extraction for preliminary processing and neural networks in next step

Results

#	Task	Class name	% Accuracy IR/IE QA System	% Accuracy Memory Net QA system [2] result/our result (3hr training)
	Basic factoid QA with single supporting fact	WhereIsActor	100.0	100.0/100.0
2	Factoid QA with two supporting facts	WhereIsObject	69.8	98.2/30.1
3	Factoid QA with three supporting facts	WhereWasObject	76.52	95.2/32.10
4	Two argument relations: subject vs. object	IsDir	100.0	100/99.20
5	Three argument relations	WhoWhatGave	-	99.3/94.20
6	Yes/No questions	IsActorThere	-	100/95.40
7	Counting	Counting	-	96.9/78.50
8	Lists/Sets	Listing	-	96.5/95.0
9	Simple Negation	Negation	-	100/89.0
10	Indefinite Knowledge	Indefinite	-	97.5/94.50
11	Basic coreference	BasicCoreference	-	99.9/76.10
12	Conjunction	Conjunction	-	100/97.20
13	Compound coreference	CompoundCoreference	-	99.8/91.30
14	Time manipulation	Time	-	100/72.30
15	Basic deduction	Deduction	-	100/47.10
16	Basic induction	Induction	-	99.4/44.40
17	Positional reasoning	PositionalReasoning	-	59.6/56.3
18	Reasoning about size	Size	-	95.3/91.20
19	Path finding	PathFinding	-	34.5/8.90
20	Reasoning about agent's motivation	Motivations	-	100/98.40

Limitations

- Information Extraction based QA is modeled for 4 bAbI tasks, can extend it more
- Modify algorithm for advanced tasks like news summarization and event extraction
  - Use hybrid algorithms- Information extraction for preliminary processing and neural networks in next step

References

1. Lita, Lucian Vlad, and Jaime Carbonell. "Unsupervised question answering data acquisition
from local corpora." Proceedings of the thirteenth ACM international conference on Information and knowledge management. ACM, 2004.
2. Kumar, Ankit, et al. "Ask me anything: Dynamic memory networks for natural language
processing." CoRR, abs/1506.07285 (2015).
3. Jason Weston, Antoine Bordes, Sumit Chopra, and Tomas Mikolov. 2015. Towards ai-complete
question answering: a set of prerequisite toy tasks. arXiv preprint arXiv:1502.05698.
4. Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. 2015. End-to-end memory
networks. arXiv preprint arXiv:1503.08895.
5. Jason Weston, Sumit Chopra, and Antoine Bordes. 2014. Memory networks. arXiv preprint
arXiv:1410.3916.

THANK YOU

Questions?

LEARNING QUESTION CLASSIFIERS FOR A QUESTION ANSWERING SYSTEM

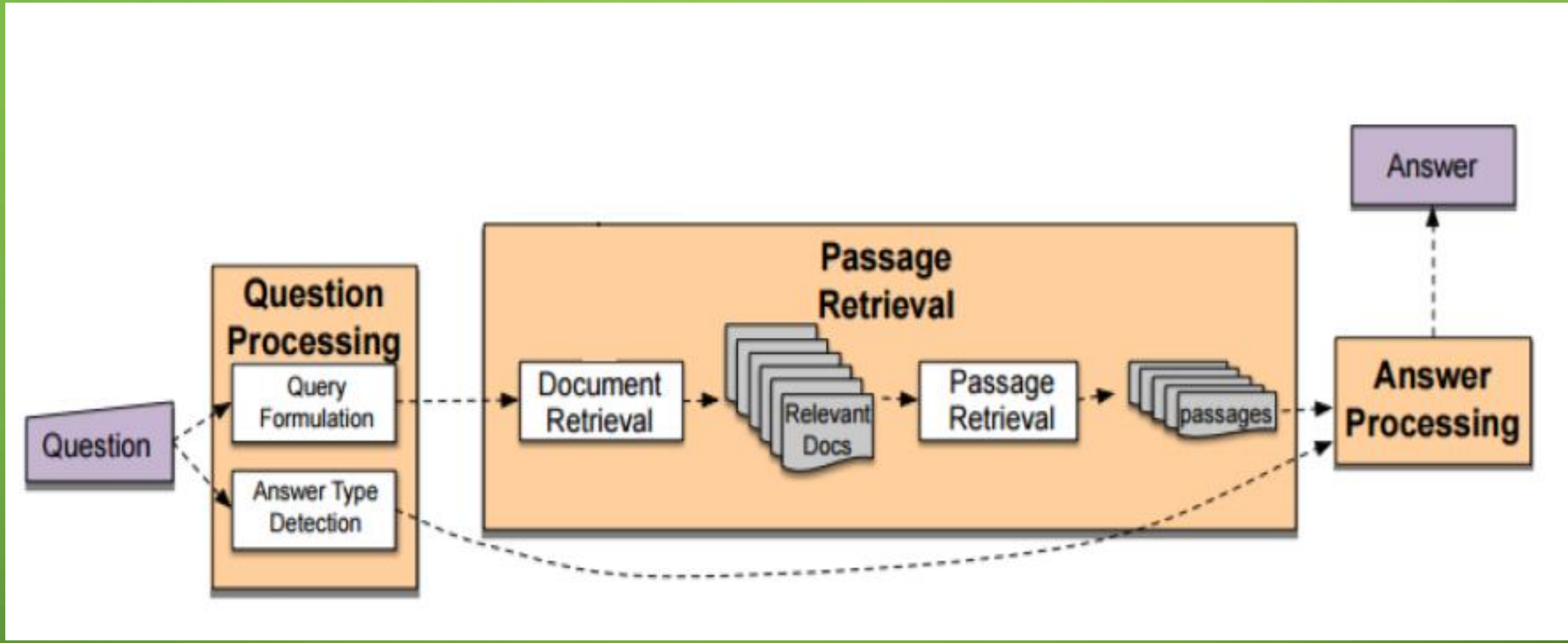
-SUSHIRDEEP NARAYANA

UIN: 124005538

OUTLINE OF PRESENTATION

- Introduction to the Problem
- Question Classification as a multiclass Classification
- Feature Extraction
- Multiclass Support Vector Machines
- Experiment
- Inferences and Conclusions

INTRODUCTION TO QUESTION ANSWERING SYSTEM



QA System consists of 3 parts

1) Question Processing module

2) Information Retrieval

3) Answer Processing module

INTRODUCTION TO QUESTION CLASSIFIERS

- Locating an accurate Answer depends on filtering a wide range of candidate answers
- Two purposes of a Question Classifier
- 1) Constraints answer types to precisely identify and verify the answer
- (Used in Answer Processing Module)
- 2) Provides information that downstream processes can utilize to determine answer selection approaches
- (Used in Information Retrieval Module)

INTRODUCTION TO QUESTION CLASSIFIERS

- Constructing Rule based manual question classifier can be very difficult
- Reason : a single query can have many reformulations
- Eg: What tourists attractions are there in Chicago ?
- What do most tourists visit in Chicago ?
- What attracts tourists to Chicago?
- (All reformulations target answer type : **Location**)

INTRODUCTION TO QUESTION CLASSIFIERS

- This work focuses on Machine Learning Approach to Question Classification
- Classify the questions into different semantic categories based on the semantic type of the answers
- Two classification tasks
- 1) Coarse –grained classification (6 coarse answer types)

(ABBR, DESC, ENTY, HUM, LOC, NUM)

- 2) Fine –grained classification (47 fine classes)

(abb, exp, animal, body, color, currency, event, food, instru, lang, ENTY:other, sport, def, desc, manner, reason, ind, title, city, country, LOC:other, state, date, temp, …)

QUESTION CLASSIFICATION

- Difference between Question Classification and Text Categorization

Questions are short and contain less word based information compared to a text document

This project

- a) Compares contribution of different features to classification performance
- b) Test performance of the classifier as to how well they categorize questions into fine and coarse class labels

QUESTION CLASSIFICATION

- Question Classification – multi-class Classification task that maps

$$g: X \rightarrow \{c_1, c_2, c_3, \ldots, c_n\}$$

- where X = features collected from the questions
- c_i = class i
- Features used in Question Classification

1) Bag of Words

Syntactic Features

2) POS-tags

3) Chunk tags

Semantic Features

4) Named Entities

BAG OF WORDS FEATURES

- The words from the questions are represented in bag
- The grammar and word order are ignored
- Multiplicity of the word is taken into consideration

- Eg: What movie is John watch?
- How was the movie Jurassic Park ?
- Construct the list and convert it into a vector accordingly
- List = {“what”, “movie”, “John”, “watch”, “How”, “movie”, “Jurassic”, “Park”}

POS- TAG FEATURES

- To include syntactic features (the Part of Speech Tags were extracted)
- POS tagger of NLTK was implemented
- POS Tags of the words in the questions were annotated and the features extracted accordingly

Example

- Q: Who was the first woman killed in the Vietnam War?
- POS tagged: [WP] [VBD] [DT] [JJ] [NN][VBN] [IN] [DT] [NNP] [NNP]

CHUNK TAG FEATURES

- Chunks – non-overlapping phrases in a sentence
- Chunk tags were extracted using NLTK parser

Eg:

Q: Who was the first woman killed in the Vietnam War?

Chunking : [NP Who] [VP was] [NP the first woman] [VP killed] [PP in] [NP in the Vietnam War?]

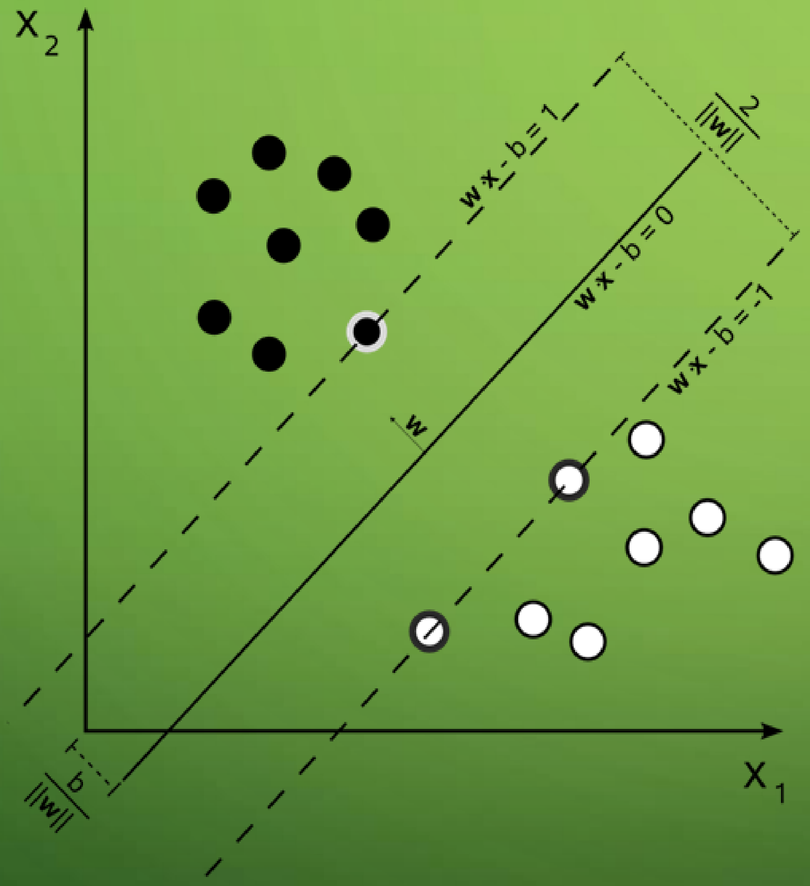
NAMED ENITITIES

- Lexical Semantic Information was Extracted in the form of Named Entities
- Named Entity Recognizer assigns a semantic category to a few noun phrases
- The Stanford Named Entity Recognizer was applied to extract feature corresponding to Named Entities
- The scope of the categories is the scope available through the Stanford NER

MULTICLASS SVM CLASSIFICATION

- Multiclass Support Vector Machines (SVM) with linear kernel were used
- Type of Multiclass classification implemented was one vs. one and one vs. rest
- N(N-1)/2 binary learners were constructed for one vs. one approach
- N binary learners were constructed for one vs. rest approach
- For each binary learner, one class is positive, another is negative, and the algorithm ignores the rest. This design exhausts all combinations of class pair assignments.

MULTICLASS SVM CLASSIFICATION



- SVM classifies the data by finding best hyperplane that separates all data points of one class from those of the other class

MULTICLASS SVM CLASSIFICATION

- Mathematical Formulation of SVM – The Primal Problem
- $min_{w,b} \frac{||w||^2}{2}$ $subject\ to$

  $y_i(wx_i \; - b) \geq 1 \; \forall \; i = 1,2, \ldots, m$

- Soft Margin – Introduce Slack variables ξ_i

  $\min_{w,b} \frac{||w||^2}{2} + C \; \sum_{i=1}^{m} \xi_i$

  $such\ that\ y_i \, (wx_i \; - b) \geq 1 - \xi_i, \; \forall \; i = 1,2, \ldots$

C – parameter (C= 1) and Linear Kernel were used

EXPERIMENTS

- Dataset collected
- Li and Roth Question Classification dataset [1]
- The dataset is a collection of questions from the TREC conference datasets
- Available through
- https://cogcomp.cs.illinois.edu/page/resource_view/49
- Training set = 5500 questions, 6 coarse labels, 47 fine grained labels
- Test Set = 500 question

- Programs composed in Python

[1] X. Li and D. Roth, Learning Question Classifiers: The Role of Semantic Information Journal of Natural Language Engineering (2005)

RESULTS OF COARSE QUESTION CLASSIFICATION

Coarse Classifier	Bag of Word Features	Bag Words + POS tag	Bag Words + POS tags + Chunk tags (Complete Syntactic Features)	Bag Words + POS tags + Chunk tags + Named Entities (Semantic _Syntactic Features)
One vs. Rest SVM (Support Vector Machines)	87.8 % (439/ 500)	87.4 % (437/ 500)	87.8% (439/ 500)	88.0 % (440/500)
One vs. One SVM (Support Vector Machines)	86.6 % (433/ 500)	86.6 % (433/ 500)	87.6 % (438/ 500)	88.0 % (440/500)
[1] SNoW (Sparse Network of Winnows)	85.10 %	91.80 %	91.80 %	93 %

Class Labels = 6 Coarse labels

[1] X. Li and D. Roth, Learning Question Classifiers: The Role of Semantic Information Journal of Natural Language Engineering (2005)

RESULTS OF COARSE QUESTION CLASSIFICATION

True Class Labels

Predicted labels

	ABBR	DESC	ENTY	HUM	LOC	NUM
ABBR	7	0	0	0	0	0
DESC	2	136	12	1	8	7
ENTY	0	1	70	3	4	0
HUM	0	0	6	61	1	0
LOC	0	0	6	0	67	1
NUM	0	1	0	0	1	99

Class	Precision
ABBR	100 %
DESC	81.93 %
ENTY	86.42 %
HUM	89.70 %
LOC	88.15 %
NUM	98.02 %

Confusion Matrix for One vs. Rest SVM classification with Bag Words + POS tags + Chunk tags + Named Entities Features

$$Precison\ [c] = \frac{\#\ correct\ predictions\ for\ class\ c}{\#\ of\ predictions\ for\ class\ c}$$

RESULTS OF COARSE QUESTION CLASSIFICATION

True Coarse Class labels

Predicted labels

	ABBR	DESC	ENTY	HUM	LOC	NUM
ABBR	7	0	0	0	0	0
DESC	2	136	11	2	7	9
ENTY	0	2	72	4	5	2
HUM	0	0	6	59	0	0
LOC	0	0	5	0	68	3
NUM	0	0	0	0	1	98

Class	Precision
ABBR	100 %
DESC	81.44 %
ENTY	84.70 %
HUM	90.76 %
LOC	89.47 %
NUM	98.98 %

Confusion Matrix for One vs. One SVM classification with Bag Words + POS tags + Chunk tags + Named Entities Features

$$Precison\,[c] = \frac{\#\ correct\ predictions\ for\ class\ c}{\#\ of\ predictions\ for\ class\ c}$$

RESULTS OF FINE-GRAINED QUESTION CLASSIFICATION

Fine Grained Classifier	**Bag of Word Features**	**Bag Words + POS tag**	**Bag Words + POS tags + Chunk tags (Complete Syntactic Features)**	**Bag Words + POS tags + Chunk tags + Named Entities (Semantic _Syntactic Features)**
One vs. Rest SVM (Support Vector Machines)	82.00 % (410/ 500)	82.00 % (410/ 500)	82.00% (410/ 500)	82.80 % (414/500)
One vs. One SVM (Support Vector Machines)	81.20 % (406/ 500)	81.60 % (408/ 500)	81.00 % (405/ 500)	80.40 % (402/500)
[1] SNoW (Sparse Network of Winnows)	82.60 %	84.90 %	84.00 %	89.3 %

Class Labels = 6 Coarse labels

[1] X. Li and D. Roth, Learning Question Classifiers: The Role of Semantic Information Journal of Natural Language Engineering (2005)

INFERENCES

- Using all the features collected (Bag of Words + POS tags + Chunk tags + Named Entities) 88.0 % coarse question label classification (6 class labels) is obtained and 82.80 % for fine-grained question label classification (47 class labels) gives the best performance with one vs. rest SVM classification
- POS tags on their own don't contribute much to classification performance, Chunk tags with POS tags give a little improvement in performance
- Minor improvements in classifying questions are acquired with Named Entity Semantic features
- Extracting semantic features related to Wordnet Senses, Class-Specific Related words and Distributional similarity might provide better contribution compared to Named Entities

CONCLUSION

- This project explores a machine learning approach to question classification as a multiclass Classification with 6 coarse labels and 47 fine labels
- The classification is achieved using multiclass SVM strategies with features extracted representing Bag of Words, POS tags, Chunk tags and Named Entity features.

REFERENCES

- [1] X. Li and D. Roth, Learning Question Classifiers: The Role of Semantic Information Journal of Natural Language Engineering (2005).
- [2] Dragomir R. Radev, John Prager, and Valerie Samn. Ranking suspected answers to natural language questions using predictive annotation. In Proceedings of the 6th Conference on Applied Natural Language Processing, Seattle, WA, May 2000.
- [3] Boris Katz and Jimmy Lin. Selectively Using Relations to Improve Precision in Question Answering. Proceedings of the EACL-2003 Workshop on Natural Language Processing for Question Answering, April, 2003.
- [4] Voorhees, E. 2002. Overview of the TREC-2002 question answering track. In Proceedings of the 11th Text Retrieval Conference, NIST, pages 115–123
- [5] Dan Jurafsky and James Martin, “Speech and Language Processing”, Prentice Hall; 2nd edition (May 16, 2008).
- [6] K. P. Murphy, “Machine Learning : A Probabilistic Perspective”, The MIT Press, 2013

THANK YOU

Deep representation of Data for Similar Question Retrieval

Shaojin Ding

Department of CSE

April 20, 2017

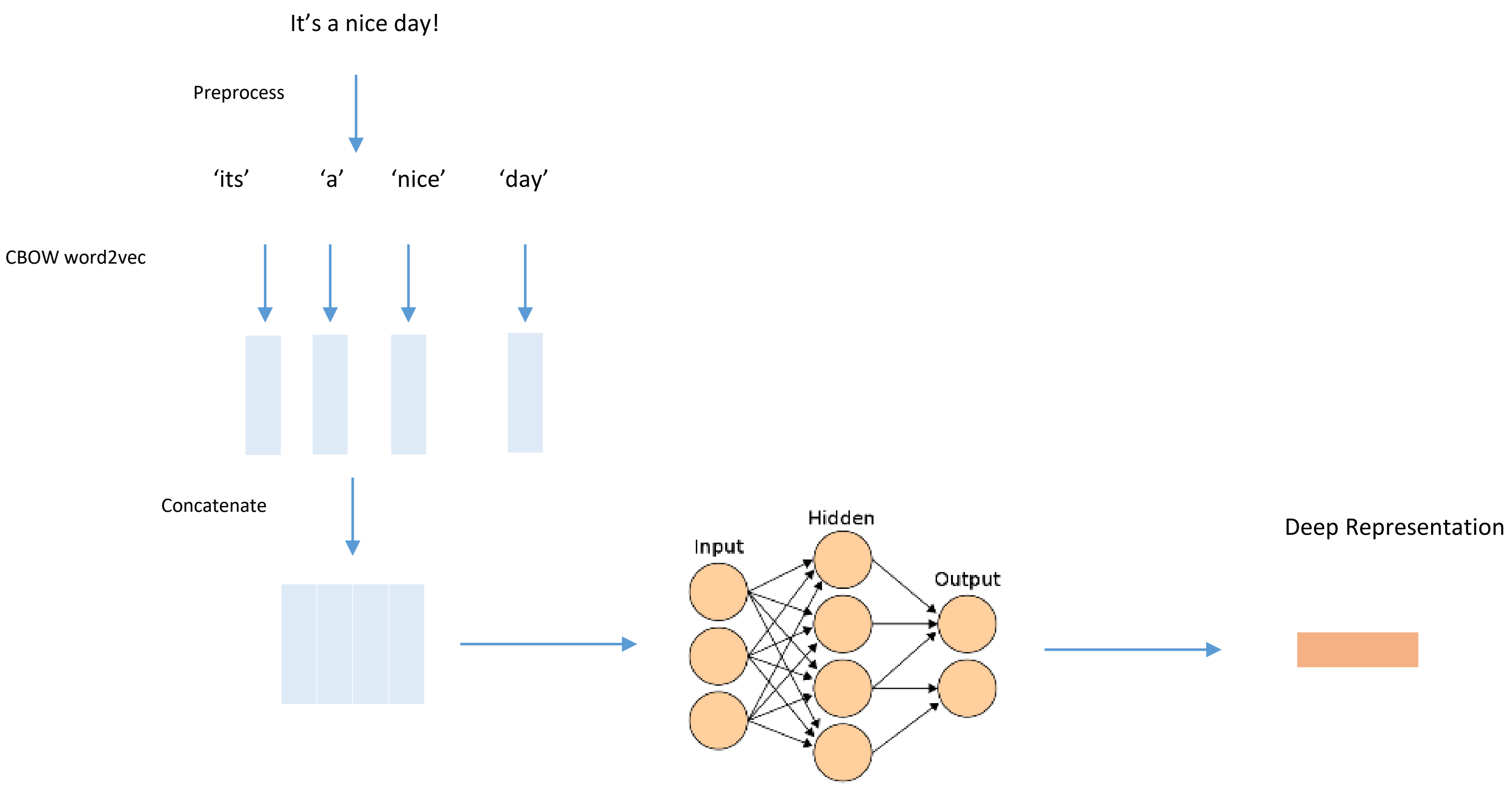
Introduction

- Similar question retrieval is a kernel problem in Community Question Answering.
- Current challenge is there is no reliable data representation for sentences in measuring question similarity
- Objective:
  - Develop a new sentence representation for similar question retrieval

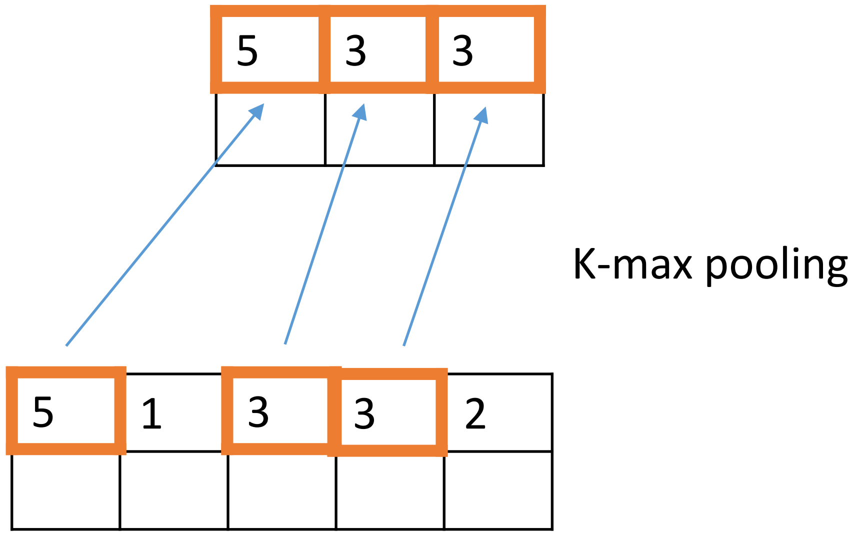
Literature Review

- Translation model
  - Measure the probability of translating one question to another
- Latent topic space
  - learn the similarity between questions in latent topic space from question-answer pairs
- Neural Network
  - Use neural Network to model question-question pair similarity

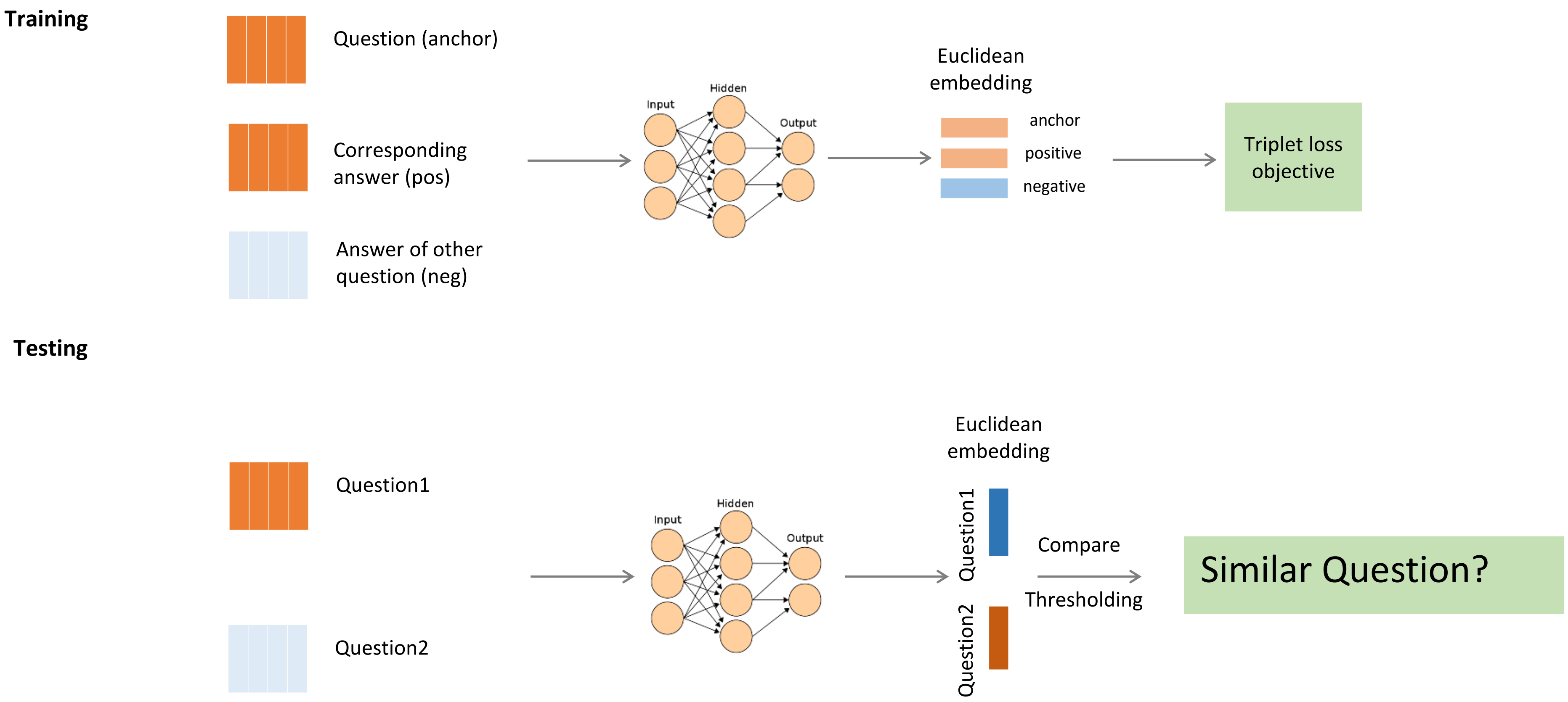
Method Overview



K-max Pooling



Triplet-Net


Training
Question (anchor)
Corresponding answer (pos)
Answer of other question (neg)
Input
Hidden
Output
Euclidean embedding
anchor
positive
negative
Triplet loss objective
Testing
Question1
Question2
Input
Hidden
Output
Euclidean embedding
Question1
Question2
Compare
Thresholding
Similar Question?

Dataset

- Training:
  - Yahoo! Answers Manner Questions, version 2.0
  - 140,000 questions and their corresponding best answers and other answers.

- Testing
  - 24,644 question pairs
  - 9938 pairs are positive and 14706 pairs are negative

Result

System	Recall	Precision	F1-score
BoW + SVD	0.9898	0.4009	0.5707
Proposed	0.8260	0.4650	0.5950

Conclusion

- Bow + SVD representation is not distinguishable
- Proposed method is better than Bow + SVD but not good enough
  - questions and the corresponding answer may not share the same words
  - training data and testing data comes from different source, unseen problem
- Future work
  - Pretrain the model on question classification, then finetune it one similar question retrieval

Thanks

Q&A

CS689 Project

Exploiting community question-answering platforms as a KB for new queries
(SemEval 2017, Task 3)

PRESENTED BY : GIRISH K

The Power of Community QA

- Quora
- Yahoo! Answers
- Stackoverflow
- Pop culture forums
- Facebook threads
- TripAdvisor
- Other localized communities

Problem Statement

Given a new question, and a huge knowledge base of existing QA threads, can we retrieve the most relevant answers to this question?

SemEval 2017, Task 3

- SemEval is a set of competitions held on semantic evaluation where multiple teams build and evaluate systems on expert-annotated data.
- This year, task 3 was on Community Question Answering
- This project addresses two subtasks of SemEval' 17, and emphasize on the first task.

SemEval ’17 : Subtask A

- Given a question, and 10 replies to the question, rank the answers in terms of their relevance to the question

- The training data is annotated with the following labels – “Good”, “Potentially Useful” and “Bad”

Example of Question-Comment Relevance

Question : Where can I get the best Thai Food in College Station?

Answers:

(1) Jins Asian Café is great!

(2) Go to Thailand…

(3) There are some nice food trucks in Northgate, haven't tried them but heard they are good!

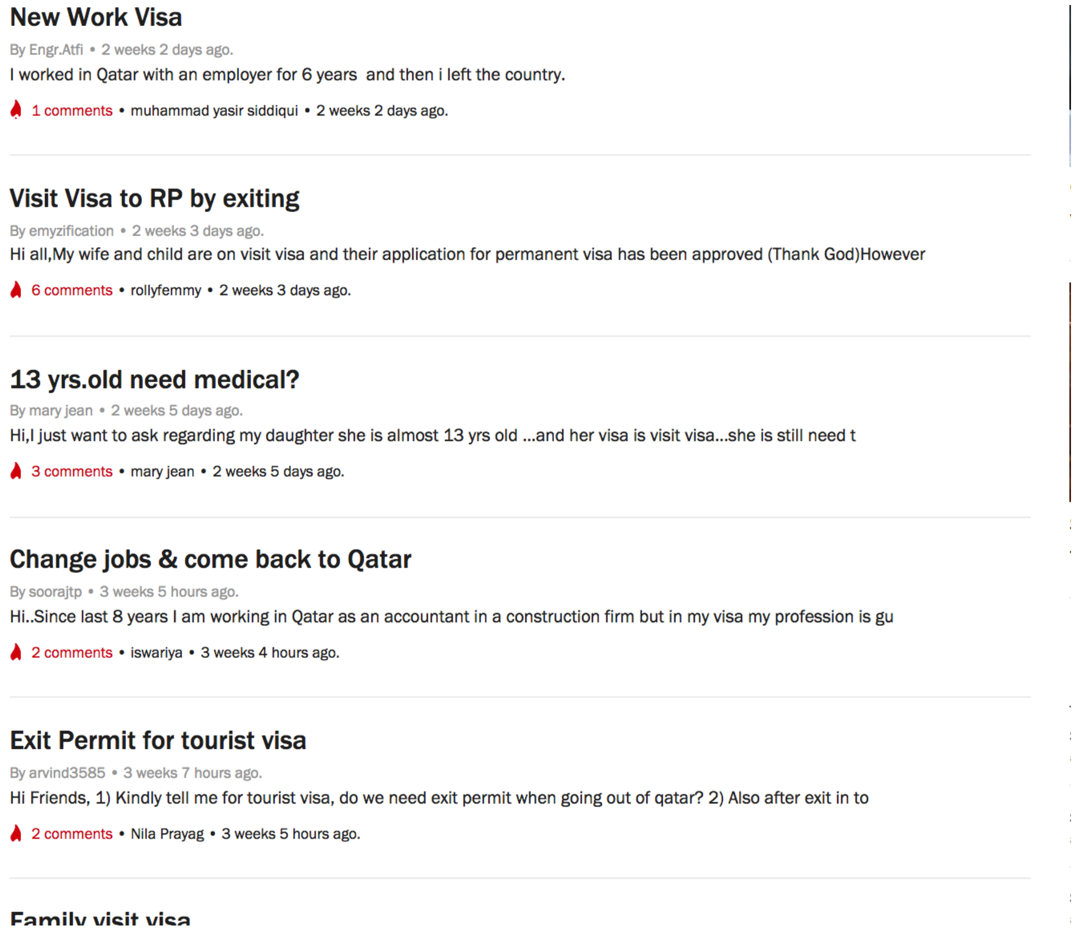
(4) Coaching services for physics, maths and comp. sci – contact 997988232

(5) McDonalds!

Dataset

- Over 40,000 comments on 6000+ questions from Qatar Living Forums

- XML files that need to be processed



Example of Question-Question Relevance

Question : Where can I get the best Thai Food in College Station?

Related Candidates:

(1) In the mood for some Thai cusine, any ideas?

(2) Where can I get the best burgers in College Station?

(3) Where can I find some spicy Asian cuisine?

(4) Searching for Thai Massages, any tips?

(5) Planning a Thai trip soon, any must-see places?

Related Work

- Detecting experts in community forums
- Identifying spam in communities
- Answer ranking for Yahoo Answers!
- Machine Reading (SQuAD and Microsoft MARCO)

Approach

❖ For each question – comment pair, extract 5 sets of features:

1. Lexical Features (word count, ARI and Flesch reading score, punctuation count)
2. General Thread Features (time, no_replies, user_reputation, etc.)
3. Text-based Similarity Features (cosine similarity, set of common words)
4. Syntactic Features (noun phrase count, ners, part-of-speech tags)
5. Centroid of Word Embedding Features (200-dimensional vectors trained on the entire Qatar Living Forum data and available at https://github.com/tbmihailov/semeval2016-task3-cqa)

Training Algorithms used

1. Logistic Regression
2. Support Vector Machine
3. Random Forest
4. Adaboost (SAMME.R)s

For all these algorithms, we use the dev set to optimize over hyperparameters using Grid Search

Ranking using Classifiers

- Given a question and list of answers to it , each of the probabilistic classifiers outputs a probability of a <question, comment> pair being RELEVANT.

- We use this probability as input to a softmax layer to compute the relevance score to rank the comments

Evaluation

- Classification Scores – the usual accuracy, precision, recall, F-Score
- Ranking Scores (used to assess IR systems)
  - **Mean Average Precision –** Precision@K averaged out over a range of k, for multiple queries
  - Mean Reciprocal Rank
  - Average Recall

Results

Approach	Mean Average Precision
IR Baseline (Provided by SemEval Organizers)	0.726
Only Cosine Similarity	0.632
Cosine Similarity + General Thread Features + Lexical Features	0.835
All features + AdaBoost	**0.8635**
All features + SVM	0.847
All features + Random Forest	0.851
All features + Logistic Regression	0.856

Comparision with other teams in SemEval 2017

Team	MAP on test set
KeLP	0.8843
Beihang-MSRA	0.8824
IIT-UHH	0.8688
ECNU	0.8672
BUNJI	0.8658
EICA	0.8653
This Work	**0.8635**
SwissAlps	0.8624
FA3L	0.8342
SnowMan	0.8184
TakeLab-QA	0.8114
LS2NSEMEVAL	0.8099
qwaider	0.7856

Challenges

- Every dataset is unique, and training models that overfit one task does not address the requirements of true QA systems
- Over-reliance on annotated data for training, not to mention opinions can be highly subjective
- Building efficient models that can handle web scale forum data
- One can argue that the task is nothing beyond conventional vector space document retrieval

Conclusion

- There is a huge trade-off between the speed of Information Retrieval, and the complexity of Natural Language Processing.

- Using these massive volumes of expert QA data to train systems can really help us in building truly intelligent QA systems that can understand what a good answer means.

Future Work

- Suggest a domain independent approach that does not rely heavily on annotated data

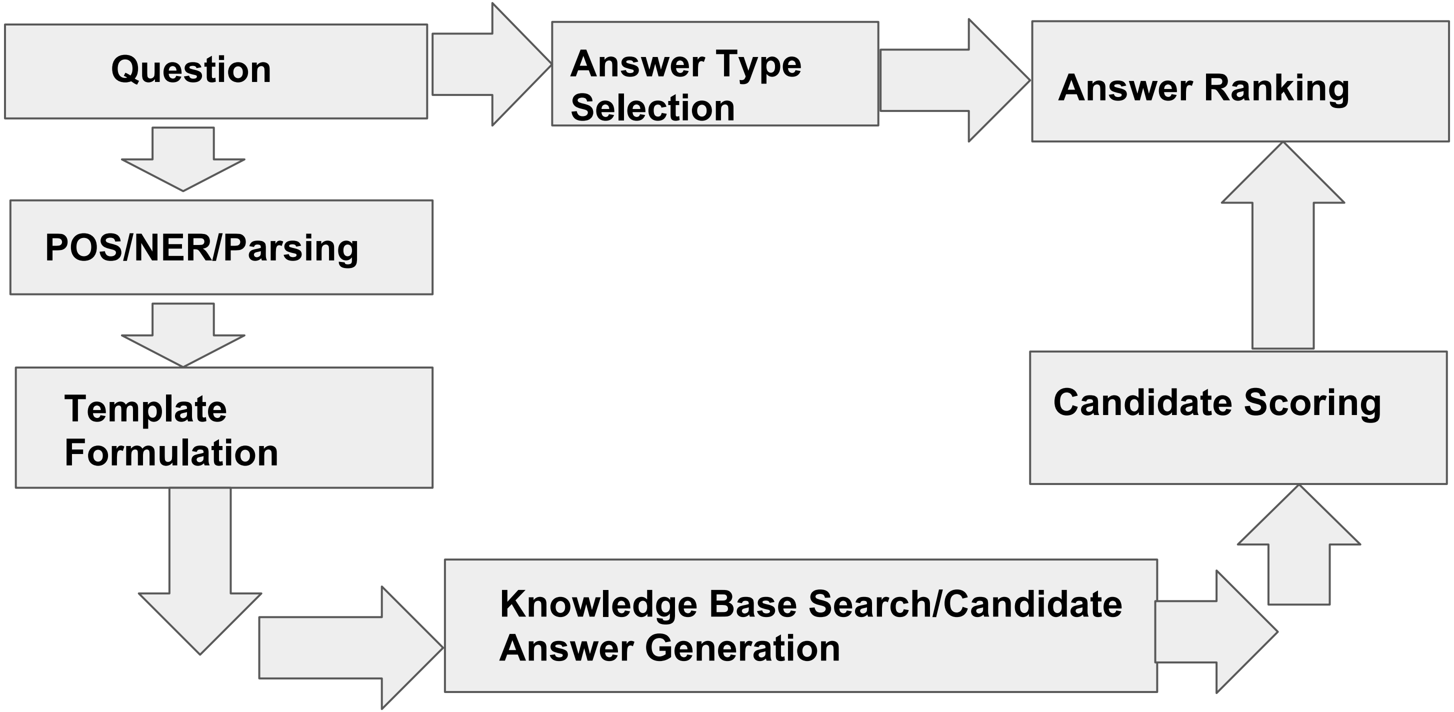
- Improve pipeline to support hyperparameter learning on MAP as opposed to accuracy on dev set

- Identify a set of better semantic features or parse features that act as strong indicators of similarity between a question and answer

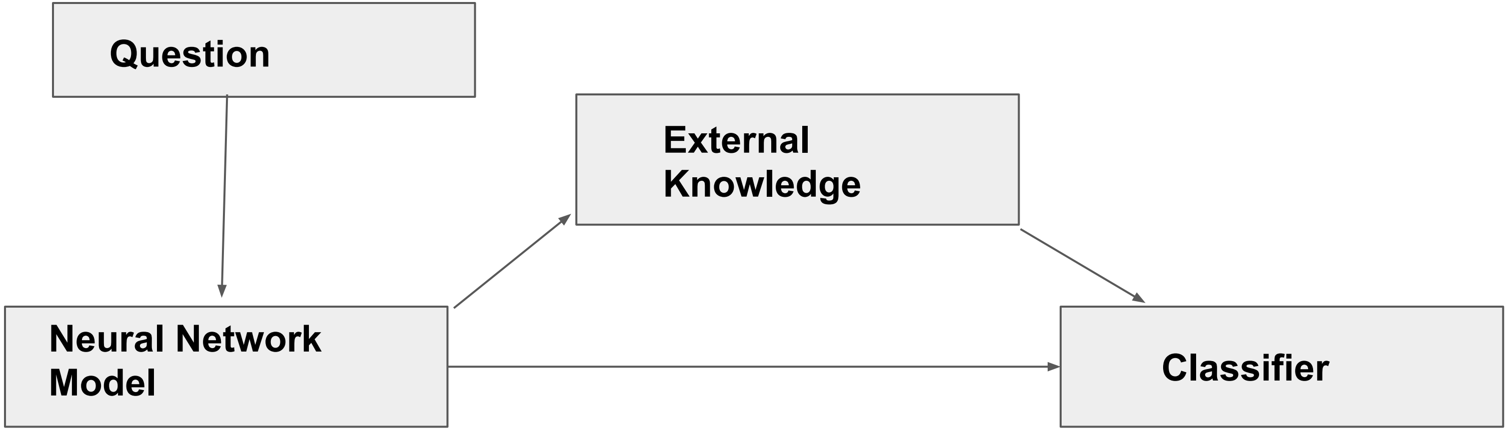
Question Answering Using Deep Learning

Nitin Bansal
Karthik Suresh
Dr. Ruihong Huang

Information Retrieval based Method



Neural Network Model



Fundamentals of NN models for NLP-Word Vectors

- Word vectors are dense vector representations for each word in the vocabulary.
- It is better compared to other word representations because it captures semantic relations between words

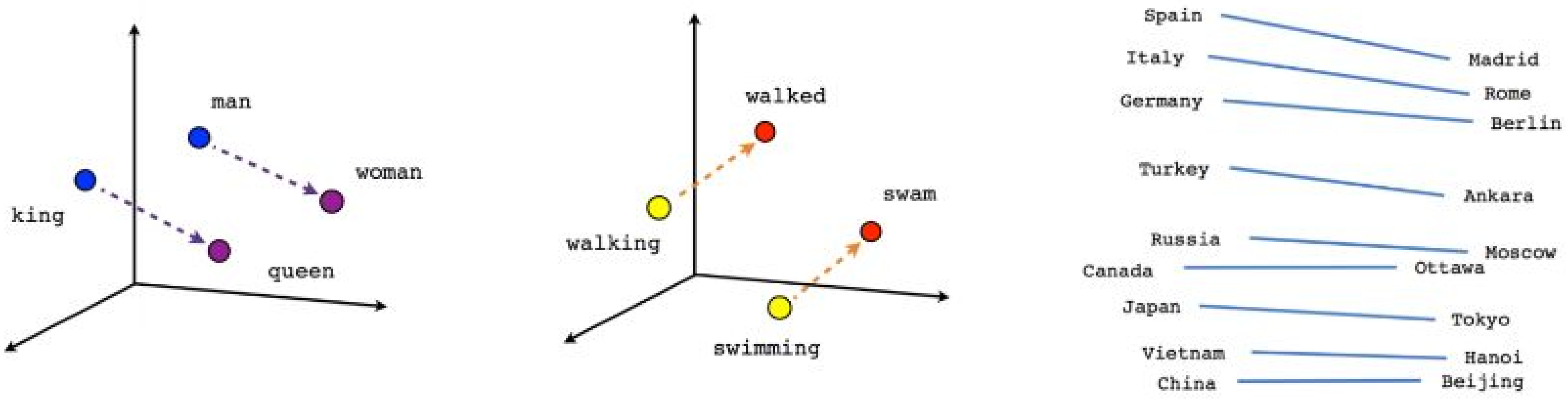


Image Source: https://www.tensorflow.org/tutorials/word2vec

Sequence Modelling using Neural Networks

- Vanilla Neural Network models cannot be used for sequential data like text or voice.
- Hence, a sequential model with a feedback component called a Recurrent Neural Network (RNN) is used
- Feedback helps the model to “remember” the previous inputs

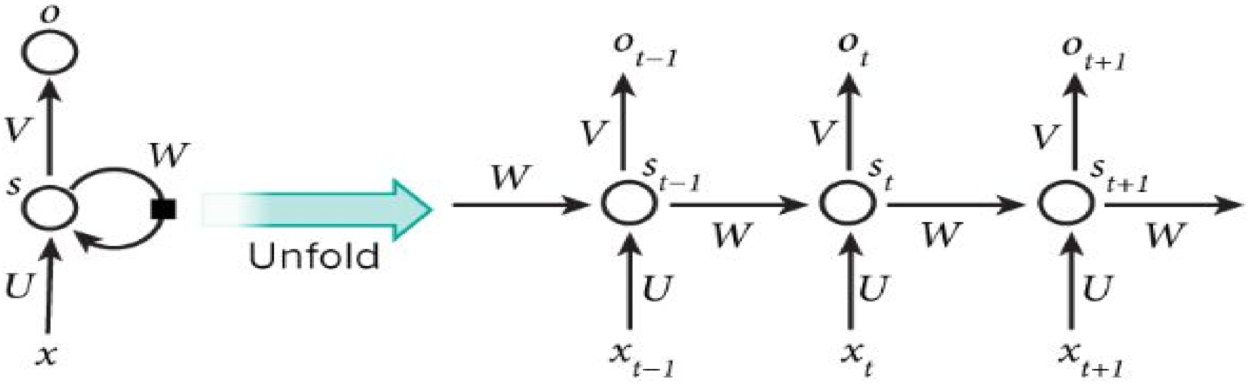


Image Source: Nature

Problems with RNN

- Vanilla RNNs have difficulties capturing long-term dependencies in the data.
- This is because of the vanishing/exploding gradient problem when training the RNN
- This can prove to be costly, especially in case of NLP tasks where long-term dependencies are common
- Hence, variants of RNNs such as Long Short Term Memory (LSTM), Gated Recurrence Unit (GRU) etc were developed

LSTMs and GRUs

- LSTMs have gated structures which enable them to capture long-term dependencies in the data
- GRUs are variants of LSTM where the input and forget gates are combined into a single update gate.

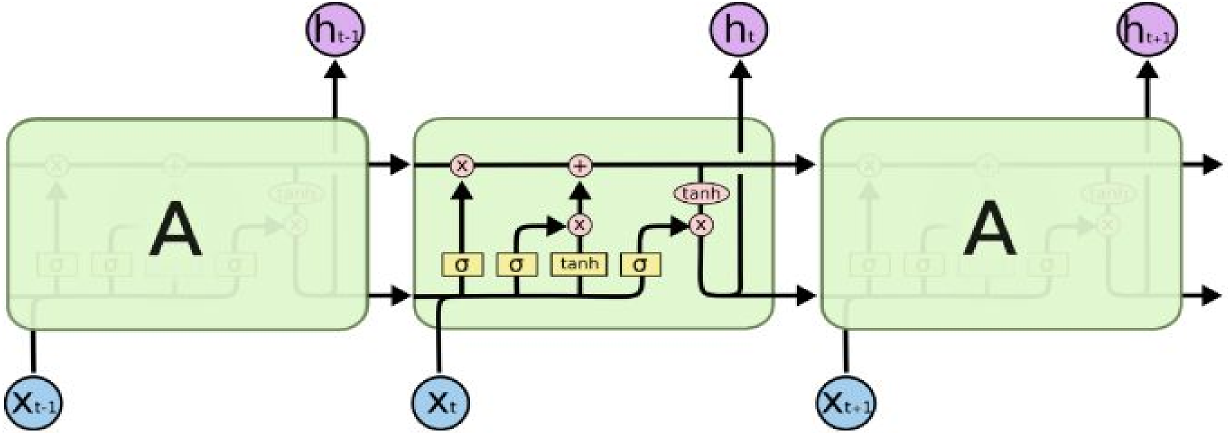


Image source: Colah’s blog

The bAbi dataset

- bAbi is a synthetic dataset created by Facebook.
- It contains 20 tasks each varying in type and difficulty
- Each task consists of a story, a query and an answer. Additionally, the training dataset consists of supporting fact IDs.
- For example, the second task in the dataset looks like the below:

1 John moved to the bedroom.
2 **Mary grabbed the football there.**
3 Sandra journeyed to the bedroom.
4 Sandra went back to the hallway.
5 Mary moved to the garden.
6 **Mary journeyed to the office.**
7 Where is the **football**? office 2 6

Image source: Smerity

QA model with RNN, LSTM and GRU

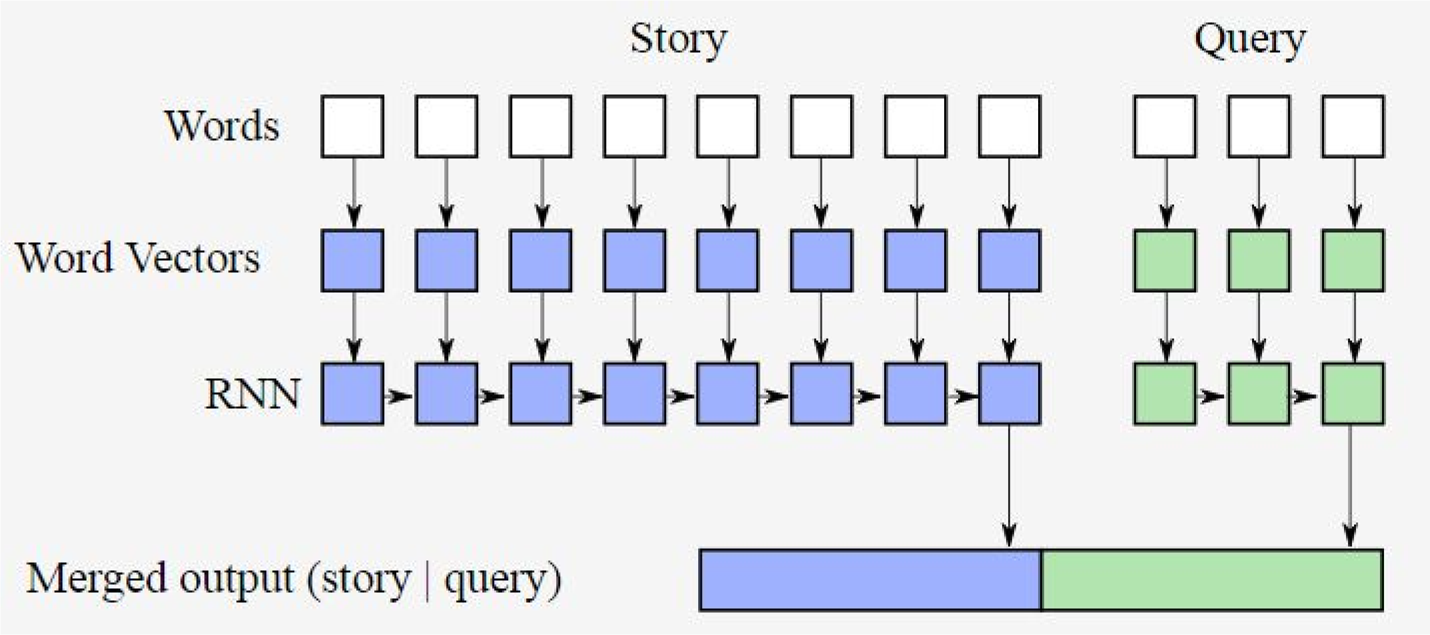


Image source: Smerity

End-to-End Memory Networks

- Networks being able to retain long-term dependencies is not enough to guarantee good efficiency
- We need networks to be able to focus on the important parts of the story for answers. This is called attention mechanism
- End-to-End networks(MemN2N) use soft attention mechanism and form a differentiable model. Hence, they can easily be trained by backpropagation

MemN2N Model

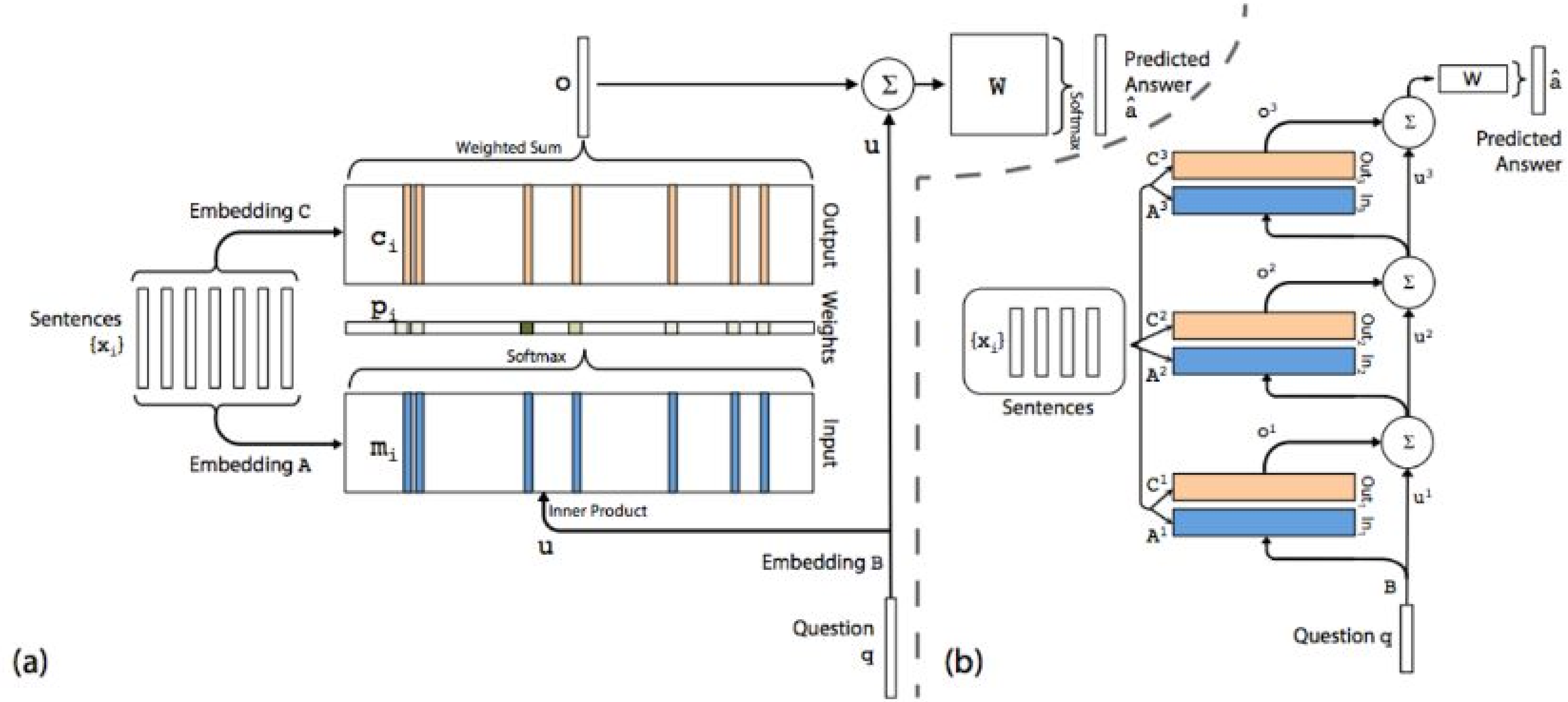
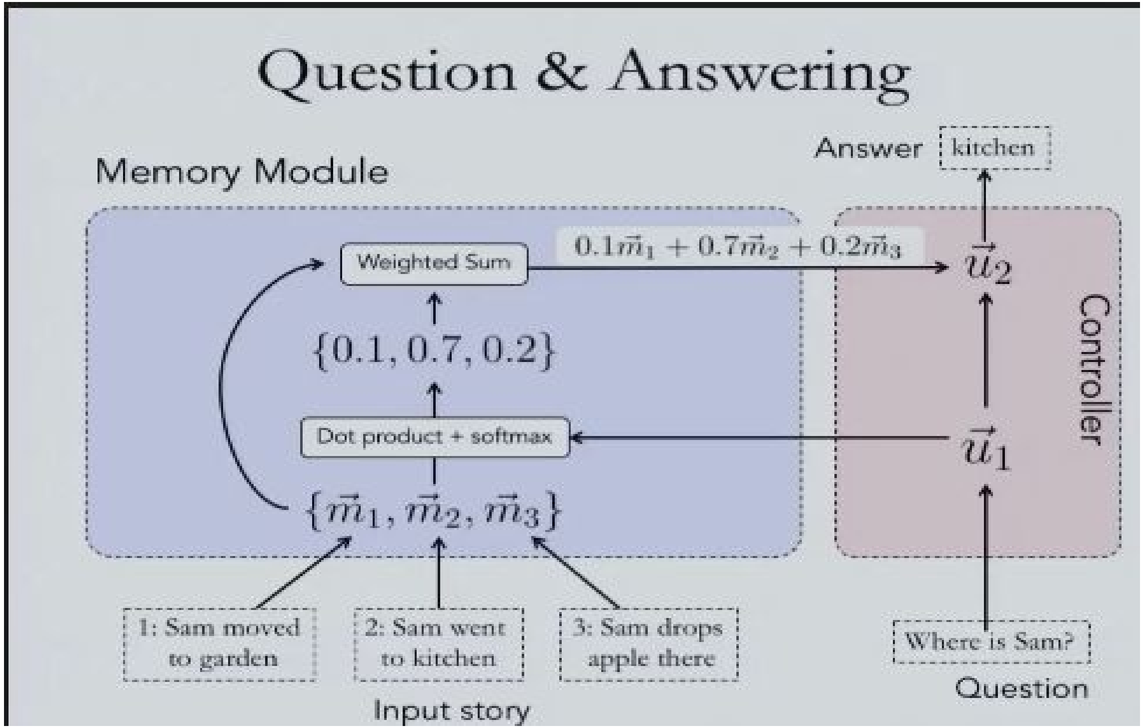
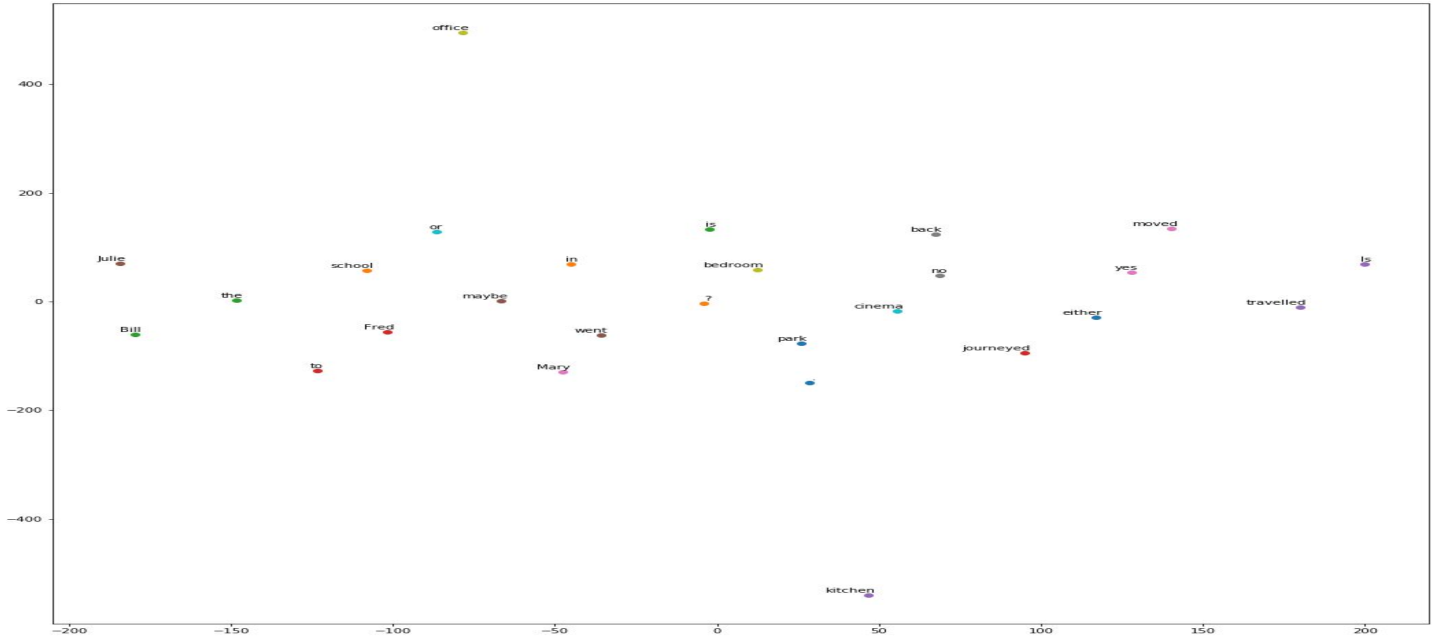


Image Source: “End-to-End Memory Networks by Sukhbaatar et al.

Mem2NN Example



t-SNE Representation for task number 10



Results

TASK	GRU	RNN	LSTM	MemNN
single-support-fact	52.60	52.70	35.50	41.20
two-support-facts	32.30	27.30	34.00	18.70
three-support-facts	13.70	16.80	16.10	20.30
two-arg-relations	21.70	39.90	19.10	66.60
three-arg-relations	52.30	50.10	28.30	52.70
yes-no-questions	69.90	63.00	50.30	69.60
counting	66.90	46.30	56.90	74.80
lists-sets	53.70	33.10	25.00	65.60
simple-negation	64.40	73.50	63.80	64.20

Results-Continued

Indefinite-knowledge	41.30	54.20	43.90	45.70
basic-coreference	75.10	75.10	75.10	38.10
conjunction	77.20	77.20	44.80	34.60
compound-coreference	94.40	94.40	94.40	37.90
time-reasoning	25.00	25.50	25.30	43.10
basic-deduction	17.40	21.90	20.50	52.50
basic-induction	50.10	40.50	44.10	45.40
positional-reasoning	52.00	52.70	48.00	58.50
size-reasoning	50.80	72.50	50.00	91.60
path-finding	11.10	9.80	9.6	9.30
agents-motivations	93.40	90.90	61.30	90.30

Acknowledgements

We are thankful to **Prof. Huang** for constant support and her permission to use the high computing facility.

We are thankful to **Jason Weston**, the lead author of “Memory Networks” and **Sainbayaar Sukhbaatur**, the lead author of “End-to-End Memory Networks” for their valuable suggestions.

We are also thankful to **TAMU High Performance Research in Computing** for letting us run our code on the supercomputer.