Question Answering for Machine Reading
with Lexical Chain

Ling Cao, Xipeng Qiu and Xuanjing Huang

{11210240044, xpqiu, xjhuang}@fudan.edu.cn

School of Computer Science and Technology

Fudan University, Shanghai 201203, China

Abstract. Question answering for machine reading (QA4MR) is a task to understand the meaning communicated by a text. In this paper, we present our system in QA4MRE[1]. The system follows the steps of reading comprehension as a language learner. Lexical chain is used to estimate the semantic relation between texts. Natural language processing (NLP) techniques are also widely used, such as: POS tagging, name entity recognition, coreference. On the QA4MRE test dataset, our system achieves the c@1 measure of 0.28 and 0.26 for the two submissions, respectively.

Keywords: Question Answering; Machine Reading; Lexical Chain; WordNet; Natural Language Processing

1   Introduction

Machine Reading [1] is the automatic, unsupervised understanding of texts, which builds a bridge between natural language and knowledge understandable by machines. The Machine Reading task focus on the deep understanding of small number of texts, which is different from text mining [2], where the system reads and extract knowledge from hundreds or thousands of texts.

Question answering for machine reading (QA4MR) [12] is a task to answer questions by reading of single documents. To understand the meaning of a text on semantic level, system should identify a set of multiple choices related to it, where correct answers require inference in all kinds, i.e., lexical (acronymy, synonymy, hyperonymy), syntactic (nominalization / verbalization, causative, paraphrase, active/passive), discourse (coreference, anaphora ellipsis) [3].

Here is an example for QA4MRE task:

Text:

Annie Lennox: Why I am an HIVAIDS activist, I'm going to share with you the story as to how I have become an HIV/AIDS campaigner. And this is the name of my

---

campaign, SING Campaign.

Question:

Who is the founder of the SING campaign?

Candidate Answers:

1) Nelson Mandela

2)Youssou N'Dour

3)Michel Sidibe

4)Zackie Archmat

5)Annie Lennox

By machine reading, the answer "*Annie Lennox*" could be chosen.

In this paper, we propose a method for question answering for machine reading system with lexical chain. Our system is similar to the scenario of humans learning a new language and dealing with reading comprehension tests. Humans always do reading comprehensions in 3 steps:

1. **Locating**: Reading the question and extracting sentences from the passage which may related to the question.
2. **Answering**: Reading these sentences in details to select which sentences are most likely to be the answer.
3. **Choosing**: Reading all the choices and choosing the one has the same meaning as the answer.

So our system explores the possibilities of following the above steps as a language learner doing Reading Comprehensions. Lexical chain [6], proposed by LCC, performs well in finding topic relations between words based on WordNet. Our system uses lexical chain to estimate the semantic relation between texts.

The rest of the paper is organized in the following way: Section 2 is to present a brief overview of related works. Section 3 provides the architecture of our system. Section 4 introduced lexical chain in details. We present our experiments and results in section 5, and conclude in section 6.

2 Related Works

QA4MR [12] is related to some topics in the fields of information retrieval and natural language processing (NLP), such as question answering [11], reading comprehension [9] and recognizing textual entailment [7].

Question answering (QA) [8] is the task of automatically answering a question posed in natural language. In contrast to QA4MR, QA systems are designed to extract answers in large corpus. QA systems seldom focus on the deeply understanding of corpus. What's more, QA systems tend to answer every question as they can even though they might not confident about the correctness of the answers.

Reading comprehension (RC) [9] system attempts to understand a document and returns an answer sentence when posed with a question. RC resembles the ad hoc question answering (QA) task that aims to extract an answer from a collection of documents when posed with a question. However, since RC focuses only on a single document, the system needs to draw upon external knowledge sources to achieve deep analysis of passage sentences for answer sentence extraction.

Recognizing Textual Entailment (RTE) [7] has been proposed recently as a generic task that captures major semantic inference needs across many NLP applications. This task requires to recognize, given two text fragments, whether the meaning of one text is entailed (can be inferred) from the other text. Semantic understanding and logic understanding of text is indispensable for RTE system.

3   System Overview

The framework of our QA4MR system is presented in Fig. 1.
In our system, text (including passage, question and choices) is sent to preprocessing module initially.
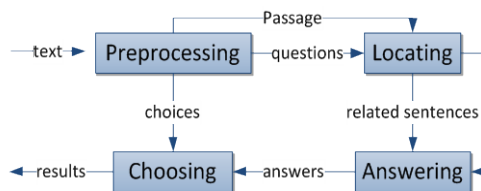

Fig. 1.   System architecture

Preprocessing module do following works:
  (1) Sentence segmentation for each passage.
  (2) Tokenization for each sentence, question and choice.
  (3) POS and parsing.
  (4) Named entities recognition.
  (5) Coreference annotation for each pronominal or abbreviate expression.
  Our system uses Illinois coreference package[2] [4] to preprocess text.

After preprocessing, annotated text is separated into 3 pieces: passage, question and choices. The passage and question are then sent to locating module. As described in section 1, locating module extracts sentences from passage which are related to the question.

While doing reading comprehension, language learners always locate sentences in a passage by the name entities mentioned in the question, such as a name of a person, a special place or an organization. For instance, in test set of QA4MRE task [3], passage1, question 2:

Who is the founder of the SING campaign?

To answer this question, language learners will focus on the organization "SING campaign". While reading, they will only pay attention to sentences talking about "SING campaign" to find the answer and neglect the others. So in locating module, system extracted all these sentences such as:

---

[2] http://cogcomp.cs.illinois.edu/page/software_view/18

And this is the name of my campaign, SING Campaign.
And yes, my SING Campaign has supported Treatment Action Campaign in the way that I have tried to raise awareness and to try to also raise funds.
SING Campaign is basically just me and about three or four wonderful people who help to support me.

It is worth to say that we can find out that the founder of SING campaign is the author of the passage (Annie Lennox) only in the sentences above. If without the recognition of name entities, the following sentence might mislead the answer of this question semantically, which is talking about the foundation of a campaign, but not SING campaign:

I was very very fortunate, a couple of years later, to have met Zackie Achmat, the founder of Treatment Action Campaign, an incredible campaigner and activist.

Sentences extracted by locating module are then submitted to answering module, with the question. Answering module, as its name implies, gives the answer to the question. Instead of giving an answer directly, answering module reads every sentence from input and select which one is most likely to be the answer. The method of the selection is by lexical chain. Details of lexical chain are presented in section 3.

In the last step, answer sentences extracted by answering module are submitted to choosing module with the choices of the question. Choosing module gives the final result by name entity matching and lexical chain.

4   Lexical Chain

As is mention in previous, lexical chain [6] is used in answering module and choosing module to estimate semantic relation between texts. WordNet [5] [10] relations including: Hypernym, Hyponym, Synonym, Meronym, Holonym, Attribute, Cause, Entailment. In addition, to build connection between synsets of different POS, gloss relation, defined by 6 is used.

Following lexical chain [6], which is put forward by LCC, system scores the semantic relation between words by the total score of each relation path. Equation (1) shows how system scores words:

$$Relation(s_i, s_j) = \sum_i Score(r_i) \qquad (1)$$

where:

$$Score(r) = I \times \prod_{i=1}^{length(r)} (W_{Ri} * MG_{Ci}) \qquad (2)$$

where:
I is the initial score.
WRi is the weight of the relation presented in table 1.
MGci can be calculated by equation (3).

Table 1. Weight of WordNet relations

WordNet Relation	Weight	WordNet Relation	Weight
Hypernym	0.8	Attribute	0.5
Hyponym	0.7	Cause	0.5
Synonym	0.9	Entailment	0.7
Meronym	0.5	Gloss	0.6
Holonym	0.5	R-Gloss	0.2

$$MG_C = \frac{CONST}{CONST + N_{R-Gloss}} \qquad (3)$$

where Nr-gloss is the total amount of r-gloss relation of a word (r-gloss relation is reverse to gloss relation).

Semantic relation between texts can be estimated by the score of their key words. For example, in passage 1, question 5:

What is Annie Lennox's profession?

Sentences are extracted by locating module, among them, following two sentences are most likely to be the answer:

……because I am a woman, and I am a mother……

……talking and using my platform as a musician, with my commitment to

Mandela......

With WordNet gloss relation, system finds out that musician is a kind of profession. As the gloss of musician#2 is "artist who composes or conducts music as a profession". While mother is not a kind of profession.

5 Experiments and Results

The evaluation measure of QA4MR task is C@13. This measure rewards systems that, while maintaining the number of correct answers are able to reduce the incorrect ones by leaving some questions unanswered.

C@1 measure is presented in Equation (4):

$$C@1 = \frac{(nr + nu*(nr/n))}{n} \qquad (4)$$

where:

---

[3] Anselmo Peñas, Pamela Forner, Richard Sutcliffe, Alvaro Rodrigo, Corina Forascu, Iñaki Alegria, Danilo Giampiccolo, Nicolas Moreau and Petya Osenova. Overview of ResPubliQA 2009: Question Answering Evaluation over European Legislation. In C. Peters, G. Di Nunzio, M. Kurimo, Th Mandl, D. Mostefa, A. Peñas, G. Roda. (Eds) Multilingual Information Access Evaluation I. Text Retrieval Experiments, 10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009, Corfu, Greece, Revised Selected Papers. Lecture Notes in Computer Science 6241. Springer-Verlag, 2010.

nr: is the number of correctly answered questions
nu: is the number of unanswered questions
n: is the total number of questions

We evaluates 2 runs on the dataset of QA4MRE 2011[3]. The significant difference between two runs is the processing of locating module. The first run keeps the question unanswered if system failed in locating, that is, failed to recognize any name entity in the question, while the second run send all sentences in the passage to answering module. And also, the length of the lexical chain in the second run is less than the first run by 1.
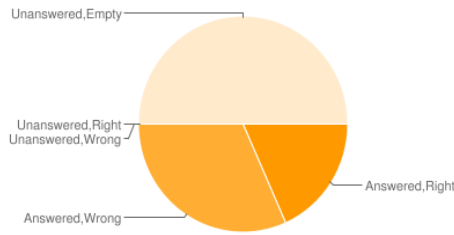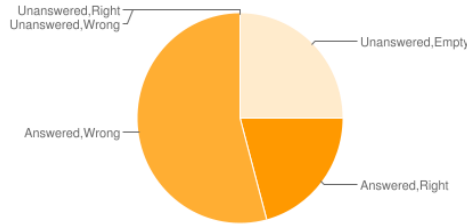

Fig. 2.   Result of submission 1

Fig. 3. Result of submission 2

Table 2. Overall C@1 measure

	Run 1	Run 2	Best	Worst	Average
C@1	0.28	0.26	0.57	0.02	0.21

Table 3. C@1 measure by topics

Topics	Run 1	Run 2
AIDS	0.28	0.26
Climate Change	0.19	0.17
Music and Society	0.36	0.34

Table 4. Evaluation at question-answering level

	Run 1	Run 2
Correctly answered	22	25
Wrongly answered	38	65
unanswered	60	30

Table 2 shows the overall C@1 measure of our two results, comparing with the best run, the worst run and the average C@1 measure for QA4MRE. Table 3 illustrates more details about the C@1 measure of each topic. Table 4, Fig. 2 and Fig. 3 show the result of both run at question-answering level. According to Fig. 2 and Fig. 3, system does not performance well without the locating of name entities.

Notice that 30 questions are not answered in both runs, for two reasons. One is that current approach gives up all questions about date, time and digit such as: For how long did people applaud at performances of the Bolivar Youth Orchestra? The other is our current approach do not use background collection, which leads to the failure in answering questions such as: What is Nelson Mandela's country of origin? (post-apartheid Rainbow Nation).

6 Conclusions and Future Works

This paper introduced the architecture of our system in QA4MRE task. System chooses correct answers by name entity locating and lexical chain. Experiments revealed that name entity locating reduce the rate of wrong answers misleaded by lexical relation, and lexical chain estimate the semantic relation of texts by scoring the lexical relation of words.

Our current approach gets an overall c@1 measure as 0.28. In the future, we will have two further works to do. One is developing a module recognizing date, time and digit. The other is trying to use world knowledge in our system, including background collection, and web-based resources such as Wikipedia.

Acknowledgement

References

1. Etzioni, M. Banko, and M. J. Cafarella. Machine reading. In Proceedings of the 21th National Conference on Artificial Intelligence (AAAI)2006.

2. Ah-hwee Tan, Text Mining: The state of the art and the challenges, in Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases, 1999.

3. Question Answering for Machine Reading Evaluation at CLEF 2011. http://celct.fbk.eu/QA4MRE/

4. Bengtson, Eric and Dan Roth. Understanding the Value of Features for Coreference Resolution. In proc. EMNLP-08. 2008.

5. Christiane Fellbaum (1998, ed.) WordNet : An Electronic Lexical Database. Cambridge, MA: MIT Press.

6. Dan I. Moldovan and Adrian Novischi. 2002. Lexical chains for Question Answering. In Proceedings of COLING, Taipei, Taiwan, August.

7. Danilo Giampiccolo, Bernardo Magnini, Ido Dagan,Bill Dolan. 2007. The Third PASCAL Recognizing Textual Entailment Challenge, In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, Prague, Czech Republic.

8. Ellen M. Voorhees. Overview of the TREC 2002 Question Answering Track (2002). In Proceedings of the Eleventh Text REtrieval Conference

9. Lynette Hirschman, Marc Light, Eric Breck, John D. Burger, Deep Read: a reading comprehension system, in Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, p.325-332, June 20-26, 1999.

10. George A. Miller (1995). WordNet : A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.

11. John Prager, Eric Brown, Anni Coden, and Dragomir Radev. Question-answering by predictive annotation. In Proceedings, 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece, July 2000.

12. Lucy Vanderwende, Answering and Questioning for Machine Reading, American Association for Artificial Intelligence, March 2007.