Rhetoric Map of an Answer to Compound Queries

Boris Galitsky
Knowledge Trail Inc.
San-Francisco, USA

bgalitsky@hotmail.com

Dmitry Ilvovsky
National Research University Higher School of Economics, Moscow, Russia

dilvovsky@hse.ru

Sergey O. Kuznetsov
National Research University Higher School of Economics, Moscow, Russia

skuznetsov@hse.ru

Abstract

Given a discourse tree for a text as a candidate answer to a compound query, we propose a rule system for valid and invalid occurrence of the query keywords in this tree. To be a valid answer to a query, its keywords need to occur in a chain of elementary discourse unit of this answer so that these units are fully ordered and connected by nucleus – satellite relations. An answer might be invalid if the queries' keywords occur in the answer's satellite discourse units only. We build the rhetoric map of an answer to prevent it from firing by queries whose keywords occur in non-adjacent areas of the Answer Map. We evaluate the improvement of search relevance by filtering out search results not satisfying the proposed rule system, demonstrating a 4% increase of accuracy with respect to the nearest neighbor learning approach which does not use the discourse tree structure.

1 Introduction

Answering compound queries, where its keywords are distributed through text of a candidate answer, is a sophisticated problem requiring deep linguistic analysis. If the query keywords occur in an answer text in a linguistically connected manner, this answer is most likely relevant. This is usually true when all these keywords occur in the same sentence: they should be connected syntactically. For the inter-sentence connections, these keywords need to be connected via anaphora, refer to the same entity or sub-entity, or be linked via rhetoric discourse.

If the query keywords occur in different sentences, there should be linguistic cues for some sort of connections between these occurrences. If there is no connection, then different constraints for an object expressed by a query might be applied to different objects in the answer text, therefore, this answer is perhaps irrelevant. There are following possibilities of such connections.

Anaphora. If two areas of keyword occurrences are connected with anaphoric relation, the answer is most likely relevant.

Communicative actions. If the text contains a dialogue, and some question keywords are in a request and other are in the reply to this request, then these keywords are connected and the answer is relevant. To identify such situation, one needs to find a pair of communicative actions and to confirm that this pair is of request-reply kind.

Rhetoric relations. They indicate the coherence structure of a text (Mann and Thompson, 1988). Rhetoric relations for text can be represented by a Discourse tree (DT) which is a labeled tree. The leaves of this tree correspond to contiguous units for clauses (elementary discourse units, EDU). Adjacent EDUs as well as higher-level (larger) discourse units are organized in a hierarchy by rhetoric relation (e.g., background, attribution). Anti-symmetric relation takes a pair of EDUs: nuclei, which are core parts of the relation, and satellites, the supportive parts of the rhetoric relation.

The most important class of connections we focus in this study is rhetoric. Once an answer text is split into EDUs, and rhetoric relations are established between them, it is possible to establish rules for whether query keywords occurring in text are connected by rhetoric relations (and therefore, this answer is likely relevant) or not connected (and this answer is most likely irrelevant). Hence we use the DT as a base for an ***Answer Map*** of a text: certain sets of nodes in DT correspond to queries so that this text is a valid answer, and certain sets of nodes correspond to an invalid answer. Our definition of the Answer Map follows the methodology of inverse index for search: instead of taking queries and considering all valid answers for it from a set of text,

we take a text (answer) and consider the totality of valid and invalid queries consisting of the keywords from this text.

Usually, the main clause of a compound query includes the main entity and some of its constraints, and the supplementary clause includes the other constraint. In the most straightforward way, the main clause of a query is mapped into a nucleus and the supplementary clause is mapped into a satellite of RST relation such as *elaboration*. Connection by other RST relation, where a satellite introduces additional constraints for a nucleus, has the same meaning for answer validity. This validity still holds when two EDUs are connected with a symmetric relation such as joint. However, when the images of the main and supplementary clause of the query are satellites of different nucleus, it most likely means that they express constraints for different entities, and therefore constitute an irrelevant answer for this query.

There is a number of recent studies employing RST features for passage re-ranking under question answering (Joty and Moschitti, 2014; Surdeanu et al., 2014). In the former study, the feature space of subtrees of parse trees includes the RST relations to improve question answer accuracy. In the latter project, RST features contributed to the totality of features learned to re-rank the answers. In (Galitsky et al., 2014) rhetoric structure, in particular, was used to broaden the set of parse trees to enrich the feature space by taking into account overall discourse structure of candidate answers. Statistical learning in these studies demonstrated that rhetoric relation can be leveraged for better search relevance. In the current study, we formulate the explicit rules for how a query can be mapped into the answer DT and the relevance of this map can be verified.

2   Example of an Answer Map

Ex. 1. DT including 6 nodes {e1...e6} is shown in Fig 1 (Joty and Moschitti, 2014). Text is split into six EDUs:

```
[what's      more,]e1      [he      be-
lieves]e2   [seasonal   swings   in
the   auto   industry   this   year
aren't   occurring   at   the   same
time   in   the   past,]e3   [because
of   production   and   pricing   dif-
ferences]e4   [that   are   curbing
the   accuracy   of   seasonal   ad-
justments]e5  ]  [built   into   the
employment data.]e6
```
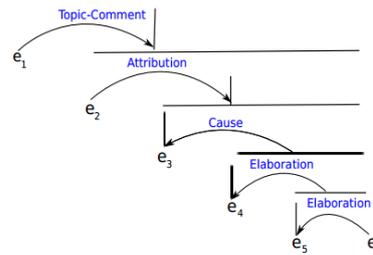

Fig.1. Discourse tree for the Example 1

Horizontal lines indicate text segments; satellites are connected to their nuclei by curved arrows. One can see that this text is a relevant answer to the query

```
Are seasonal swings in the auto
industry due to pricing differ-
ences?
```

but is an irrelevant answer to the query
```
Are  pricing  differences  built
into employment data?
```
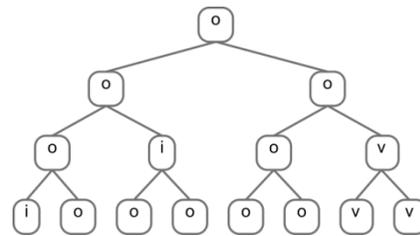

Fig. 2. An Answer Map and its areas for valid and invalid answers

A valid set of nodes of an Answer Map is defined as the one closed under common ancestor relations in a DT. For example, the *i*-nodes on the bottom-left of DT in Fig. 2 constitute the invalid set, and the *v*-nodes on the right of DT constitute the valid set.

Ex. 2.
```
I went to watch a movie because
I had nothing else to do. I en-
joyed the movie which was about
animals finding food in a de-
sert. To feed in a desert envi-
ronment, zebras run hundreds of
miles in search of sources of
water.
```

This answer is valid for the following queries (phrases) since their keywords form *v*-set:
- enjoy movie watched when nothing else to do
- I went to watch a movie about feeding in desert en-vironment
- I went to watch a movie about zebras run hundreds of miles

```
-  I  went  to  watch  a  movie
   about  searching  sources  of
   water
```

And this text is not a correct answer for the
following queries (phrases), since their keywords
form *i*-sets:

```
-  animals  find  food  in  desert
   when  have  nothing  else  to  do
-  I  had  nothing  else  except
   finding  food  in  a  desert
-  I  had  nothing  else  to  do  but
   run  hundreds  of  miles  in
   search  of  water
-  finding  food  in  a  desert  -  a
   good  thing  to  do
```

3 Definition and Construction Algo-rithm

Discourse tree includes directed arcs for anti-
symmetric rhetoric relation and undirected arcs
for symmetric rhetoric relations such as *joint*,
time sequence, and others. For two nodes of DT
we define its directed common ancestor as a
common ancestor node which is connected with
these nodes via directed arcs.

The valid set of EDUs which is a result of
mapping of a query is closed under common di-
rected ancestor relation: it should contain the set
of all directed common ancestor for all EDUs.
Hence this constraint is applied for antisymmet-
ric RST relations; query terms can occur in
symmetric EDU nodes in an arbitrary way.

To construct an Answer Map from DT, firstly,
we need to map keywords and phrases of a query
into EDUs of an answer. For each noun phrase
for a query, we find one or more EDUs which
include noun phrases with the same head noun.
Not each keyword has to be mapped, but there
should be not more than a single EDU each key-
word is mapped under a given mapping. For ex-
ample, noun phrase from the query `family do-
ing its taxes` is mapped into the EDU `in-
cluding how individuals and families
file their taxes` since they have the same
head noun `tax`. If a multiple mapping exists for
a query, we need to find at least one valid occur-
rence to conclude that this query is a valid one
for the given map.

For a query Q, if its keywords occur in candi-
date answer A and the set of EDUs Q_{edu}, then
commonAncestorsDT$(A)(Q_{edu}) \subseteq Q_{edu}$.

For a real-word search system, the enforce-
ment of RST rules occurs at indexing time, since
RST parsing is rather slow.

For answer text A, we produce a sequence of
texts $A_e < \{A$ directed common ancestor I$\}$ for
all pairs of EDU nodes connected with their par-
ents by directed arcs. Then the match of the set
of keyword occurs with the extended index in the
regular manner: there is no element A_e for inva-
lid mapping Q to Q_{edu}.

4 Approach Scalability

In terms of search engineering, enforcing of the
condition of the Rhetoric Map of an answer re-
quires additional part of the index besides the
inverse one. Building this additional index re-
quires enumeration of all maximal sequences of
keywords from Rhetoric Map for every docu-
ment (potential answer A). Once A is determined
to be fired by query Q using the regular search
index, there should be an entry in Rhetoric Map
which is fired by a query formed as a conjunc-
tion of terms in Q.

Since application of Rhetoric Map rules oc-
curs via an inverse index, the search time is con-
stant with respect to the size of the overall RM
index and size of a given document. The index-
ing time is significantly higher due to rhetoric
parsing, and the size of index is increased ap-
proximately by the number of average maximal
paths in a DT graph, which is 3-5. Hence alt-
hough the performance of search will not signifi-
cantly change, the amount of infrastructure ef-
forts associated with RM technology is substan-
tial.

5 Evaluation

We used the TREC evaluation dataset as a list of
topics: http://trec.nist.gov/data/qa/. Given a short
factoid question for entity, person, organization,
event, etc. such as `#EVENT Pakistan earth-
quakes of October 2005#` we ran a web
search and automatically (using shallow parsing
provided by Stanford NLP) extracted compound
sentences from search expressions, such as `A
massive earthquake struck Pakistan
and parts of India and Afghanistan
on Saturday morning October 8, 2005.
This was the strongest earthquake in
the area during the last hundred
years.`

Ten to twenty such queries were derived for a
topic. Those portions of text were selected with
obvious rhetoric relation between the clauses.
We then fed Bing Search Engine API such que-
ries and built the Answer Map for each candidate
answer. We then ran the Answer Map - based

filter. Finally, we manually verify that these filtered answers are relevant to the initial questions and to the queries.

We evaluated improvement of search relevance for compound queries by applying the DT rules. These rules provide Boolean decisions for candidate answers, but we compare them with score-based answer re-ranking based on ML of baseline SVM tree kernel (Moschitti, 2006), discourse-based SVM (Ilvovsky, 2014) and nearest-neighbor Parse Thicket-based approach (Galitsky et al., 2013).

The approach based on SVM tree kernel takes question-answer pairs (also from TREC dataset) and forms the positive set from the correct pairs and negative set from the incorrect pairs. The tree kernel learning (Duffy and Collins, 2002) for the pairs of extended parse trees produces multiple parse trees for each sentence, linking them by discourse relations of anaphora, communicative actions, "same entity" relation and rhetoric relations (Galitsky et al., 2014).

In the Nearest Neighbor approach to question – answer classification one takes the same data of parse trees connected by discourse relations and instead of applying SVM learning to pairs, compare these data for question and answer directly, finding the highest similarity.

To compare the score-based answer re-ranking approaches with the rule-based answer filtering one, we took first 20 Bing answers and classified them as valid (top 10) and invalid (bottom 10) under the former set of approaches and selected up to 10 acceptable (using the original ranking) under the latter approach. Hence the order of these selected set of 10 answers is irrelevant for our evaluation and we measured the percentage of valid answers among them (the focus of evaluation is search precision, not recall).

Answer validity was assessed by Amazon Mechanical Turk. The assessors were asked to choose relevant answers from the randomly sorted list of candidate answers. Table 1 shows the evaluation results.

Table 1. Evaluation results

Filtering method		Baseline Bing search, %	SVM TK learning of QA pairs (baseline improvement), %	SVM TK learning for the pairs for extended parse trees, %	Nearest neighbor for question – answer, %	Answer Map, %
Sources / Query types	Source of discourse information	-	-	Anaphora, same entity, selected discourse relations		Discourse Tree
Clauses connected with *elaboration*		68.3	69.4	73.9	74.6	**79.2**
Clauses connected with *attribution*		67.5	70.1	72.7	75.1	**78.8**
Clauses connected with *summary*		64.9	66.3	70.2	74.0	**78.0**
Clauses in *joint/sequence* relation		64.1	65.2	68.1	72.3	**76.3**
Average		66.2	67.8	71.2	74.0	**78.0**

The top two rows show the answer filtering methods and sources of discourse information. Bottom rows show evaluation results for queries with various rhetoric relations between clauses.

One can observe just a 1.5% improvement by using SVM tree kernel without discourse, further 3.5% improvement by using discourse-enabled SVM tree kernel, and further improvement of 2.8% by using nearest neighbor learning. The latter is still 4% lower than the Answer Map ap-

proach, which is the focus of this study. We observe that the baseline search improvement, SVM tree kernel approach has a limited capability of filtering out irrelevant search results in our evaluation settings. Also, the role of discourse information in improving search results for queries with symmetric rhetoric relation between clauses is lower than that of the anti-symmetric relations.

Code and examples are available at code.google.com/p/relevance-based-on-parse-trees/ (package *opennlp.tools.parse_thicket.external_rst*).

6   Discussion and Conclusion

Overall, our evaluation settings are focused on compound queries where most answers correctly belong to the topic of interest in a query and there is usually sufficient number of keywords to assure this. However, in the selected search domain irrelevant answers are those based on foreign entities or mismatched attributes of these entities. Hence augmenting keyword statistics with the structured information of parse trees is not critical to search accuracy improvement. At the same time, discourse information for candidate answers is essential to properly form and interpret the constraints expressed in queries.

Although there has been a substantial advancement in document-level RST parsing, including the rich linguistic features-based of (Feng and Hirst, 2012) and powerful parsing models (Joty et al., 2013), document level discourse analysis has not found a broad range of applications such as search. The most valuable information from DT includes global discourse features and long range structural dependencies between DT constituents.

Despite other studies (Surdeanu et al., 2014) showed that discourse information is beneficial for search via learning, we believe this is the first study demonstrating how Answer Map affects search directly. To be a valid answer for a question, its keywords need to occur in adjacent EDU chain of this answer so that these EDUs are fully ordered and connected by nucleus – satellite relations. Note the difference between the proximity in text as a sequence of words and proximity in DT (Croft et al., 2009). An answer is expected to be invalid if the questions' keywords occur in the answer's satellite EDUs and not in their nucleus EDUs. The purpose of the rhetoric map of an answer is to prevent it from being fired by questions whose keywords occur in non-adjacent areas of this map.

References

S. Joty and A. Moschitti. 2014. Discriminative Reranking of Discourse Parses Using Tree Kernels. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 2049–2060, October 25-29, 2014, Doha, Qatar.

V. Wei Feng and G. Hirst. 2012. Text-level discourse parsing with rich linguistic features. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL-2012), pages 60-68, Jeju, Korea.

P. Jansen, M. Surdeanu, and P. Clark. 2014. Discourse Complements Lexical Semantics for Non-factoid Answer Reranking. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL).

S. Joty, G. Carenini, and R. T. Ng. 2012. A Novel Discriminative Framework for Sentence-Level Discourse Analysis. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL'12, pages 904–915, Jeju Island, Korea. Association for Computational Linguistics.

W. Mann, S. Thompson. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. Text, 8(3):243–281.

S. Joty, G. Carenini, R. Ng, Y. Mehdad. 2013. Combining Intra- and Multi-sentential Rhetorical Parsing for Document-level Discourse Analysis. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria.

B. Galitsky, D. Ilvovsky, S.O. Kuznetsov, F. Strok. 2013. Matching sets of parse trees for answering multi-sentence questions. In Proceedings of the Recent Advances in Natural Language Processing (RANLP), Shoumen, Bulgaria, pages 285–294.

D. Ilvovsky. 2014. Going beyond sentences when applying tree kernels. Proceedings of the Student Research Workshop ACL 2014, pp. 56-63.

B. Galitsky, D. Usikov, S.O. Kuznetsov. 2013. Parse Thicket Representations for Answering Multi-sentence questions. 20th International Conference on Conceptual Structures, ICCS 2013.

B. Galitsky, S.O. Kuznetsov. 2008. Learning communicative actions of conflicting human agents. J. Exp. Theor. Artif. Intell. *20(4): 277-317.*

B. Galitsky. 2012. Machine Learning of Syntactic Parse Trees for Search and Classification of Text. Engineering Application of AI.

A. Moschitti. 2006. Efficient Convolution Kernels for Dependency and Constituent Syntactic Trees. In Proceedings of the 17th European Conference on Machine Learning, Berlin, Germany.

A. Severyn, A. Moschitti. 2012. Structural relationships for large-scale learning of answer re-ranking. SIGIR 2012: 741-750.

A. Severyn, A. Moschitti. 2012. Fast Support Vector Machines for Convolution Tree Kernels. Data Mining Knowledge Discovery 25: 325-357.

M. Collins and N. Duffy. 2002. Convolution kernels for natural language. In Proceedings of NIPS, 625–632.

H. Lee, A. Chang, Y. Peirsman, N. Chambers, Mihai Surdeanu and Dan Jurafsky. 2013. *Deterministic coreference resolution based on entity-centric, precision-ranked rules*. Computational Linguistics 39(4).

B. Croft, D. Metzler, T. Strohman. 2009. Search Engines - Information Retrieval in Practice. Pearson Education. North America.

V. Vapnik. 1995. The Nature of Statistical Learning Theory. – Springer-Verlag.