Predicting Lexical Answer Types in Open Domain QA

Alfio Massimiliano Gliozzo, IBM T.J. Watson Research Center, USA
Aditya Kalyanpur, IBM T.J. Watson Research Center, USA

ABSTRACT
Automatic open-domain Question Answering has been a long standing research challenge in the AI community. IBM Research undertook this challenge with the design of the DeepQA architecture and the implementation of Watson. This paper addresses a specific subtask of Deep QA, consisting of predicting the Lexical Answer Type (LAT) of a question. Our approach is completely unsupervised and is based on PRISMATIC, a large-scale lexical knowledge base automatically extracted from a Web corpus. Experiments on the Jeopardy! data shows that it is possible to correctly predict the LAT in a substantial number of questions. This approach can be used for general purpose knowledge acquisition tasks such as frame induction from text.

Keywords: Association for Computing Machinery (ACM) Proceedings, Information Systems Applications, Miscellaneous, Software Engineering Metrics Theory, Text Tagging

1. INTRODUCTION
Open domain Question Answering (QA) is a long standing research problem that has been pursued for decades (Green, Wolf, Chomsky, & Laughery, 1961; Woods, 1977; Voorhees, 2002; Katz et al., 2002). Recently, IBM took on this challenge in the context of the Jeopardy! game. Jeopardy! is a well-known TV quiz show that has been airing on television in the United States for more than 25 years. It pits three human contestants against one another in a competition that requires answering rich natural language questions over a broad domain of topics, with penalties for wrong answers.

Jeopardy! clues are straightforward assertional forms of questions. So where a question might read, What drug has been shown to relieve the symptoms of ADD with relatively few side effects? the corresponding Jeopardy clue might read This drug has been shown to relieve the symptoms of ADD with relatively few side effects. The correct Jeopardy! response would be What is Ritalin?. The bulk of Jeopardy! clues represent what we would consider factoid questions (i.e. questions whose answers are based on factual information about one or more individual entities). Some more complex clues contain multiple facts about the answer, all of which are required to arrive at the correct response (e.g., the question above contains This...
drug has been shown to relieve the symptoms of ADD and This drug has relatively few side effects).

The development of a system able to compete against grand champions in the Jeopardy! challenge led to the design of the DeepQA architecture and the implementation of Watson (Ferrucci et al., 2010). The Deep QA architecture advances and incorporates a variety of QA technologies including parsing, question classification, question decomposition, automatic source acquisition and evaluation, entity and relation detection, logical form generation, knowledge representation and reasoning. Among those components, type coercion plays a crucial role in filtering out irrelevant candidate answers and providing useful features to assess the confidence level of answers based on their degree of match to the expected answer type specified in the question. Type coercion can be done once the Lexical Answer Type (LAT) of the answer has been identified, for example by looking at knowledge bases like Yago and WordNet, or semi-structured resources such as Wikipedia Categories (see Murdock et al., 2011).

However, for a relatively large set (approx 19%) of Jeopardy! clues the LAT is not explicitly given. In those cases, type coercion cannot be applied unless we figure out a way to infer possible LATs from the clues themselves (for example from the possible types of predicate arguments). This process can be illustrated by the example above assuming that the term This drug in the clue is replaced by this. The two target facts become This has been shown to relieve the symptoms of ADD and This has relatively few side effects. Any English speaker would be able to infer that possible types for X are drugs, medications, treatments, by looking at the two predicates relieve(X,symptoms) and has_few_side_effects(X) and finding the intersection of possible types for X in both cases. However, this problem is particularly hard for a machine, since it involves many ontological and cognitive issues such as:

1. Finding the right level of abstraction (e.g., Thing is not a very useful LAT);
2. The possible set of types is almost unlimited (see section 2);
3. Predicates are usually very ambiguous in the language and often allow multiple types.

This paper addresses this specific subtask of Deep QA, consisting of predicting the LAT of questions. To this aim, we present an unsupervised approach based on PRISMATIC (Fan, Kalyanpur, Gondek, & Ferrucci, 2011) a large scale lexical knowledge base automatically extracted from the web.

2. LEXICAL ANSWER TYPES IN THE DEEP QA ARCHITECTURE

We define a LAT to be a word in the clue that indicates the type of the answer, independent of assigning semantics to that word. For example in the clue Invented in the 1500s to speed up the game, this maneuver involves two pieces of the same color the LAT is the string maneuver (Figure 1).

Determining whether or not a candidate answer can be considered an instance of the LAT is an important kind of scoring and a common source of critical errors. This capability is implemented in the DeepQA architecture by a component called tycor, whose goal is to estimate the extent to which an entity can be coerced into (or matched with) a specific type. Functionally, tycor takes as input any two English terms, the first representing the LAT, the second being the candidate, and returns a numeric value corresponding to the candidate being an instance of the LAT. Therefore its output can be used to generate features exploited at different levels of the Deep QA architecture. As a design issue, to allow generality in the DeepQA architecture, LATs do not belong to a fixed type system. Any term (word or phrase in the language) is then a possible type.
13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:
www.igi-global.com/article/predicting-lexical-answer-types-open/74340?camid=4v1

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Computer Science, Security, and Information Technology. Recommend this product to your librarian:
www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

Elementary: Large-Scale Knowledge-Base Construction via Machine Learning and Statistical Inference
Feng Niu, Ce Zhang, Christopher Ré and Jude Shavlik (2012). International Journal on Semantic Web and Information Systems (pp. 42-73).
www.igi-global.com/article/elementary-large-scale-knowledge-base/74339?camid=4v1a

Semantic Policies for Modeling Regulatory Process Compliance
www.igi-global.com/chapter/semantic-policies-modeling-regulatory-process/60068?camid=4v1a

The Reflexive Practitioner: Knowledge Discovery through Action Research
www.igi-global.com/chapter/reflexive-practitioner-knowledge-discovery-through/65691?camid=4v1a
From Overview to Facets and Pivoting for Interactive Exploration of Semantic Web Data
www.igi-global.com/article/overview-facets-pivoting-interactive-exploration/77822?camid=4v1a