Analysis and Modeling of Effective Passage Retrieval Mechanisms in QAS

Hans Ulrich Christensen and Daniel Ortiz-Arroyo
Computer Science Department
Aalborg University Esbjerg
Niels Bohrs Vej 8, 6700 Denmark
E-mail: huc1405@student.aau.dk, do@cs.aau.dk

Abstract—Effective mechanisms for passage retrieval have a direct impact in the performance of Question Answering Systems (QAS). In this paper we investigate the application of fuzzy logic models in Passage Retrieval (PR), and describe FuzzyPR, our passage retrieval system that applies these models. Additionally, a performance analysis of diverse mechanisms employed in passage retrieval for open domain QAS is presented. Preliminary performance results of FuzzyPR, when applied to retrieving answering passages in the TREC corpora, are provided.

Keywords: Information Retrieval, Question Answering Systems, Passage Retrieval, Fuzzy Logic

I. INTRODUCTION AND RELATED WORK

A Question Answering System (QAS) is one type of information retrieval (IR) system that attempts to find exact answers to user’s questions expressed in natural language. In an Open-Domain Question Answering System (ODQAS), questions are not restricted to certain topics and answers have to be found in an unstructured document collection. Passage Retrieval (PR), one component of a QAS, extracts text segments from a group of retrieved documents and ranks these passages in decreasing order of computed likelihood for containing the correct answer. Typically, such text segments are referred to as candidate passages.

Applying fuzzy logic in IR systems —especially similarity measures— is a promising approach since this mathematical framework models naturally the uncertainty and vagueness involved in the process of retrieving information. Examples of successful applications of fuzzy logic in IR include Larsen’s Query Answering System [1] for libraries and Szczepaniak and Gil’s experimental IR system for retrieving documents written in Polish [2]. Kong et al. [4] explored the use of fuzzy aggregation operators in a passage-based retrieval system for documents, where the relevance of a document is re-calculated taking into account the retrieved passages. Brøndsted et al. describe in [3] a fuzzy logic based implementation of a document retrieval system that employs concept clusters and statistical query term expansion in a closed domain spoken QA system for mobile devices.

This paper investigates the application of fuzzy logic based models for PR in ODQAS. A performance analysis of some proposed mechanisms for PR system is presented jointly with the preliminary performance results of FuzzyPR, our PR system. The paper is organized as follows. Section II describes and analyzes the main component mechanisms of a passage retrieval system. Section III briefly describes FuzzyPR and presents its preliminary performance results. Finally, Section IV presents some conclusions and future work.

II. ANALYSIS OF MAIN COMPONENT MECHANISMS IN A PASSAGE RETRIEVAL SYSTEM

The intuition “frequently, an answer to a (factoid) question can be found as a reformulation of the same question” has been applied successfully in passage retrieval systems for QA [5], [6]. An example of the application of this reformulation intuition approach in PR is the question “How much is the international space station expected to cost?” of QA@TREC 11 (QID: 1645) 1. The answering passage contains the snippet “(...)United States and Russia, are working together to build the SPACE STATION, which is EXPECTED TO COST between $40 billion and $60 billion(...)”. Successful applications of this intuition include Gómes-Soriano et al.’s [5] n-gram based passage retrieval system for QAS, where passages containing larger sequences of terms of the questions are ranked higher and Brill et al.’s [6] Web QAS, which poses queries constructed as permutations of the terms employed in the question. In the following subsections we briefly analyze three of the main modeling components that are used in our fuzzy logic modeling implementation of the reformulation intuition. Further details can be found in [7].

A. Automatic Detection of Term Variations

A QAS requires an automatic mean to detect term variations occurring in documents and questions written in natural language. Term variations are lexical differences, in terms of meaning and spelling, between a word of the question typed by a user and an equivalent word contained in a document in the corpus. Reasons for term variations include grammatical inflection and spelling mistakes. These vocabulary mismatches have a negative impact on the effectiveness of an IR system when it is not able to recognize them. Two main features are needed in a mechanism to handle term variations effectively: 1) language-independence and 2) effectiveness, measured in terms of tolerance toward common misspellings and grammatical inflections (interpreted as a type of misspelling). Contrary

1TREC’s Question Answering collections are available from: http://trec.nist.gov/data/qa.html
to Boolean algorithms for term matching, fuzzy term similarity algorithms determine the degree of similarity between two strings. Reflexivity and symmetry are desired properties of these algorithms. To select the most adequate fuzzy term similarity algorithm for FuzzyPR, we performed a comparative evaluation on the effectiveness of six different algorithms when set to calculate the similarity between 300 English homophone pairs. The average of the similarity computations yields the score of the fuzzy term matching algorithm.

![Correct answers vs. consulted passages](image1)

**Fig. 1:** Comparative evaluation of 6 fuzzy terms similarity algorithms

As illustrated in Fig. 1, the inverse normalized Damerau-Levenshtein (InDD) [7] distance performed best, giving an average homophone pair similarity rate of 0.5552.

![Coverage of the 3 fuzzy proximity measures](image2)

**Fig. 2:** The Coverage of the 3 fuzzy proximity measures

**Table I:** MRR of 3 fuzzy proximity measures

<table>
<thead>
<tr>
<th>Fuzzy proximity measure</th>
<th>Mean Reciprocal Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Span Size Ratio</td>
<td>0.2933</td>
</tr>
<tr>
<td>Fuzzy Proximity Measure w. k = 70</td>
<td>0.3463</td>
</tr>
<tr>
<td>Extended Distance Factor w. s = 0.1 and α = 0.75</td>
<td>0.3137</td>
</tr>
</tbody>
</table>

A homophone pair is two terms pronounced the same but differing in meaning and spelling, thus reflecting misspellings and typos. Examples include “advice vs. advise” and “cite vs. site”.

**B. Proximity of Question Terms Occurring in the Passage**

The proximity of question terms occurring within a passage indicates the probable location of an answer in a document. Fuzzy proximity measures calculate the degree of proximity within a document of two or more question terms, based on the following two intuitions:

- If all matching document terms are juxtaposed then the measure yields 1, and;
- The farther away the matching document terms occur, the lower the degree of proximity.

We evaluated three different fuzzy proximity measures as to their ability in finding answering passages for the first 50 questions of TREC 11’s question set using the AQUAINT corpus.

![Image showing MRR comparison](image3)

**TABLE I:** MRR of 3 fuzzy proximity measures

- **Multiplicative importance weighted Averaging Operators**, where importance weights are sum-normalized and membership degrees multiplied by the importance weights, and;
- **Implicative importance weighted Averaging Operators**, where importance weights are max-normalized and applied using fuzzy implication: $v \Rightarrow a$—i.e., a high weight implies a high degree of satisfaction.

We chose the class of implicative importance weighted Averaging Operators since this type of operators have the desirable property that importance weights have maximum influence on the result at both high and low ANDness.
Among the operators available, two were used in our experiments: Andness Directed Implicative-Importance Weighting (AIWA) and Importance-Weighted Maximum-Entropy Ordered Weighted Averaging (IW-MEOWA) [11]. Both operators satisfy the requirements of efficiency and effectiveness and have complementary strengths and weaknesses as shown in Table II.

III. PASSAGE RETRIEVAL SYSTEM AND PRELIMINARY PERFORMANCE RESULTS

FuzzyPR employs a document retrieval system based on Apache Lucene, a popular open source vector space search engine, to efficiently retrieve documents in response to a question. Additionally, FuzzyPR includes a special PR mechanism that identifies and processes the passages included within the retrieved documents. The PR mechanism consists of two main components: 1) a question–passage similarity measure module and 2) a passage identification and extraction mechanism adapted to the special needs of QAS. The following subsections describe these components.

A. Similarity Measure

The similarity measure used in FuzzyPR is based on the fuzzy logic interpretation of the intuition: "a passage is relevant to the question posed if many question terms or variations of these question terms occur in close proximity". Equation 1 describes the measure that models such intuition.

\[
\mu_{rel}(p, q) = w \text{Min}((v_1, \mu_f(p, q), v_2, \mu_p(p, q)))
\]  

where \( w \) is the term influence function and \( \mu_f(p, q) \) and \( \mu_p(p, q) \) are the weighted fraction of question terms occurring in the passage and the question respectively. The term influence function \( \mu_f(p, q) \) is determined experimentally. The value of \( w \) is determined based on the distance of the question terms from the question and the values of \( v_1 \) and \( v_2 \) are determined experimentally. The similarity measure combines lexical and statistical data extracted at term-level into the two fuzzy measures: \( \mu_f(p, q) \) and \( \mu_p(p, q) \) into a single fuzzy value: \( \mu_{rel}(p, q) \).

B. Mechanism for Passage Identification and Extraction

FuzzyPR employs a fuzzified variation of the concept arbitrary passages\(^3\). Arbitrary passages are modeled by its membership function in the ideal set of passage size as stated in equation 4.

\[
\mu_{ideal\ size} = \begin{cases} 
1 & \text{if } 0 \leq x \leq d \\
\frac{x-b}{d-b} & \text{if } d < x < b \\
0 & \text{if } x \geq b 
\end{cases}
\]  

where \( d \) and \( b \) are used to adjust the crisp support and the fuzzy support values respectively. Due to efficiency concerns, the membership function of the ideal passage size set is transformed into an equivalent symmetric membership function, where the center term of a passage is required to have a question term similarity greater than \( \alpha \) and a NIDF greater than \( \beta \). The justification for this restriction is the intuition that a passage that contains none or very few of the terms in the question is unlikely to provide an answer to the question.

\(^3\)Arbitrary passages are defined as: "any sequence of words of any length starting at any word in the document".
C. Preliminary Performance Results

We measured the effectiveness of FuzzyPR, by comparing its ability to find correct answers to questions in a document corpora against an adapted PR system that we have integrated within Lucene. This adapted PR system implements an index of 3 sentence passages with 1 sentence overlapping. Llopis et al. in [12] report that this approach achieves good results. The PR system allows Lucene to be used as QAS by employing a simple query expansion method. In this method the question term with the lowest IDF is removed until \( \geq 20 \) passages are retrieved from the index of 3 sentence passages. As test data we used TREC-12’s set of 495 questions and the corpus called AQUAINT, which consists of approximately 1 million documents of English news text. To check automatically for correct answers to questions, using Ken Ltitkowsky’s regular expression patterns, the question set was reduced to 380, since 115 questions do not have a pattern. As evaluation metrics we used Coverage, and the Mean Reciprocal Rank (MRR). As is done in the JIRS system [5], we measure coverage on the first top 20 passages.

Fig. 3 and table III show that FuzzyPR consistently outperforms the vector space PR system in terms of coverage (by a margin of 5-10%) and MRR, independently of the number of top-ranked passages consulted.

![Fig. 3: The coverage of Lucene PR and FuzzyPR tested with TREC 12's QA test data](image)

### TABLE III: MRR obtained with TREC-12’s QA test data

<table>
<thead>
<tr>
<th>Passage retrieval system</th>
<th>MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lucene PR</td>
<td>0.2855</td>
</tr>
<tr>
<td>FuzzyPR</td>
<td>0.3099</td>
</tr>
</tbody>
</table>

IV. Conclusions and Future Work

In this paper we presented FuzzyPR, a novel PR system that implements fuzzy logic models for passage retrieval. The main component mechanisms included in FuzzyPR are: 1) automatic detection of term variations, 2) proximity of question terms, and 3) fraction of question terms occurring in the passage. Using these components we created a fuzzy logic model based interpretation of the reformulation intuition. FuzzyPR has three main advantages: 1) its passage identification and extraction methods enables it to retrieve candidate passages from documents at retrieval retrieval time thus avoiding the time-consuming indexing process\(^5\) 2) its language independence property, and 3) its capability for handling term variations due to spelling errors and grammatical reflections.

Our preliminary evaluation shows that FuzzyPR achieves a consistently higher coverage and MRR than a PR system adapted within Lucene. To test its performance on another language, FuzzyPR will be further evaluated with the Spanish CLEF corpus. We also plan to perform a comparative evaluation of FuzzyPR against JIRS [5]. Finally, we will explore the effect in performance of combining different fuzzy logic based PR mechanisms with machine learning techniques.

### REFERENCES


\(^5\)An unoptimized method in Java for segmenting and indexing the AQUAINT corpus took 4 hours on an AMD64 3400+ w. 2 GB RAM and RAID 0.