Question Classification in English-Chinese Cross-Language Question Answering: An Integrated Genetic Algorithm and Machine Learning Approach

Min-Yuh Day 1, 2, Chorng-Shyong Ong 2, and Wen-Lian Hsu 1,* , Fellow, IEEE

1 Institute of Information Science, Academia Sinica, Taiwan
2 Department of Information Management, National Taiwan University, Taiwan

{myday, hsu}@iis.sinica.edu.tw; ongcs@im.ntu.edu.tw

Introduction

- Question classification (QC) plays an important role in cross-language question answering (CLQA)
  - QC: Accurately classify a question into a question type and then map it to an expected answer type
  - “What is the biggest city in the United States?”
  - Question Type: “Q_LOCATION_CITY”
  - Extract and filter answers in order to improve the overall accuracy of a cross-language question answering system
Introduction (cont.)

Question informers (QI) play a key role in enhancing question classification for factual question answering

- **QI**: Choosing a minimal, appropriate contiguous span of a question token, or tokens, as the informer span of a question, which is adequate for question classification.

- “What is the biggest city in the United States?”
- Question informer: “city”
- “city” is the most important clue in the question for question classification.
Introduction (cont.)

- **Feature Selection** in Machine Learning
  - Optimization problem that involves choosing an appropriate feature subset.
  - Hybrid approach that integrates Genetic Algorithm (GA) and Conditional Random Fields (CRF) improves the accuracy of question informer prediction in traditional CRF models (Day et al., 2006)

- We propose an integrated Genetic Algorithm (GA) and Machine Learning (ML) approach for question classification in cross-language question answering.
Research Background

- Cross Language Question Answering
  - International Question Answering (QA) contests
    - TREC QA: 1999~
      - Monolingual QA in English
    - QA@CLEF: 2003~
      - European languages in both non-English monolingual and cross-language
    - NTCIR CLQA: 2005~
      - Asian languages in both monolingual and cross-language
  
- Question Classification
  - Rule-based method
  - Machine Learning based method
Two strategies for question classification in English-Chinese cross-language question answering

1) **Chinese QC (CQC)** for both English and Chinese queries.
   - English source language has to be translated into the Chinese target language in advance.

2) **English QC (EQC)** for English queries and Chinese QC (CQC) for Chinese queries.

We focus on question classification in English-Chinese cross-language question answering

- Bilingual QA system for English source language queries and Chinese target document collections.
Methods

- Hybrid GA-CRF-SVM Architecture
  - GA for CRF Feature Selection
  - GA-CRF Question Informer Prediction
  - SVM-based Question Classification using GA-CRF Question Informer
Hybrid GA-CRF-SVM Architecture

- **GA**
  - GA : Feature Selection
    - GA for CRF Feature Selection
    - Near Optimal Feature Subset of CRF

- **CRF**
  - GA-CRF Question Informer Prediction
    - Near Optimal CRF Prediction Model
    - CRF-based Question Informer Prediction

- **SVM**
  - SVM-based Question Classification
    - Question Informer
    - SVM-based Question Classification

- Question Type
Encoding a Feature Subset of CRF with the structure of chromosomes

Initialization

Population

Evaluate (Fitness Function)

Stopping criteria Satisfied?

Yes

Near Optimal Feature Subset of CRF

Near Optimal CRF Prediction Model

No

GA Operators: Reproduction, Crossover, Mutation

CRF model 10-fold Cross Validation

x: Feature subset

F(x): Fitness Function

Training dataset

Test dataset

CRF-based Question Informer Prediction

GA-CRF Question Informer Prediction
Experiment Design

- **Data set for English Question Classification**
  - Training dataset (5288E)
    - 4,204 questions from UIUC QC dataset (E)
    - + 500 questions from the NTCIR-5 CLQA development set (E)
    - + 200 questions from the NTCIR-5 CLQA test set (E)
    - + 384 questions from TREC2002 questions (E)
  - Test dataset (CLQA2T150E)
    - 150 English questions from NTCIR-6 CLQA’s formal run

- **Data set for Chinese Question Classification**
  - Training dataset (2322C)
    - 1238 question from IASL (C)
    - + 500 questions from the NTCIR-5 CLQA development set (C)
    - + 200 questions from the NTCIR-5 CLQA test set (C)
    - + 384 questions from TREC2002 questions (translated) (C)
  - Test dataset (CLQA2T150C)
    - 150 Chinese questions from NTCIR-6 CLQA’s formal run
Experiment Design (cont.)

- Features for English Question Classification
  - Syntactic features
    - Word-based bi-grams of the question (WB)
    - First word of the question (F1)
    - First two words of the question (F2)
    - Wh-word of the question (WH)
      - i.e., 6W1H1O: who, what, when, where, which, why, how, and other
  - Semantic features
    - Question informers predicted by the GA-CRF model (QIF)
    - Question informer bi-grams predicted by the GA-CRF model (QIFB)
Experiment Design (cont.)

- Features for Chinese Question Classification
  - Syntactic features
    - Bag-of-Words
      - character-based bi-grams (CB)
      - word-based bi-grams (WB).
    - Part-of-Speech (POS)
  - Semantic Features
    - HowNet Senses
      - HowNet Main Definition (HNMD)
      - HowNet Definition (HND).
    - TongYiCi CiLin (TYC)
Experiment Design (cont.)

- Performance Metrics

  - Accuracy

  \[
  \text{Accuracy} = \frac{\text{Number of corrected question types}}{\text{Total number of questions}}
  \]

- MRR (mean reciprocal rank)

  \[
  \text{MRR} = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{\text{rank}_i}
  \]

  where \(\text{rank}_i\) is the rank of the first corrected question type of the \(i^{th}\) question, and \(M\) is total number of questions.
Experimental Results

- Question informer prediction
  - Using GA to optimize the selection of the feature subset in CRF-based question informer prediction improves the F-score from 88.9% to 93.87%, and reduces the number of features from 105 to 40.
    - Training dataset (UIUC Q5500)
    - Test dataset (UIUC Q500)
  - The accuracy of our proposed GA-CRF model for the UIUC dataset is 95.58% compared to 87% for the traditional CRF model reported by Krishnan et al. (2005)
    - The proposed hybrid GA-CRF model for question informer prediction significantly outperforms the traditional CRF model.
**Experimental Results**

- **English Question Classification (EQC) using SVM**

![Graph showing accuracy rates for different combinations of features.](image)

- **Accuracy**
  - **Top 1 Accuracy (Fine)**
  - **Top 1 Accuracy (Coarse)**

- **Results**
  - **WB**
    - 86.00% 88.67% 90.67% 92.00% 94.00% 95.33%
  - **WB+WH**
    - 86.67% 89.33% 91.33% 93.00% 94.00% 95.33%
  - **WB+WH+QIF**
    - 89.33% 91.33% 93.00% 94.00% 95.33% 95.33%
  - **WB+WH+QIF+QIF+QF**
    - 92.00% 94.00% 95.33% 95.33% 95.33% 95.33%
  - **FB+F1+F2**
    - 94.00% 95.33% 95.33% 95.33% 95.33% 95.33%
## Experimental Results of Chinese Question Classification (CQC) using SVM with different features

<table>
<thead>
<tr>
<th>Feature Used</th>
<th>Top 1 Accuracy (Fine)</th>
<th>Top 1 Accuracy (Coarse)</th>
<th>Top 5 MRR (Fine)</th>
<th>Top 5 MRR (Coarse)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POS</td>
<td>53.33%</td>
<td>65.33%</td>
<td>0.5732</td>
<td>0.7533</td>
</tr>
<tr>
<td>POSB</td>
<td>60.00%</td>
<td>74.00%</td>
<td>0.6469</td>
<td>0.7970</td>
</tr>
<tr>
<td>HNMD</td>
<td>71.33%</td>
<td>81.33%</td>
<td>0.7480</td>
<td>0.8832</td>
</tr>
<tr>
<td>CB</td>
<td>74.00%</td>
<td>84.67%</td>
<td>0.7934</td>
<td>0.9130</td>
</tr>
<tr>
<td>HNMDB</td>
<td>74.00%</td>
<td>86.00%</td>
<td>0.7916</td>
<td>0.9117</td>
</tr>
<tr>
<td>C</td>
<td>74.67%</td>
<td>84.67%</td>
<td>0.7979</td>
<td>0.9152</td>
</tr>
<tr>
<td>TYCB</td>
<td>74.67%</td>
<td>86.00%</td>
<td>0.7880</td>
<td>0.9062</td>
</tr>
<tr>
<td>HND</td>
<td>74.67%</td>
<td>86.67%</td>
<td>0.7860</td>
<td>0.9102</td>
</tr>
<tr>
<td>W</td>
<td>76.00%</td>
<td>88.00%</td>
<td>0.7901</td>
<td>0.9208</td>
</tr>
<tr>
<td>HNDB</td>
<td>76.67%</td>
<td>88.00%</td>
<td>0.8000</td>
<td>0.9162</td>
</tr>
<tr>
<td>WB</td>
<td>77.33%</td>
<td>88.00%</td>
<td>0.8067</td>
<td>0.9162</td>
</tr>
<tr>
<td>TYC</td>
<td>77.33%</td>
<td>88.67%</td>
<td>0.8019</td>
<td>0.9240</td>
</tr>
</tbody>
</table>
Experimental Results (cont.)

- Chinese Question Classification (CQC) using SVM

![Graph showing Chinese Question Classification accuracy for different methods: CB, CB+HNMD, CB+HNMD+HND, CB+HNMD+HND+TYC.

- Top 1 Accuracy (Fine)
- Top 1 Accuracy (Coarse)
Conclusions

- We have proposed a hybrid genetic algorithm and machine learning approach for cross-language question classification.

- The major contribution of this paper is that the proposed approach enhances cross-language question classification by using the GA-CRF question informer feature with Support Vector Machines (SVM).

- The results of experiments on NTCIR-6 CLQA question sets demonstrate the efficacy of the approach in improving the accuracy of question classification in English-Chinese cross-language question answering.
Applications: ASQA (Academia Sinica Question Answering System)

- ASQA (IASL-IIS-SINICA-TAIWAN)
  - ASQA is the best performing Chinese question answering system.
  - The first place in the English-Chinese (E-C) subtask of the NTCIR-6 Cross-Lingual Question Answering (CLQA) task. (2007)
  - The first place in the Chinese-Chinese (C-C) subtask of the NTCIR-6 Cross-Lingual Question Answering (CLQA) task. (2007)
  - The first place in the Chinese-Chinese (C-C) subtask of the NTCIR-5 Cross-Lingual Question Answering (CLQA) task. (2005)

http://asqa.iis.sinica.edu.tw
Question Classification in English-Chinese Cross-Language Question Answering: An Integrated Genetic Algorithm and Machine Learning Approach

Min-Yuh Day ¹, ², Chorng-Shyong Ong ², and Wen-Lian Hsu ¹,*

¹ Institute of Information Science, Academia Sinica, Taiwan
² Department of Information Management, National Taiwan University, Taiwan

{myday, hsu}@iis.sinica.edu.tw; ongcs@im.ntu.edu.tw