Information Retrieval and Extraction Project 1

NTU CSIE, Fall 2018 陳信希 教授
顏安孜 陳重吉 {d04922005, d05922016}@ntu.edu.tw

CONTENTS

- Task Description
- Data
- Evaluation
- Report

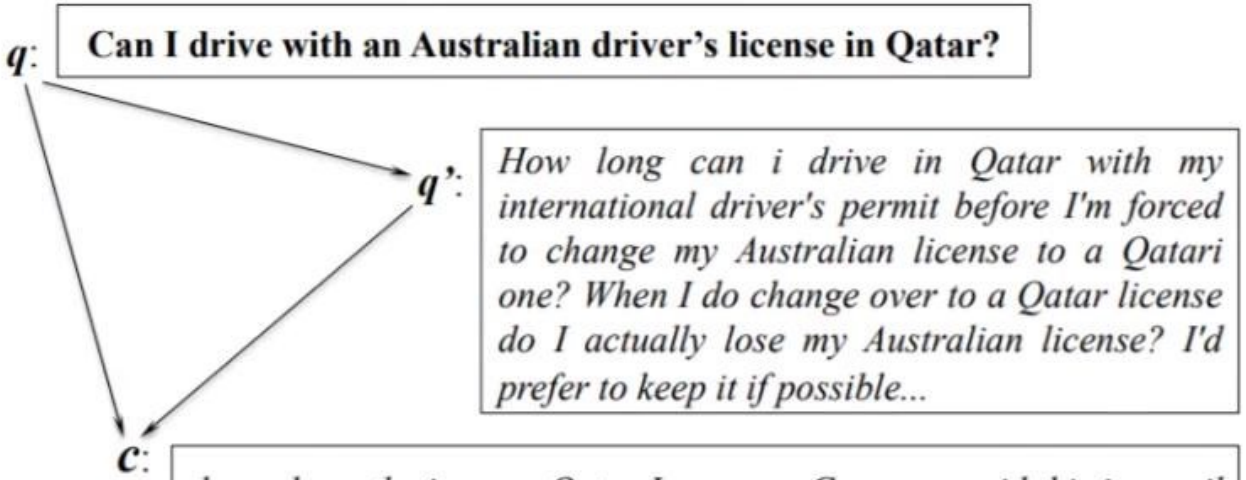
1

Task Description

SemEval-2017 Task 3
Community Question Answering

http://alt.qcri.org/semeval2017/task3/

http://aclweb.org/anthology/S17-2003


q:
Can I drive with an Australian driver's license in Qatar?
q':
How long can i drive in Qatar with my international driver's permit before I'm forced to change my Australian license to a Qatari one? When I do change over to a Qatar license do I actually lose my Australian license? I'd prefer to keep it if possible...
c:
depends on the insurer, Qatar Insurance Company said this in email to me:"Thank you for your email! With regards to your query below, a foreigner is valid to drive in Doha with the following conditions: Foreign driver with his country valid driving license allowed driving only for one week from entry date Foreign driver with international valid driving license allowed driving for 6 months from entry date Foreign driver with GCC driving license allowed driving for 3 months from entry". As an Aussie your driving licence should be transferable to a Qatar one with only the eyetest (temporary, then permanent once RP sorted).

Task Description – Subtask A

English subtask A

- Question-Comment Similarity
- Given a question and the first 10 comments in its question thread, rerank these 10 comments according to their relevance with respect to the question.

Task Description – Subtask B

English subtask B

- Question-Question Similarity
- Given a new question (aka original question) and the set of the first 10 related questions (retrieved by a search engine), rerank the related questions according to their similarity with the original question.

Task Description – Subtask C

English subtask C

- Question-External Comment Similarity
- Given a new question (aka the original question) and the set of the first 10 related questions (retrieved by a search engine), each associated with its first 10 comments appearing in its thread, rerank the 100 comments (10 questions x 10 comments) according to their relevance with respect to the original question.

2

Data

Data – Distribution

Train

- SemEval2015-Task3-CQA-QL-dev-reformatted-...
- SemEval2015-Task3-CQA-QL-test-reformatted-...
- SemEval2015-Task3-CQA-QL-train-reformatted...
- SemEval2016-Task3-CQA-QL-dev.xml
- SemEval2016-Task3-CQA-QL-test.xml
- SemEval2016-Task3-CQA-QL-train-part1.xml
- SemEval2016-Task3-CQA-QL-train-part2.xml

Test

- SemEval2017-task3-English-test-input.xml

Category	Train+Dev+Test from SemEval-2015	Train(1,2)+Dev+Test from SemEval-2016	Test
Original Questions	–	(200+67)+50+70	88
Related Questions	2,480+291+319	(1,999+670)+500+700	880
– Perfect Match	–	(181+54)+59+81	24
– Relevant	–	(606+242)+155+152	139
– Irrelevant	–	(1,212+374)+286+467	717
Related Comments (with respect to Original Question)	–	(19,990+6,700)+5,000+7,000	8,800
– Good	–	(1,988+849)+345+654	246
– Bad	–	(16,319+5,154)+4,061+5,943	8,291
– Potentially Useful	–	(1,683+697)+594+403	263
Related Comments (with respect to Related Question)	14,893+1,529+1,876	(14,110+3,790)+2,440+3,270	2,930
– Good	7,418+813+946	(5,287+1,364)+818+1,329	1,523
– Bad	5,971+544+774	(6,362+1,777)+1,209+1,485	1,407
– Potentially Useful	1,504+172+156	(2,461+649)+413+456	0

Data Format – Original Questions

A dataset file is a sequence of examples (original questions):

<root>
  <OrgQuestion> ... </OrgQuestion>
  <OrgQuestion> ... </OrgQuestion>
  ...
  <OrgQuestion> ... </OrgQuestion>
</root>

Each OrgQuestion has an ID, e.g., <OrgQuestion ORGQ_ID="Q1“>

Data Format – The structure of an OrgQuestion

The structure of an OrgQuestion is the following:

```
<OrgQuestion ...>
  <OrgQSubject> text </OrgQSubject>
  <OrgQBody> text </OrgQBody>
  <Thread ...>
    <RelQuestion ...> ... </RelQuestion>
    <RelComment ...> ... </RelComment>
    ...
    <RelComment ...> ... </RelComment>
</OrgQuestion>
```

- <OrgQSubject> text </OrgQSubject> → The subject of the original question
- <OrgQBody> text </OrgQBody> → The main body of the question
- <Thread ...> → A Thread consists of a potentially relevant question RelQuestion, together with 10 comments RelComment for it.

Data Format – RelQuestion

Each RelQuestion tag has a list of attributes, as in the following example:

```
<RelQuestion RELQ_ID="Q1_R4" RELQ_RANKING_ORDER="4"
RELQ_CATEGORY="Advice and Help" RELQ_DATE="2013-05-02 19:43:00"
RELQ_USERID="U1" RELQ_USERNAME="ankukuma"
RELQ_RELEVANCE2ORGQ="PerfectMatch">
```

RELQ_RELEVANCE2ORGQ: relevance of the thread of this RelQuestion with respect to the OrgQuestion. This label could be:

- PerfectMatch: RelQuestion matches OrgQuestion (almost) perfectly (at test time, this is to be merged with Relevant)
- Relevant: RelQuestion covers some aspects of OrgQuestion
- Irrelevant: RelQuestion covers no aspects of OrgQuestion

Data Format – RelComment

Each RelComment tag has a list of attributes, as in the following example:

<RelComment RELC_ID="Q104_R22_C1" RELC_DATE="2012-01-09 11:39:52" RELC_USERID="U2011" RELC_USERNAME="drsam" RELC_RELEVANCE2ORGQ="Bad" RELC_RELEVANCE2RELQ="Good">

RELC_RELEVANCE2ORGQ: human assessment about whether the comment is "Good", "Bad", or "PotentiallyUseful" with respect to the **OrgQuestion**

- Good: at least one subquestion is directly answered by a portion of the comment

- PotentiallyUseful: no subquestion is directly answered, but the comment gives potentially useful information about one or more subquestions (at test time, this class will be merged with "Bad")

- Bad: no subquestion is answered and no useful information is provided

Data Format – RelComment

Each RelComment tag has a list of attributes, as in the following example:

<RelComment RELC_ID="Q104_R22_C1" RELC_DATE="2012-01-09 11:39:52"
RELC_USERID="U2011" RELC_USERNAME="drsam"
RELC_RELEVANCE2ORGQ="Bad" RELC_RELEVANCE2RELQ="Good">

RELC_RELEVANCE2RELQ:

human assessment about whether the comment is "Good", "Bad", or "PotentiallyUseful" (the latter two will be merged under "Bad" at test time) with respect to the **RelQuestion**

3

Evaluation

Evaluation

The scorer takes as input a <GOLD_FILE> and a <PREDICTIONS_FILE>

Both files should contain one prediction per line in the following format:

"<Question_ID> <Answer_ID> <RANK> <SCORE> <LABEL>“

where tabulation is used as a separator.

Evaluation

<Question_ID>

- RELQ_ID in subtask A
- ORGQ_ID in subtasks B and C

<Answer_ID>

- RELC_ID in subtasks A and C
- RELQ_ID in subtask B

<RANK>

A positive integer, reflecting the rank of the answer with respect to the question.

<SCORE>

A real number reflecting the relevance of the answer with respect to the question.

A higher value means higher relevance of the answer with respect to the question.

<LABEL>

There are only two possible values: 'true' and 'false'.

- subtasks A and C: 'true' covers 'Good', 'false' covers 'Bad' and 'PotentiallyUseful‘
- subtask B: 'true' covers 'PerfectMatch' and 'Relevant, 'false' covers 'Irrelevant'

Evaluation

An example <PREDICTIONS_FILE> for subtask C:

Q104 Q104_R1_C1 0 0.341656072212579 false

Q104 Q104_R1_C2 0 1.49980396096993 true

Q104 Q104_R1_C3 0 0.192743311247135 false

…

Evaluation

The official score for the competition for all subtasks is **MAP**.

The scorer calculates and outputs the following scores:

A. Classification scores
- accuracy
- precision (P), recall (R), and F1-score with respect to the positive ('true') class

B. Overall ranking scores
- mean average precision (MAP) -- the official score
- average recall (AvgRec)
- MRR (mean reciprocal rank)

C. Ranking scores at each ranking position r (1 <= r <= 10)
- REC-1: percentage of questions with at least 1 correct answer in the top @X positions
- ACC: accuracy, i.e., number of correct answers retrieved at rank @X normalized by the rank and the total number of questions

Evaluation

Use of the scorer:

python MAP_scripts/ev.py <GOLD_FILE> <PREDICTIONS_FILE>

Example use:

- python MAP_scripts/ev.py SemEval2017-task3-English-test.xml.subtaskA.relevancy subtaskA.pred
- python MAP_scripts/ev.py SemEval2017-task3-English-test.xml.subtaskB.relevancy subtaskB.pred
- python MAP_scripts/ev.py SemEval2017-task3-English-test.xml.subtaskC.relevancy subtaskC.pred

```
(py2.7) azyen@nlg-wks-d:/nfs/nas-5.1/azyen/IRIE2018/scorer_v2.3$ python MAP_scripts/ev.py ../SemEval2017_task3_submissions_and_scores/_gold/SemEval2017-Task3
-CQA-QL-test-subtaskA.xml.subtaskA.relevancy ../SemEval2017_task3_submissions_and_scores/KeLP/subtask_A_primary.txt

*** Official score (MAP for SYS): 0.8843


******************************
*** Classification results ***
******************************

Acc = 0.7389
P   = 0.8730
R   = 0.5824
F1  = 0.6987


********************************
*** Detailed ranking results ***
********************************

IR  -- Score for the output of the IR system (baseline).
SYS -- Score for the output of the tested system.

          IR   SYS
MAP   : 0.7261 0.8843
AvgRec: 0.7932 0.9379
MRR   :  82.37  92.82
             IR    SYS             IR    SYS             IR    SYS           IR  SYS
REC-1@01:  73.38  89.08  ACC@01:  73.38  89.08  AC1@01:   0.75   0.91  AC2@01:  215  261
REC-1@02:  83.62  94.88  ACC@02:  65.53  85.84  AC1@02:   0.69   0.90  AC2@02:  384  503
REC-1@03:  91.13  96.59  ACC@03:  64.28  81.91  AC1@03:   0.70   0.90  AC2@03:  565  720
REC-1@04:  93.52  96.93  ACC@04:  61.77  78.41  AC1@04:   0.71   0.90  AC2@04:  724  919
REC-1@05:  96.25  97.61  ACC@05:  59.52  75.22  AC1@05:   0.73   0.92  AC2@05:  872 1102
REC-1@06:  96.93  97.61  ACC@06:  57.91  70.82  AC1@06:   0.76   0.93  AC2@06: 1018 1245
REC-1@07:  96.93  97.95  ACC@07:  55.97  66.55  AC1@07:   0.80   0.96  AC2@07: 1148 1365
REC-1@08:  96.93  97.95  ACC@08:  54.48  61.56  AC1@08:   0.86   0.97  AC2@08: 1277 1443
REC-1@09:  97.61  97.95  ACC@09:  53.01  56.66  AC1@09:   0.92   0.99  AC2@09: 1398 1494
REC-1@10:  97.95  97.95  ACC@10:  51.98  51.98  AC1@10:   1.00   1.00  AC2@10: 1523 1523

REC-1 - percentage of questions with at least 1 correct answer in the top @X positions (useful for tasks where questions have at most one correct answer)
ACC   - accuracy, i.e., number of correct answers retrieved at rank @X normalized by the rank and the total number of questions
AC1   - the number of correct answers at @X normalized by the number of maximum possible answers (perfect re-ranker)
AC2   - the absolute number of correct answers at @X
```

4

Report

Report

- Language: Chinese or English (Be clear in meaning!)
- Pages: no more than 6 (with readable font size)
- Format: PDF
- Must include:
  - Name and ID
  - Division of work
  - Methods
  - Evaluation
  - Discussion
  - Conclusion

Code

- Describe your code
  - write the proper comment for each part and function

Uncommented Code

```
city=raw_input("Enter a city: ")
while city[-1]==" ":
    city = city[:-1]
temp=raw_input("Enter a temperature in Farenheit: ")
temp = float(temp)
temp = (temp - 32.0)*(100.0/180.0)
temp = round(temp,3)
temp = str(temp)
print "In "+city+" it is "+temp+" degrees Celcius!"
```

Commented Code

```
#Alyssa P. Hacker
#fah_to_celsius.py

#collect a city name from user
city=raw_input("Enter a city: ")

#trunacte whitespace
while city[-1]==" ":
    city = city[:-1]

#collect a temp from user
temp=raw_input("Enter a temperature in Farenheit: ")

#convert string to float
temp = float(temp)

#convert Farenheit temp to Celsius temp
temp = (temp - 32.0)*(100.0/180.0)

#truncate to 3 decimal places
temp = round(temp,3)

#recast as string so we can concatenate
temp = str(temp)

#print result!
print "In "+city+" it is "+temp+" degrees Celcius!"
```

Submission Format

Project1_team_<team number>.zip
- Report_team_<team number>.pdf
- Code_team_<team number>
  - Readme(description of each script)
  - $script_1$
  - …
  - $script_n$

Ex:
Project1_team_0.zip
- Report_team_0.pdf
- Code_team_0
  - Readme.txt
  - functions.py
  - main.py

Project 1 presentation

- Date: 2018/12/06
- Please submit your presentation slides to CEIBA before **2018/12/05 21:59**
- **5 minutes** per group
- Judging Criteria
  - Content
  - State your idea, methodology, evaluation and conclusion clearly and logically

Project 1 Schedule

11/30 23:59

Report & Code Submission Due

12/5 21:59

Slide Submission Due

12/6

Presentation in class

Submit via CIEBA

One submission per group

Delay: -5% per day

Should you have any question, please let TAs know.

Grading Policy

- Report: 60%
- Presentation: 30%
- Performance: 10%
- For each subtask, you will get 3% bonus if your performance beat the best result in paper.☺

Have Fun !

Any questions?

Detail of Dataset

Data Format – Thread (obligatory)

A thread has one obligatory and one optional attribute as in the following example:

```
<Thread THREAD_SEQUENCE="Q1_R4">
```

THREAD_SEQUENCE: (obligatory) internal identifier for the related question.

- Qxx_Ryy: Qxx is the ID of the original question, and Ryy is the rank of the thread in the list of results returned by a search engine for the original question Qxx.

Data Format – Thread (optional)

A thread has one obligatory and one optional attribute as in the following example:

```
<Thread THREAD_SEQUENCE="Q3_R1"
SubtaskA_Skip_Because_Same_As_RelQuestion_ID="Q2_R21">
```

SubtaskA_Skip_Because_Same_As_RelQuestion_ID: (optional) present when the current related question thread has already appeared as a related question to some other OrgQuestion; the value of the attribute is the THREAD_SEQUENCE of the first occurrence of this thread in the dataset

Data Format – RelQuestion

The structure of a RelQuestion is the following:

<RelQuestion ...>
  <RelQSubject> text </RelQSubject> → The subject of the related question
  <RelQBody> text </RelQBody> → The main body of the related question
</RelQuestion>

Data Format – RelQuestion

Each RelQuestion tag has a list of attributes, as in the following example:

```
<RelQuestion RELQ_ID="Q1_R4" RELQ_RANKING_ORDER="4"
RELQ_CATEGORY="Advice and Help" RELQ_DATE="2013-05-02 19:43:00"
RELQ_USERID="U1" RELQ_USERNAME="ankukuma"
RELQ_RELEVANCE2ORGQ="PerfectMatch“>
```

Data Format – RelQuestion

Each RelQuestion tag has a list of attributes, as in the following example:

<RelQuestion RELQ_ID="Q1_R4" RELQ_RANKING_ORDER="4"
RELQ_CATEGORY="Advice and Help" RELQ_DATE="2013-05-02 19:43:00"
RELQ_USERID="U1" RELQ_USERNAME="ankukuma"
RELQ_RELEVANCE2ORGQ="PerfectMatch">

RELQ_ID: the same as for the thread (as there is a 1:1 correspondence between a RelQuestion and its thread)

Data Format – RelQuestion

Each RelQuestion tag has a list of attributes, as in the following example:

<RelQuestion RELQ_ID="Q1_R4" RELQ_RANKING_ORDER="4"
RELQ_CATEGORY="Advice and Help" RELQ_DATE="2013-05-02 19:43:00"
RELQ_USERID="U1" RELQ_USERNAME="ankukuma"
RELQ_RELEVANCE2ORGQ="PerfectMatch">

RELQ_RANKING_ORDER: the rank of the related question thread in the list of results returned by a search engine for the original question

Data Format – RelQuestion

Each RelQuestion tag has a list of attributes, as in the following example:

<RelQuestion RELQ_ID="Q1_R4" RELQ_RANKING_ORDER="4"
RELQ_CATEGORY="Advice and Help" RELQ_DATE="2013-05-02 19:43:00"
RELQ_USERID="U1" RELQ_USERNAME="ankukuma"
RELQ_RELEVANCE2ORGQ="PerfectMatch">

RELQ_CATEGORY: the question category

According to the Qatar Living taxonomy, the examples of these categories: Advice and Help, Beauty and Style, Cars and driving, Computers and Internet, Doha Shopping, Education, Environment, Family Life in Qatar, Funnies, Health and Fitness, Investment and Finance, Language, Moving to Qatar, Opportunities, Pets and Animals, Politics, Qatar Living Lounge, Qatari Culture, Salary and Allowances, Sightseeing and Tourist attractions, Socialising, Sports in Qatar, Visas and Permits, Welcome to Qatar, Working in Qatar.

Data Format – RelQuestion

Each RelQuestion tag has a list of attributes, as in the following example:

<RelQuestion RELQ_ID="Q1_R4" RELQ_RANKING_ORDER="4"
RELQ_CATEGORY="Advice and Help" RELQ_DATE="2013-05-02 19:43:00"
RELQ_USERID="U1" RELQ_USERNAME="ankukuma"
RELQ_RELEVANCE2ORGQ="PerfectMatch">

RELQ_DATE: date of posting

Data Format – RelQuestion

Each RelQuestion tag has a list of attributes, as in the following example:

<RelQuestion RELQ_ID="Q1_R4" RELQ_RANKING_ORDER="4"
RELQ_CATEGORY="Advice and Help" RELQ_DATE="2013-05-02 19:43:00"
RELQ_USERID="U1" RELQ_USERNAME="ankukuma"
RELQ_RELEVANCE2ORGQ="PerfectMatch">

RELQ_USERID: internal identifier for the user who posted the question; consistent across all questions and across all datasets

Data Format – RelQuestion

Each RelQuestion tag has a list of attributes, as in the following example:

```
<RelQuestion RELQ_ID="Q1_R4" RELQ_RANKING_ORDER="4"
RELQ_CATEGORY="Advice and Help" RELQ_DATE="2013-05-02 19:43:00"
RELQ_USERID="U1" RELQ_USERNAME="ankukuma"
RELQ_RELEVANCE2ORGQ="PerfectMatch">
```

RELQ_USERNAME: the name of the user who posted the question; consistent across questions and comments; note that users can change their names over time, and this field shows the latest name the user used (but this name is consistent across the questions, comments and the datasets)

Data Format – RelComment

Each RelComment tag has a list of attributes, as in the following example:

```
<RelComment RELC_ID="Q104_R22_C1" RELC_DATE="2012-01-09 11:39:52"
RELC_USERID="U2011" RELC_USERNAME="drsam"
RELC_RELEVANCE2ORGQ="Bad" RELC_RELEVANCE2RELQ="Good">
```

Data Format – RelComment

Each RelComment tag has a list of attributes, as in the following example:

```
<RelComment RELC_ID="Q104_R22_C1" RELC_DATE="2012-01-09 11:39:52"
RELC_USERID="U2011" RELC_USERNAME="drsam"
RELC_RELEVANCE2ORGQ="Bad" RELC_RELEVANCE2RELQ="Good">
```

RELC_ID: internal identifier of the comment

Data Format – RelComment

Each RelComment tag has a list of attributes, as in the following example:

```
<RelComment RELC_ID="Q104_R22_C1" RELC_DATE="2012-01-09 11:39:52"
RELC_USERID="U2011" RELC_USERNAME="drsam"
RELC_RELEVANCE2ORGQ="Bad" RELC_RELEVANCE2RELQ="Good">
```

RELC_USERID: internal identifier of the user posting the comment

Data Format – RelComment

Each RelComment tag has a list of attributes, as in the following example:

<RelComment RELC_ID="Q104_R22_C1" RELC_DATE="2012-01-09 11:39:52"
RELC_USERID="U2011" RELC_USERNAME="drsam"
RELC_RELEVANCE2ORGQ="Bad" RELC_RELEVANCE2RELQ="Good">

RELC_USERNAME: the name of the user who posted the comment; consistent across questions and comments; note that users can change their names over time, and this field shows the latest name the user used

Data Format – The structure of Comments

Comments are structured as follows:

<RelComment ...>

  <RelCText> text </RelCText> → The text of the comment

</RelComment>