

Applying Learning To Rank Techniques in a Translator-Expert Retrieval Framework

Mihai Lupu

Information and Software Engineering Group
Vienna University of Technology
Favoritenstrasse 9-11/188
A-1040 Vienna, Austria
Email: lupu@ifs.tuwien.ac.at

Navid Rekabsaz

Faculty of Informatics
Vienna University of Technology
Email: rekabsaz.n@gmail.com

Abstract—*TODO: Abstract*****

I. INTRODUCTION

*** TODO: Introduction ***

II. CASE STUDY

The designed framework can be categorized in domain of on-line translation. It aims to find the best translator-corrector pair for a client who seeks for a reasonable offer to translate a document.

Every expert creates a profile by registering in the framework. The expert can play the role of translator or corrector. The profile consists of offered price per word for translation/correction, translation/correction time per word, experience in a special context and proficiency in a specific language pair (source and destination languages). Based on client's document, the system figures out the value of each aspect and offers the first three most related translator-corrector pairs.

In order to evaluate the proficiency of experts, an evaluation system based on feedbacks of correctors, clients and evaluators is provided. After finishing translator's task, the corrector starts revising the text as well as assessing translation quality. The feedback of corrector which is a number between one to five is used for estimation the proficiency of translator. Then by delivering the translation, client's feedback is asked. The feedback of client is used to deploy and leverage ranking system. Apparently, an evaluation team consisted of business experts are consistently evaluating the quality of ranking system.

In the rest of the section, we denote the framework's elements. As it is depicted in the following formula, the profile of each expert related to the query is a vector where q is the corresponding query. In the formula, t_q and c_q stand for estimated time and cost proposed by the expert. The prefix T and C stand for role of expert as translator or corrector. s_q denotes the experience or knowledge of expert in a special context which is calculated based on similarity between query document and all previous-translated documents. The technique applied for calculating s_q is explained further. p shows the proficiency of expert and how previous tasks have been estimated by other experts. It is being achieved by

calculating the average value of correctors' feedbacks. The value is between one (very bad) to five (perfect).

$$\overrightarrow{Profile}_q = \{t_{T,q}, t_{C,q}, c_{T,q}, c_{C,q}, s_q, p\}$$

Question: How should we handle the underlying relation between s_q and p ?

The evaluation results are also denoted as follows. In the following formula e stands for evaluation value which can be between one (very bad) and five (perfect). Because of some practical considerations, delivery time t is used in evaluation process (not separated schedule time of translator and corrector).

$$\overrightarrow{Evaluation}_q = \{t, c_{T,q}, c_{C,q}, s_{T,q}, s_{C,q}, p_T, p_C, e\}$$

III. METHODS

The subjects which are studied during design of the framework can be classified into two main topics. First aggregating results of similarity between query document and previous translated documents in order to estimate expert's proficiency. Second applying a learning to rank method to find and offer the best experts.

A. Aggregation Functions

In order to calculate experience of expert related to client's document, we need to figure out the similarity between query document and all previous-translated documents.

***TODO: Introducing aggregation methods and choosing one for the framework ***

B. Learning To Rank

***TODO: Ranking methods goes here. Comparing between Point-wise, Pair-wise and List-wise and choosing the most related one for the framework ***

IV. APPLY THE METHODS AND RESULTS

Question: Is it necessary to consider just a specific language pair? Or ML can be applied on whole the data.

***Question: Based on discussion, there are current formulas:

$$SCORE_{T,q} = \{c_{T,q}, s_{T,q}, p\}$$

$$SCORE_{C,q} = \{c_{C,q}, s_{C,q}, p\}$$

$$SCORE_q = \alpha.SCOR_{E_{T,q}} + \beta.SCOR_{E_{C,q}}$$

$$RANK_q = \{t_q, e_q, SCORE_q\}$$

How can we achieve $SCORE_{T,q}$ and $SCORE_{C,q}$? ***

***TODO: Evaluation ***

V. CONCLUSION

***TODO: conclusion ***

REFERENCES

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.