Expert Retrieval Report - No. 1

Seminar and Project in Software Engineering and Internet Computing

Navid Rekabsaz

(1129057, rekabsaz.n@gmail.com)

19th September 2013

Contents

1 Feedback Data **2**

2 Overview On Reviewed Papers **2**

2.1 CSIRO ICT Centre [3] 2

2.2 Georgetown University [1] 2

2.3 ICTNET [2] 3

2.4 IRIT [4] 3

2.5 University of Delaware [5] 3

3 Ideas and Brainstorming **3**

4 Appendix A - Feedback data samples **6**

1 Feedback Data

After every translation, a pre-selected corrector revises the translation. After reviewing, the corrector assesses the translation and leave a feedback on it. This feedback consists of a comment and number which evaluate the quality of translation from one to five (five is the best). These feedbacks are also categorized in different aspects of translation like content, grammar, style, accuracy and etc. For each translation, the value related to each criterion (price, time, text matching and cooperation) can be fetched from database.

Since the website is working from 2011, there are about 2000 feedback records in database which related to different translators and language pairs.

Some sample records of database are shown in Appendix A.

2 Overview On Reviewed Papers

Since the defined problem in 2012 TREC Contextual Suggestion Track is some how similar to the current problem, I focused on reviewing the solution proposed by participants of this Track. In TREC Contextual Suggestion Track, different factors like personal preferences, location and time influence on ranking of venues. I have tried to concentrate on the way different papers have encountered with the ranking problem in order to gain some ideas for current problem.

Here is a short brief of some of them:

2.1 CSIRO ICT Centre [3]

Besides the solution presented for accumulating the data of venues, it offers a vector-space approach for identifying personal preferences. Armed with a preference vector for each profile, it scores each venue according to its cosine distance.

2.2 Georgetown University [1]

The represented solution for ranking seems very simple and straightforward. First, it creates a list of venues in different categories. Then based on scores for each category which is filled in user's profile, it sums per-user category scores with the scores for every venues in each category. Finally it returns the first top ten venues. Boosting well-known attractions, using learning techniques (SVMRank) for ranking venues in each category and defining a limit for each set are some special points in the paper.

2.3 ICTNET [2]

The final formula for ranking consisted of some fixed factors which are selected using feedback method. Although the formula seems very clear and practical, the way to achieve the factors is not explained clearly.

2.4 IRIT [4]

I found the paper quite interesting as it acquired the first ranking between other participants. Like some other papers, they also create a Vector Space Model and calculate similarity based on cosine measure. The interesting point is that they distinguish between positive and negative preferences and subtract the similarity of negative preferences from positive ones. Therefore the results are most similar to global positive preferences and most dissimilar to global negative preferences.

2.5 University of Delaware [5]

The similarity calculation in the paper is very similar to [4]. The main difference is that it uses average of similarity for negative and positive preferences. Time factor is not considered in rating. After rating the final results are filtered based on time.

3 Ideas and Brainstorming

Using data represented in Appendix A for every translation, a vector can be created that shows the relation between four criteria (time, price, text matching and cooperation) and final feedback. My idea for translators ranking consisted of two elements which both are based on the defined vector. It is formulated as follows where R_T ,H_T and P_T denote ranking value of translator, first element (Harmonic) and second element (Privilege). α is the tuning factor and can be between 0 and 1.

$$R_T = \alpha H_T + (1 - \alpha) P_T$$

The first element (H_T) which can be called Harmonic element shows how similar the vector is to well-ranked results and how far it is from bad-ranked results. The similarity is calculated by cosine measure. In the following formula O_T denotes translator's offer and F_i stands for vector of feedback. For each feedback, the feedback rate is shown in r_i. The formula is depicted as follows:

$$H_T = \sum_{i=0}^{numberoffeedbacks} r_i SIM(O_T, F_i)$$

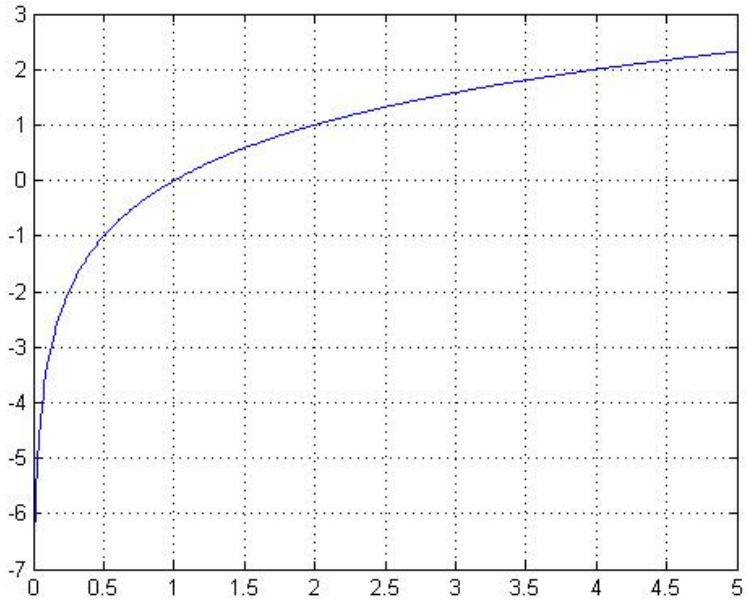

Figure 1: Suggested function for normalizing the criteria.

The second element (P_T) which can be called Privilege element shows how good the offered criterion in comparison to a standard value is. In order to find the value, we need a Standard Vector (S) which can be achieved by calculating the average of all offered values by translators in every criterion. Here is the formulas for calculating S and G_T. $O_{T,c}$ stands for offer value by a translator for a specific criterion. The function f is used for normalizing the value. Finally n denotes number of offers by translators

$$S = \sum_{i=0}^{n} \frac{O_{T,c}}{n}$$

$$G_T = \sum_{c=0}^{numberofcriteria} \frac{f(O_{T,c})}{S_c}$$

I suggest using $\log_2(x)$ as function f. The reason for selecting base two is decreasing radically the effect of division when offer is more than two times bigger than standard value. The suggested function is depicted in Figure 1 :

References

[1] Hui Yang Nazli Goharian Steve Kunath Ophir Frieder Andrew Yates, Dave DeBoer. (not too) personalized learning to rank for contextual suggestion. *TREC Conference*, 2012.

[2] Tong Wu Bingyang Liu. Ictnet at context suggestion track trec 2012. *TREC Conference*, 2012.

[3] Cecile Paris David Milne, Paul Thomas. Finding, weighting and describing venues: Csiro at the 2012 trec contextual suggestion track. *TREC Conference*, 2012.

[4] Guillaume Cabanac Gilles Hubert. Irit at trec 2012 contextual suggestion track. *TREC Conference*, 2012.

[5] Peilin Yang and Hui Fang. An exploration of ranking-based strategy for contextual suggestion. *TREC Conference*, 2012.

4 Appendix A - Feedback data samples

Table 1: Feedback data sample.

Corrector	Translator	Language	Category	Rating	Comment
Bettina	Sabine	fr->de-DE	Content	5	
Bettina	Sabine	fr->de-DE	Formatierung	5	
Bettina	Sabine	fr->de-DE	Level of language	3	Korrekturen siehe Änderungen in doc-Datei, die sich hier leider über "upload a file" nicht hochladen lässt, sondern ausnahmsweise per E-mail folgt.
Brigitte	Jakarta	de-DE->jv	Grammar	5	
Brigitte	Jakarta	de-DE->jv	Formatierung	4	
Brigitte	Jakarta	de-DE->jv	Completeness	5	
Brigitte	Jakarta	de-DE->jv	Style	5	
Brigitte	Jakarta	de-DE->jv	Content	3	
Brigitte	Jakarta	de-DE->jv	Level of language	3	
Brigitte	Jakarta	de-DE->jv	Spelling	5	
Brigitte	Jakarta	de-DE->jv	Accuracy	5	
Brigitte	Yogyakarta	de-DE->jv	Accuracy	5	
Brigitte	Yogyakarta	de-DE->jv	Completeness	5	
Brigitte	Yogyakarta	de-DE->jv	Style	5	
Brigitte	Yogyakarta	de-DE->jv	Grammar	5	
Brigitte	Yogyakarta	de-DE->jv	Spelling	5	
Heinz	Olga	de-DE->ru	Formatierung	4	Die Eigennamen (lateinische Buchstaben - z.B. nativy) sind in einer vom sonstigen russischen Text abweichenden Schriftart wiedergegeben - das fällt leider auf.
Heinz	Olga	de-DE->ru	Level of language	5	Sehr gutes Russisch!
					Continued on next page

Table 1 – continued from previous page

Corrector	Translator	Language	Category	Rating	Comment
Heinz	Olga	de-DE->ru	Content	3	Einige Wörter wurden nicht ganz korrekt übertragen (statt "Auftrag" - "Projekt"); eine wesentliche Wortgruppe wurde weggelassen (beim Rabatt: "für 4 Monate"). Die Präzision der Übertragung lässt an einigen Stellen zu wünschen übrig.
Jakarta	Yogyakarta	de-DE->jv	Style	4	
Jakarta	Yogyakarta	de-DE->jv	Accuracy	4	
Jakarta	Yogyakarta	de-DE->jv	Completeness	4	
Jakarta	Yogyakarta	de-DE->jv	Spelling	4	
Jakarta	Yogyakarta	de-DE->jv	Grammar	4	
Sonja	Nicole	de-DE->en	Content	0	
Sonja	Nicole	de-DE->en	Level of language	0	
Sonja	Nicole	de-DE->en	Formatierung	0	
Sonja	Sabine	de-DE->en	Level of language	0	
Sonja	Sabine	de-DE->en	Content	0	
Sonja	Sabine	de-DE->en	Formatierung	0	
Yogyakarta	Brigitte	de-DE->jv	Completeness	5	
Yogyakarta	Brigitte	de-DE->jv	Accuracy	5	
Yogyakarta	Brigitte	de-DE->jv	Style	5	
Yogyakarta	Brigitte	de-DE->jv	Spelling	5	
Yogyakarta	Brigitte	de-DE->jv	Grammar	5	
Yogyakarta	Semarang	de-DE->jv	Level of language	2	
Yogyakarta	Semarang	de-DE->jv	Formatierung	2	
Yogyakarta	Semarang	de-DE->jv	Content	3	