

# Are Two Heads Better than One? Crowdsourced Translation via a Two-Step Collaboration of Non-Professional Translators and Editors

**Rui Yan, Mingkun Gao, Ellie Pavlick and Chris Callison-Burch**

Computer and Information Science Department,

University of Pennsylvania, Philadelphia, PA19104, USA

**{ruiyan, gmingkun, epavlick, ccb} @seas.upenn.edu**

## Experiment and Evaluation

**Crowdsourcing Preliminary**

## Graph based Crowdsourcing Translation Model

* **Evaluation metric**
  + BLEU: Bilingual Evaluation Understudy score
  + 4 references from professional translator as ground truth set
* **Baselines**
  + Random
  + Oracle: based on segment level and Turker level separately, and both based on translation only and translation plus post-editings
  + Lowest TER: to select the translation with the minimum average TER
  + Linear combined regression (results directly reported\*)
  + Graph based ranking based on translation only
  + Graph based ranking based on translator/editor collaboration
* **Experimental results**
* **Problem formulation**
  + Given a set of candidate translation for a particular source sentence, the goal is to choose the best output translation.
  + We form two graphs: the first graph (GT) represents Turkers (translator/ editor pairs) as nodes; the second graph (GC) represents candidate translated and edited as nodes.
  + The two graphs (GT and GC) are combined as sub-graphs of a third graph (GTC). Edges in GTC connect author pairs (nodes in GT) to the candidate they produced (nodes in GC).

G = (V, E)

= (VT, VC, ET, EC, ETC)

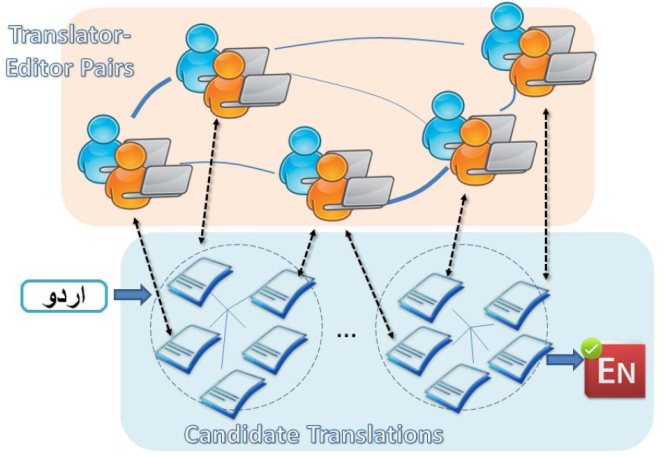
GC = (VC, EC)

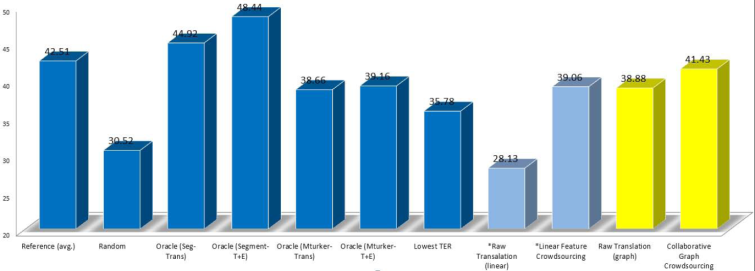
GT = (VT, ET)

GTC = (VTC, ETC)

VTC = VT U VC

* **Framework**
* **A viable mechanism for creating large-scale training data for Natural Language Processing techniques, i.e., machine translation, etc.** 
  + Low cost
  + Fast turn-around time
  + Especially useful under the scenario of aiding “low resource” languages.
* **Potential Pitfalls** 
  + Non-professionals
  + Low quality
* **Solution**
  + Automated quality control!
  + Two-Step collaboration between translators and post-editors based on graph ranking

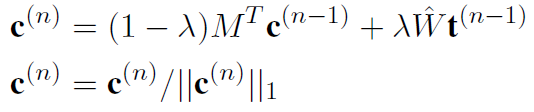
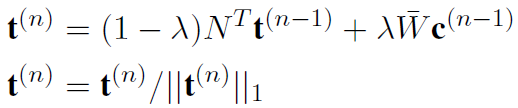
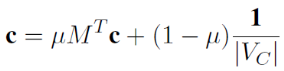
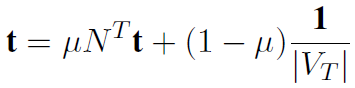
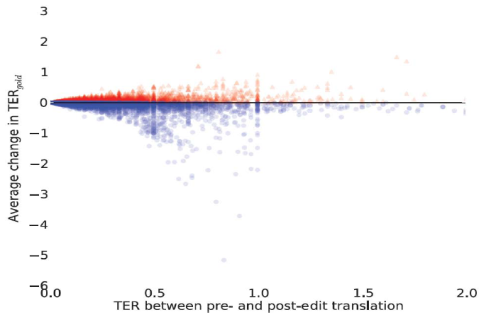




The 2-step collaborative crowdsourcing translation model based on graph ranking framework includes three sub-networks. The undirected links between users denotes translation-editing collaboration. The undirected links between candidate translations indicate lexical similarity between candidates. A bipartite graph ties candidate and Turker networks together by authorship (to make the figure clearer, some linkage is omitted). A dashed circle indicates the group of candidate translations for a single source sentence to translate.

**Crowdsourcing Translation**

* **Setups**
  + Data set: 1,792 Urdu sentences, paired with English translations.
  + Each Urdu sentence was translated redundantly by 3 distinct translators
  + Each translation was edited by 3 separate native English speakers as post-editors
  + 52 Turkers took part in the translation task, each translating 138 sentences
  + 320 Turkers participated in editing task, average 56 sentences edited each
* **Some editors make improvements but it is tricky to automatically identify the good ones**
  + Agreement picks out lazy editors!



## Example Rankings

* **Contributions**
  + An analysis of the difficulties posed by a 2-step collaboration between editors and translators in crowdsourcing environment
  + A graph based algorithm for quality control in selecting the best translation

* **Parameter effect in classification**
  + The best results for both GBG-g and GBG-s are achieved when θ is 1.
  + On NG20, GBG-g can perform slightly better with both small and large θ values than with the medium values.

## Conclusions

* **Inter-Graph Ranking**
  + A candidate sentence is important if 1) it is similar to many other proposed candidates and 2) it is authored by better qualified translators and/or post-editors
  + A translator/editor pair is better qualified if 1) the editor is collaborating with a good translator and vice versa and 2) the pair has authored important candidate translations

Introducing the saliency scores for candidate sentence and turker pairs, we can formulae as:

* + Homogeneity:
  + Heterogeneity:
  + Computing steps: 1) compute the saliency scores of candidates and then normalize and 2) compute the saliency scores of turker pairs and then normalize. Repeat until convergence.
* **Intra-Graph Ranking**
  + Pagerank schema:
* To overcome the shortcomings of the existing method, we propose an adaptive concept resolution model
* ACR can adaptively learn a concept border from an ontology taking into consideration of the characteristics of a particular document collection.
* Then this border can provide a tailor-made semantic concept representation for a document coming from the same domain.
* ACR is applicable in both classification task.
* Supported by NSFC under Grant No. 60673129 and 60773162, 863 Program under Grant No. 2007AA01Z154, the 2008/09 HP Labs Innovation Research Program, and National Key Technology R&D Pillar Program in the 11th Five-year Plan of China (Research No: 2009BAH47B00).

## Acknowledgement

* Phone: 8610-62755592
* Email: zhy@cis.pku.edu.cn
* Affiliation: Peking University

## Contact Information

## Conclusions