<Project Review>

* CS410, 2020, Fall
* Authors : Changsoo Kim, Hongseok Ha

1. Overall

LARAM(2011) improved LARA(2010) in that it allows finding latent topics without specifying the seed words of the topics. LARAM can be divided into two modules, which are 1) finding latent topic aspects (aspect modeling module) and 2) finding ratings on each identified aspect (rating analysis module).

Although we tried to understand the details of LARAM based on the original author’s source code and python code from other sources, it was difficult for us to figure out all the details of the paper.

Thus, as we mentioned in our progress report, we tried to implement the paper by replacing the aspect modeling module with LDA, and the rating analysis module with LRR (Latent Rating Regression).

However, it was difficult for us to link the LDA results to LRR. Also, although we converted the java code of the original author to implement a python-version of LRR (which we could not find other references), there were subtle numeric differences in our intermediate results.

We could not be sure whether the previous mismatch was the fundamental cause of following failures, but many unknown causes leaded us to a problem in which LBFGS (which was used as an optimizer, to  
minimize beta parameters) fails in line search, so that the beta parameters could not converge.  
  
In conclusion, the beta parameters could not be updated, so that the EM algorithm could not worked properly, resulting the maximization of log-likelihood (that we expected) could not happen.

We found that we should try to understand the paper more sufficiently, before starting the code implementation. Also, to save time, utilizing libraries and packages would be very helpful.

With this project, we were able to empathize with researchers who were making great efforts to advance their research in their fields.

1. Difficulties

There were ambiguous words, that we could not be sure whether they should be included in stopwords. For instance, the phrase ‘n’t’ was not included in stopwords of the nltk package. Although it seems to mean ‘not’ with some verbs, it would mean nothing, without any verbs.

Although the python package ‘scipy’ provides L-BFG-S option in an optimizer, structures of parameters were different, so that using the function imported from scipy was difficult. That’s why we chose to convert the original author’s java code of LRR. However, it resulted in the numerical mismatch in intermediate results.

In LARAM, the Z values from aspect segmentation should be utilized in LRR, to improve the overall performance, but we could not implement that part, for lack of understanding of the paper and limit of time.

We found that it is a tough task to implement a paper to well-structured codes only by looking at the formula and diagram in the paper. Additional resources (such as the author’s presentation materials, that provides details and easier explanation) seems to be helpful to implement the paper in detail.