Supplementary material for Domain Adapted Word Embeddings for Improved Sentiment Classification

Prathusha K Sarma, Yingyu Liang and William A Sethares

University of Wisconsin-Madison
{kameswarasar,sethares}@wisc.edu,
yliang@cs.wisc.edu

1 Derivation of α for minimizing cluster variance.

A domain adapted word embedding $\mathbf{w}_{i,DA}$ is then obtained as $\hat{\mathbf{w}}_{i,DA} = \alpha\bar{\mathbf{w}}_{i,DS} + (1-\alpha)\bar{\mathbf{w}}_{i,G}$.

Assume each document d_i is expressed as the sum of word embeddings in the document, then we have,

$$d_i = \sum_{j=1}^{n} \hat{\mathbf{w}}_j$$

$$d_i = \sum_{j=1}^{n} (\bar{\mathbf{w}}_{j,G} + \alpha(\bar{\mathbf{w}}_{j,DS} - \bar{\mathbf{w}}_{j,G}))$$

$$d_i = d_{g_i} + \alpha\bar{d}_i.$$

Now let us assume we have N documents out of which k are positive. We can express every positive document as $d_{p_i} = d_{g_{p_i}} + \alpha\bar{d_{p_i}}$. Similarly we can express negative documents as $d_n = d_{g_{n_i}} + \alpha\bar{d_{n_i}}$. We can calculate the center of each cluster as follows to get,

$$\mu_p = \frac{1}{k}\sum_{i=1}^{k}(d_{p_i})$$

$$\mu_p = \frac{1}{k}\sum_{i=1}^{k}(d_{g_{p_i}} + \alpha\bar{d_{p_i}})$$

$$\mu_p = \hat{\mu}_p + \alpha\bar{\mu}_p$$

where, $\bar{\mu}_p = \frac{1}{k}\sum_{i=1}^{k}\bar{d}_{p_i}$ and $\hat{\mu}_p = \frac{1}{k}\sum_{i=1}^{k} d_{g_{p_i}}$. Similarly we can get the negative cluter center $\mu_n = \hat{\mu}_n + \alpha\bar{\mu}_n$ with $\bar{\mu}_n = \frac{1}{N-k}\sum_{i=1}^{N-k}\bar{d}_{n_i}$ and $\hat{\mu}_n = \frac{1}{N-k}\sum_{i=1}^{N-k} d_{g_{n_i}}$.

One way to determine α is by selecting α such that the two document clusters are tightly packed, i.e the variance within each cluster is minimized. This can be cast as the following optimization problem,

$$\min_{\alpha} = \frac{1}{k}\sum_{i=1}^{k}||d_{p_i} - \mu_p||_2^2 + \frac{1}{N-k}\sum_{i=1}^{N-k}||d_{n_i} - \mu_n||_2^2,$$

$$\min_{\alpha} = \frac{1}{k}\sum_{i=1}^{k}||(d_{g_{p_i}} - \hat{\mu}_p) - \alpha(\bar{\mu}_p - \bar{d_{p_i}})||_2^2 + \frac{1}{N-k}\sum_{i=1}^{N-k}||(d_{g_n}$$

$$\min_{\alpha} = \frac{1}{k}\Big[-2\alpha(d_{g_{p_i}} - \hat{\mu}_p)^\top(\bar{\mu}_p - \bar{d}_{p_i}) + \alpha^2(\bar{\mu}_p - \bar{d}_{p_i})^\top(\bar{\mu}_p - \bar{d}$$

$$\frac{1}{N-k}\Big[-2\alpha(d_{g_{n_i}} - \hat{\mu}_n)^\top(\bar{\mu}_n - \bar{d}_{n_i}) + \alpha^2(\bar{\mu}_n - \bar{d}_{n_i})^\top(\bar{\mu}_n$$

Solving this optimization problem for α we get,

$$\alpha = \frac{\frac{1}{k}\sum_{i=1}^{k}(d_{g_{p_i}} - \hat{\mu}_p)^\top(\bar{\mu}_p - \bar{d_{p_i}}) + \frac{1}{N-k}\sum_{i=1}^{N-k}(d_{g_{n_i}} - \hat{\mu}_n)^\top}{\frac{1}{k}\sum_{i=1}^{k}(\bar{\mu}_p - \bar{d}_{p_i})^\top(\bar{\mu}_p - \bar{d}_{p_i}) + \frac{1}{N-k}\sum_{i=1}^{N-k}(\bar{\mu}_n - \bar{d}_{n_i})^\top(}$$

Finally, $\alpha = \max(0, \min(\alpha, 1))$.

2 Dimensions of word embeddings

Dimensions of generic, DS and DA embeddings used in experiments is provided in the table below.

Word embedding	Dimension
GloVe	100
word2vec	300
LSA	70
CCA-DA	68
KCCA-DA	68
GloVe common crawl	300
KCCA-DA(GlvCC)	300
concSVD	300

Table 1: This table presents the average dimensions of LSA, generic and DA word embeddings.