# Linking Tweets with Monolingual and Cross-Lingual News using Transformed Word Embeddings

Aditya Mogadala[1], Dominik Jung[2] and Achim Rettinger[1]

[1] Institute AIFB, Karlsruhe Institute of Technology, Germany,
aditya.mogadala@kit.edu, rettinger@kit.edu
[2] Institute IISM, Karlsruhe Institute of Technology, Germany,
dominik.jung2@kit.edu

**Abstract.** Social media platforms have grown into an important medium to spread information about an event published by the traditional media, such as news articles. Grouping such diverse sources of information that discuss the same topic in varied perspectives provide new insights. But the gap in word usage between informal social media content such as tweets and diligently written content (e.g. news articles) make such assembling difficult. In this paper, we propose a transformation framework to bridge the word usage gap between tweets and online news articles across languages by leveraging their word embeddings. Using our framework, word embeddings extracted from tweets and news articles are aligned closer to each other across languages, thus facilitating the identification of similarity between news articles and tweets. Experimental results show a notable improvement over baselines for monolingual tweets and news articles comparison, while new findings are reported for cross-lingual comparison.

## 1 Introduction

On the web, growth of social media platforms has offered numerous opportunities with several challenges to solve. Twitter [3] is one such social media platform that allows its users to share 140 characters of text messages (popularly known as **tweets**) in multiple languages with their friends or followers. Tweets may contain personal information or a confined description about an event motivated by the traditional media such as online news articles. Studies [1] have shown that 85% of the tweets are news affiliated. Though only some tweets acknowledge news articles by explicitly linking them, most of them do not. This implicit linking of tweets with the news topics provide novel insights. For example, most of the traditional media companies that publish online news write only facts about an event. However, identifying relevant tweets for the corresponding news will append people opinion. Furthermore, attaching tweets with the news articles will allow to understand the multi-dimensional view about controversial topics, thus

[3] https://twitter.com/

empowering the editor of an article to modify upcoming or following articles based on veracity.

Howbeit, due to the differences in word usage across informal tweets and the attentively drafted writings like news articles make this linking challenging. Nevertheless, different approaches are pursued to solve the problem. Initially, monolingual comparison of tweets with news articles is achieved by comprehending commonality between the topics using unsupervised topic models [2]. Although a scalable approach, it fails to capture importance of words and their differences across corpora. A graph-based latent variable model [3] was further introduced for finding short text correlations using microblog hashtags and news articles named entities. Even though it addresses earlier drawbacks by giving importance to keywords such as named entities in news articles. It still ignores other large chunk of vocabulary. Krestel et al. [4] followed a different path by posing the comparison of tweets with news as relevance assessment problem and designed supervised binary classifier with many hand-crafted features. Yet supervised, the hand-crafted features limit its scalability. Further, aforementioned approaches ignore the multilingual aspect of the published news. Nowadays most of the online news about any event is multilingual. Identification of a news article belonging to single language is not enough to cover the collective views about an event.

In this paper, we overcome the limitations of earlier approaches and propose a new scalable framework to support tweets with monolingual and cross-lingual news article comparison. Our framework leverages monolingual [5] and bilingual [6] word embeddings acquired from tweets and news articles as basic units for bridging the word usage gap across these collections. Furthermore, non-linear transformation of tweet word embeddings is performed to make it closer to the news article word embeddings using manifold alignment with Procrustes analysis [7]. Work closely related to our approach is by Tan et al. [8] who perform lexical comparison of words observed in tweets and Wikipedia belonging to same language with only linear transformation, while we perform non-linear transformation and also across languages. Three main contributions are summarized as follows:

1. Proposed an approach to classify tweets as to how relevant they are for a given news article in more than one language.
2. New evaluation corpora is created for monolingual and cross-lingual tweets to news article comparison.
3. Lexical and task specific evaluation results are presented on two different datasets.

## 2 Related Work

Most of our research is closely related to the work that identifies relevance of tweets with online news or perform event detection. We divide each of the related works into separate categories.

### 2.1 Event Detection in Tweets

Analyzing information flow about the events as they emerge is an important aspect of event detection in tweets. Several works used this information in various ways. Some approaches [9] collocated emerging events and classified them into different categories, while some [10] found sentiment from the detected events. Others detected events as trends to track public health [11], political abuse [12] and crisis communication [13].

### 2.2 News and Relevant Tweets

Several approaches have been explored to identify relevant noisy tweets with the lengthy news articles. Initially, a semantic enrichment framework [14] was built to link news articles and tweets by identifying possible correlations to provide personalized news recommendations. Jin et al. [15] viewed the problem from different perspective and introduced a dual latent Dirichlet allocation model to jointly learn two sets of topics. Later, a more sophisticated unsupervised topic modeling [2] approach was proposed for finding overlap of topic distribution between tweets and news articles obtained from New York Times[4].

### 2.3 Distributed Representations

Distributed word representations [5] has shown significant improvements in many NLP tasks [16]. Different variations of them such as bilingual [6] and polylingual [17] are also obtained by projecting multiple or pair of languages into the shared semantic space. Also, word representations were extended to meet requirements of the short or noisy text [18, 19].

## 3 Monolingual Word Usage Characteristics

To understand the characteristics of word usage, initially news articles in German and English are collected between January, 2015 and December, 2015. To have a good overlap of topics, keywords[5] are extracted from news articles to be used as queries for collecting tweets belonging to the same period with Twitter search API[6]. Acquired tweets are then polished by removing URLs, user mentions, "#" symbol of the hashtags, and all re-tweets. Additionally, Glove[7] is used to obtain word embeddings with 400 dimensions for both collections. Size of the final document sets and the vocabulary extracted from Glove is listed in the Table 1.

Word embeddings for each collection are now used to effectively comprehend the word usage characteristics. Initially, top 10 common and frequent words observed in both collections are visualized with t-SNE [20]. We observed that the

[4] `http://www.nytimes.com/`
[5] `https://github.com/aneesha/RAKE`
[6] `https://dev.twitter.com/rest/public/search`
[7] `https://github.com/stanfordnlp/GloVe`

| Collection | Language | Documents | Vocabulary |
|---|---|---|---|
| News | English | 1027987 | 348419 |
| News | German | 198784 | 241014 |
| Tweets | English | 110731 | 47280 |
| Tweets | German | 56957 | 31887 |

**Table 1.** Collection Sizes

same words learned separately from tweets and news collection are highly separated. Furthermore to apprehend the difference in slangs, abbreviations etc., in both collections, we use frequent 5000 common vocabulary terms (both English and German) to perceive differences among their nearest neighbors. Based on rank biased overlap (RBO) measure [21, 8] which provides a comparison between incomplete and indefinite rankings, we observe a minimal average RBO measure of 0.2856 and 0.2589 for English and German respectively with parameters $\varphi = 0.9$ and $k = 100$. Thus exhibiting the difference in word usage among both collections. This motivates us to transform word embeddings learned with tweets closer to word embeddings learned using news articles or vice versa.

# 4 Transformed Word Embeddings (TWE)

Difference in the embeddings learned from two different collections such as tweets and news require bridging with embedding transformation. In this section, we formulate the problem and present our approach for monolingual and cross-lingual transformation.

## 4.1 Problem Formulation

Let, $T^l_{w_n} = \{t^l_{w_1}, t^l_{w_2}...t^l_{w_i}...t^l_{w_n}\}$ and $T^l_{e_n} = \{t^l_{e_1}, t^l_{e_2}...t^l_{e_i}...t^l_{e_n}\}$ represent set of words and their corresponding embeddings extracted from tweet collection respectively. Where $l$ is the language of tweets, $n$ is the size of vocabulary and each embedding is of dimension $t^l_{e_i} \in R^{1Xd}$. Similarly, $N^l_{w_m} = \{n^l_{w_1}, n^l_{w_2}...n^l_{w_i}...n_{w_m}\}$ and $N^l_{e_m} = \{n^l_{e_1}, n^l_{e_2}...n^l_{e_i}...n^l_{e_m}\}$ represent set of words and their embeddings of news corpora respectively. Where $l$ is the language of news corpora, $m$ is the size of vocabulary and each embedding is of dimension $n_{e_i} \in R^{1Xd}$.

Formally, now our research question is to identify common words $\{T^l_{w_c}, N^l_{w_c}\} = \{t^l_{w_i}, n^l_{w_i}\}^c_{i=1}$ and transform word embeddings in the tweet collection ($T^l_{e_c}$) closer to the embeddings of news collections ($N^l_{e_c}$) or vice versa. This transformation is based on the assumption that there prevails a transformation relationship between the vectors for the frequent words of each collection. Some approaches [8] have earlier performed this simple transformation only if the language of tweets and formal language corpora (e.g. news, Wikipedia) belong to same language. But, it is non-trivial if the language of tweets and formal language corpora differs.

In the following sections, we present the transformation of tweet embeddings closer to the monolingual or cross-lingual news embeddings.

### 4.2 Monolingual-TWE

Earlier approaches [8] assume only linear relationship between embeddings from different collections to perform transformation. Sometimes relationship needs to handle disturbances such as scaling and rotation. To cater such issues, we leverage manifold alignment using Procrustes analysis [7] to transform word embeddings of tweets closer to word embeddings of news articles with a three step procedure.

- Learning low-dimensional embeddings is cue for transformation. We already have low-dimensional embeddings $\{T^l_{e_c}, N^l_{e_c}\}$ of words observed in both tweet and news collection.
- To find the optimal values of transformation, Procrustes superimposition is done by translating, rotating and scaling the objects (i.e. rows of $T^l_{e_c}$ is transformed to make it similar to the rows of $N^l_{e_c}$). Transformation is achieved by
  - **Translation:** Taking mean of all the members of set to make centroids $(\sum_{i=1}^{c} \frac{T^l_{e_i}}{c}, \sum_{i=1}^{c} \frac{N^l_{e_i}}{c})$ lie at origin.
  - **Scaling and Rotation:** The rotation and scaling that maximizes the alignment is given by orthogonal matrix ($Q$) and scaling factor ($j$). They are obtained by minimizing orthogonal Procrustes problem [22] and is provided by Equation 1.

$$\arg\min_{j,Q} ||N^l_{e_c} - T^l_{e^*_c}||_F \tag{1}$$

    where $T^l_{e^*_c}$ a matrix of transformed $T^l_{e_c}$ values given by $jT^l_{e_c}Q$ and $||.||_F$ is the Frobenius norm constrained over $Q^TQ = I$.
- If $T^l_{w^*_c}$ represents the words of $T^l_{e^*_c}$ low-dimensional embeddings, then the final sets $\{T^l_{w^*_c}, N^l_{w_c}\}$ contains closer correspondence.

To understand the effectiveness of this transformation, we perform similar experiments as of § 3 in § 6.3.

### 4.3 Cross-Lingual-TWE

Comparison of vocabulary obtained from tweets in one language ($l_1$) with the vocabulary of news articles in another language ($l_2$) is not straightforward. To subdue this concern, we propose a two step approach.

- In the first step, news articles from two different languages are acquired to learn bilingual word distributed representations(i.e. bilingual embeddings). Aim of bilingual embeddings is to capture linguistic regularities across languages into a common semantic space such that English and German words (e.g. "wonderful" and "wunderbar") are neighbors in the t-SNE visualization, thus bridging the language gap.

- In the second step, cross-lingual transformation is achieved between word embeddings obtained from tweets in $l_1$ and word embeddings of news articles in $l_2$. As bilingual word embeddings of news articles in $l_1$ also share linguistic regularities from $l_2$, mapping word embeddings of tweets closer to the bilingual word embeddings of news articles of $l_1$ will also help to incorporate linguistic regularities of $l_2$. Consequently, transformation is attained in the similar way as § 4.2 between word embeddings of tweets and bilingual word embeddings of news articles belonging to same language.

**Step-1** To learn bilingual embeddings, we leverage the approach of Gouws et al. [6] as it is fast and scalable to jointly optimize the monolingual objective $M(\cdot)$ with the cross-lingual objective $\varphi(\cdot)$ (i.e. cross-lingual regularization term) to find the overall loss $L(\cdot)$. Documents in the news collection of languages $l_1$ and $l_2$ are used to learn monolingual models along with cross-lingual regularization term learned with parallel corpora (e.g. Europarl-v7). Overall loss function $L(\cdot)$ is given by Equation 2.

$$L(\cdot) = \min_{\theta^{l_1}, \theta^{l_2}} \sum_{l \epsilon \{l_1, l_2\}} \sum_{C^l} M^l(w_t, h; \theta^l) + \frac{\lambda \varphi(\theta^{l_1}, \theta^{l_2})}{2} \tag{2}$$

$\varphi(.)$ eliminates the need for word-alignment and makes an assumption that each word observed in the document of language $l_1$ can potentially find its alignment in the document of language $l_2$. Thus, the Equation 2 is now modified into Equation 3.

$$\begin{aligned} L(\cdot) = & \min_{\theta^{l_1}, \theta^{l_2}} \sum_{l \epsilon \{l_1, l_2\}} \sum_{C^l} M^l(w_t, h; \theta^l) \\ & + \frac{\lambda || \frac{1}{m} \sum\limits_{w_i \epsilon l_1}^{m} V_i^{l_1} - \frac{1}{n} \sum\limits_{w_i \epsilon l_2}^{n} V_i^{l_2} ||^2}{2} \end{aligned} \tag{3}$$

Where $V^{l_1}$ and $V^{l_2}$ are monolingual word vectors of the words in documents of languages $l_1$ and $l_2$ respectively and $C^l$ is monolingual corpus (e.g. News). $w_t$ is the predicted word in the context $h$ of a monolingual model.

**Step-2** We follow a similar procedure as of § 4.2 but with a different set of embeddings.

- Low-dimensional embeddings that are used initially are $\{T_{e_c}^{l_1}, N_{e_c}^{l_1}\}$ of words observed in both tweet and news collection belonging to the same language. Here, $N_{e_c}^{l_1}$ represents **bilingual embeddings**.

Transformation is now achieved by translating, rotating and scaling the objects (i.e. rows of $T_{e_c}^{l_1}$ is transformed to make it similar to the rows of $N_{e_c}^{l_1}$) using the same procedure as described in § 4.2.

## 5 Experimental Setup

To evaluate our approach, we built a dataset for the cross-language and monolingual pairwise tweet and news article relevance assessment. Also, we used the existing monolingual comparisons corpora to compare with other approaches.

### 5.1 Corpus Creation

Unavailability of datasets for comparing news articles with the tweets in different languages compelled us to create our own. We created a gold standard dataset for monolingual and cross-lingual comparison across collections by acquiring some more tweets and news articles mainly in English and German in the same way as described in § 3.

Tweets with a single URL link to any news article are collected and carefully evaluated to see if it does not simply represent the news title or summary. If they only represent news title or summary then they are considered to be trivial and are removed. After basic preprocessing, using the keyword "Grexit" (the Greece exit of the European Union) around 18 tweets and 18 news articles (both English and German) are selected for further human evaluation.

### 5.2 Human Evaluation

The goal of the human evaluation is to get pairwise comparison scores between tweets and news. Thus, each participant had to rate a pair of documents with respect to their semantic similarity. Three different annotators who have English(E) and German(G) language skills were chosen for comparing pair of tweets and news based on scores listed in Table 2. At the end, a list of 628 relevance

| Score | Type | Description |
|---|---|---|
| 0 | Dissimilar | Tweet and news article are not about same topic. |
| 1 | Related | Tweet and news article share topic but important ideas in news is not represented in the tweet. |
| 2 | Similar | Tweet and news article are about same topic and important ideas in news is represented in the tweet |

**Table 2.** Similarity Scores

judgments (i.e. 162 between (E)Tweets and (E)News, 162 between (E)Tweets and (G)News and so on) were produced. A significance test with Kendall's $\tau$ is computed to test the consistency among user judgments. Results suggested that there is no significant difference in the score pairs of users (0.05 significance level). Specifically, the results showed that users have an similar understanding of the similarity assessment. To obtain the final score for each pair, similar to SemEval semantic similarity tasks[8] arithmetic mean was calculated between all user ratings. We term this resource as **Dataset-1**[9]. This dataset provides more fine-grain comparison as compared to other datasets [4] that provide only binary relevance.

### 5.3 Other Datasets

Evaluation of monolingual comparison is also performed on the other existing resources such as Krestel et al. [4]. This dataset consists of 1600 relevance judgments constituting 17 news articles covering different topics with the Tweets labeled as relevant or irrelevant for the each news article. We term this resource as **Dataset-2**.

### 5.4 Evaluation Metrics

For many pairwise semantic similarity tasks statistical correlation based measures have been used. Here, we use Pearson correlation coefficient ($r$) to evaluate our approaches on the dataset we created. While, measures like accuracy is used for other datasets.

## 6 Experimental Results

In this section, we present our experimental results on different datasets with variation in parameters.

### 6.1 Baselines

Two different baselines are used to compare with our approach.

**Latent Dirichlet Allocation (LDA)** Most of the earlier research [2, 3] have shown significant interest to compare news and tweets with LDA and its variations. We use the polylingual topic model [23] trained on English and German Wikipedia with 100 topics to support multiple languages. Similarity between tweet and news represented as topics vector is measured using cosine similarity.

**WTMF-G** Weighted Textual Matrix Factorization on Graphs (WTMF-G) [3] is one of the baseline that compare tweets and news based on a graph connected by hashtags, named entities or temporal information. To train the WTMF-G model we used regularization coefficient ($\lambda = 20$), weight of missing words as $w_n = 0.01$, number of neighbors ($k = 4$) and link weights ($\delta = 3$) as suggested in earlier research. Latent dimension of 100 is used to represent tweet and news, while similarity between them is calculated using cosine similarity.

[8] http://ixa2.si.ehu.es/stswiki/index.php/Main_Page

[9] http://people.aifb.kit.edu/amo/cicling2017/

### 6.2 TWE Implementation

Major parameters that affect training of Glove is the dimensionality of word embeddings and the size of word context window. We choose 25, 50, 100, 200, 400 word embedding dimensions and 5 words on left and right context window. Similarly, later for learning bilingual word embeddings we used Bilbowa tool[10] to learn same embedding dimensions as former with 5 word left context window and entire English-German Europarl-v7[11] as the parallel data. In both cases, count of words less than 2 in the entire corpus are discarded.

### 6.3 Monolingual Comparison

Before comparing monolingual news and tweets, we estimate the quality of embedding transformation achieved with **Monolingual-TWE** by performing similar experiments as in § 3. The transformation can be either from tweets to news (T2N) or in the opposite orientation (N2T). Though both of them have different transformation, we observed that they produce similar t-SNE visualization. Also, there is a slight decrease in distance between common words across collections as compared to without transformation. Average RBO measure using the top 5000 frequent terms observed in both tweets and news collections in German and English is recalculated to perceive the refinement. We perceived that there is an improvement of 24.4% and 21.2% for English and German respectively.

Now, tweets and news articles in Dataset-1 and Dataset-2 are represented as the tf-idf weighted average of transformed word embeddings. They are now used as input to SVM classifier[12] with default parameters to calculate accuracy and to cosine similarity for finding Pearson correlation. Furthermore, top performing embedding dimensions are identified based on Pearson correlation and accuracy measures using validation data of the datasets. Figure 1 and Figure 2 show the comparison of results with ((T2N)TWE and (N2T)TWE) and without (Non-TWE) transformation on different datasets. Once the top performing embedding dimensions are identified, testing data is used to compare different approaches with diverse measures in Table 3 and Table 4.

### 6.4 Cross-Lingual Comparison

For the cross-lingual comparison, we follow a similar procedure as in § 6.3. Since, news word embeddings incorporate bilingual information from both German and English, calculation of RBO measure between tweets and news without transformation is not appropriate. Hence, we calculate RBO measure after transformation to verify that it satisfies minimum threshold of 0.328, which in general fetch satisfactory results [8]. Now to compare tweets and news belonging to the dataset listed in § 5.1 across languages, we estimate the top performing embedding dimension based on Pearson correlation measure using the validation data

[10] `https://github.com/gouwsmeister/bilbowa`
[11] `http://www.statmt.org/europarl/`
[12] `https://www.csie.ntu.edu.tw/čjlin/libsvm/`

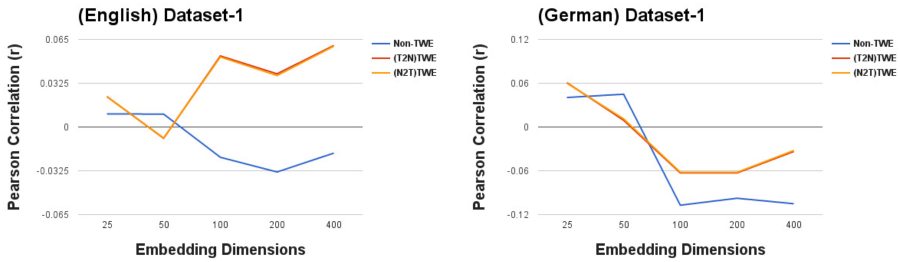


**Fig. 1.** Effect of Embedding Dimensions(Dataset-1)

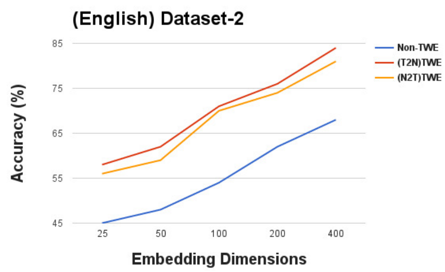


**Fig. 2.** Effect of Embedding Dimensions(Dataset-2)

| | Method | Dim | $r$ |
|---|---|---|---|
| **German** | No-Transformation | 400 | -0.1051 |
| | LDA-PTM [23] | 100 | 0.0445 |
| | WTMF-G [3] | 100 | 0.0498 |
| | (T2N)Monolingual-TWE | 25 | **0.0607** |
| | (N2T)Monolingual-TWE | 25 | 0.0601 |
| **English** | No-Transformation | 400 | -0.1193 |
| | LDA-PTM [23] | 100 | 0.0321 |
| | WTMF-G [3] | 100 | 0.0491 |
| | (T2N)Monolingual-TWE | 400 | **0.0605** |
| | (N2T)Monolingual-TWE | 400 | 0.0599 |

**Table 3.** Monolingual Tweets and News Comparison

of the dataset. Figure 3 show the comparison of results with (TWE) and without (Non-TWE) transformation. Once the top performing embedding dimension is identified, testing data is used to compare different approaches as provided in Table 5.

| Method | Dim | Accuracy |
|---|---|---|
| LDA-PTM [23] | 100 | 79.1% |
| Boosting [4] | - | 82.5% |
| (T2N)Monolingual-TWE+SVM | 400 | **83.1%** |
| (N2T)Monolingual-TWE+SVM | 400 | 81.0% |

**Table 4.** Accuracy (English)

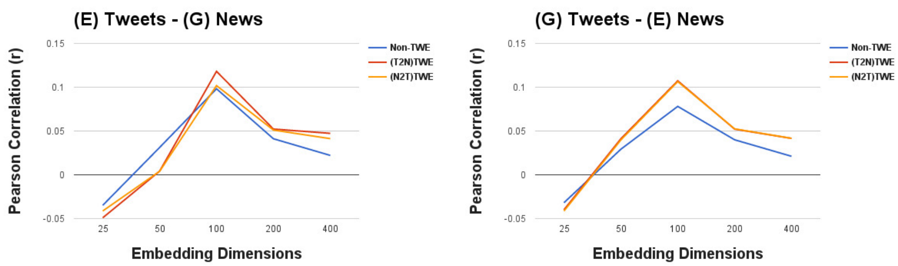


**Fig. 3.** Effect of Embedding Dimensions(Cross-Lingual)

| Method | | $r$ |
|---|---|---|
| **(E)Tweets - (G)News** | LDA-PTM [23] | 0.0821 |
| | (T2N)Cross-Lingual-TWE | **0.1181** |
| | (N2T)Cross-Lingual-TWE | 0.1018 |
| **(G)Tweets - (E)News** | LDA-PTM [23] | 0.0765 |
| | (T2N)Cross-Lingual-TWE | **0.1073** |
| | (N2T)Cross-Lingual-TWE | 0.1064 |

**Table 5.** Cross-Lingual Tweets and News Comparison With 100-Dimensions

# 7 Discussion

We start our analysis with results observed in the Table 3. It can be comprehended that the Monolingual-TWE (either T2N or N2T) achieved an commendable improvement over other approaches. However, the values for Pearson correlation are low and can be associated to the fact that Tweets and news are inherently very different and achieving high level of pairwise similarity is a complex task. But for accuracy assessment, which is mostly seen from the perspective of a classification task there is clear improvement over other approaches by using transformed embeddings as features. Table 4 shows that T2N achieved better performance as compared to N2T.

Although aforementioned analysis is perceived on a small dataset. The results show a promising direction to use Monolingual-TWE which can easily scale with the size of common vocabulary across collections. Thus giving a possibility to improve or sustain the accuracy and Pearson correlation values on larger datasets.

Similar observations can be enunciated about cross-lingual-TWE. Given the complexity associated with finding pairwise relevance between tweets and cross-language news, we compared only LDA based approaches with cross-lingual-TWE. It can be comprehended from Table 5 that T2N outperformed LDA-PTM with notable improvement. Although it may not be significant, these results only show preliminary examination to perceive research in this direction.

## 8 Conclusion and Future Work

In this paper, we focused on mapping tweets with monolingual and cross-lingual news by transforming their word embeddings closer to each other, thus bridging the lexical and word usage gap across collections. In future, we aim to improve the quality of results with more sophisticated approaches.

## References


1. Kwak, H., Lee, C., Park, H., Moon., S.: What is twitter, a social network or a news media?. In: Proceedings of WWW., ACM (2010) 591–600
2. Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.P., Yan, H., Li., X.: Comparing twitter and traditional media using topic models. In: Advances in Information Retrieval., Springer Berlin Heidelberg (2011) 338–349
3. Guo, W., Li, H., Ji, H., Diab., M.T.: Linking tweets to news: A framework to enrich short text data in social media. In: Proceddings of ACL. (2013) 239–249
4. Krestel, R., Werkmeister, T., Wiradarma, T.P., Kasneci., G.: Tweet-recommender: Finding relevant tweets for news articles. In: Proceedings of WWW, ACM (2015) 53–54
5. Pennington, J., Socher, R., Manning., C.D.: Glove: Global vectors for word representation. In: Proceedings of EMNLP. (2014) 1532–1543
6. Gouws, S., Bengio, Y., Corrado., G.: Bilbowa: Fast bilingual distributed representations without word alignments. In: arXiv preprint arXiv:1410.2455. (2014)
7. Wang, C., Mahadevan., S.: Manifold alignment using procrustes analysis. In: Proceedings of ICML, ACM (2008) 1120–1127
8. Tan, L., Zhang, H., Clarke, C.L., Smucker., M.D.: Lexical comparison between wikipedia and twitter corpora by using word embeddings. In: Proceedings of ACL. (2015)
9. Ritter, A., Etzioni, O., Clark., S.: Open domain event extraction from twitter. In: Proceedings of KDD. (2012) 1104–1112
10. Thelwall, M., Buckley, K., Paltoglou., G.: Sentiment in twitter events. Journal of the American Society for Information Science and Technology. **62** (2011) 406–418
11. Paul, M.J., Dredze., M.: You are what you tweet: Analyzing twitter for public health. In: Proceedings of ICWSM. (2011) 265–272
12. Ratkiewicz, J., Conover, M., Meiss, M., Goncalves, B., Flammini, A., Menczer., F.: Detecting and tracking political abuse in social media. In: Proceedings of ICWSM. (2011)

13. Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski., J.: #earthquake: Twitter as a distributed sensor system. Transactions in GIS. **17(1)** (2013) 124–147
14. Abel, F., Gao, Q., Houben, G.J., Tao., K.: Analyzing user modeling on twitter for personalized news recommendations. In: Proceddings of UMAP. (2011) 1–12
15. Ou, J., Liu, N.N., Zhao, K., Yu, Y., Yang., Q.: Transferring topical knowledge from auxiliary long texts for short text clustering. In: Proceddings of CIKM., ACM (2011) 775–784
16. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa., P.: Natural language processing (almost) from scratch. The Journal of Machine Learning Research **12** (2011) 2493–2537
17. Al-Rfou, R., Bryan, P., Steven., S.: Polyglot: Distributed word representations for multilingual nlp. In: Proceedings of CoNLL, ACL (2013) 183–192
18. Ramon, A.F., Amir, S., Lin, W., Silva, M., Trancoso., I.: Learning word representations from scarce and noisy data with embedding sub-spaces. In: Proceedings of ACL. (2015)
19. Kim, J., Rousseau, F., Vazirgiannis., M.: Convolutional sentence kernel from word embeddings for short text categorization. In: Proceedings of EMNLP. (2015)
20. der Maaten, L.V., Hinton., G.: Visualizing data using t-sne. The Journal of Machine Learning Research **9** (2008) 2579–2605
21. Webber, W., Moffat, A., Zobel., J.: A similarity measure for indefinite rankings. ACM Transactions on Information Systems (TOIS). **4** (2010)
22. Schönemann, P.H.: A generalized solution of the orthogonal procrustes problem. Psychometrika. **31(1)** (1966) 1–10
23. Mimno, D., Wallach, H.M., Naradowsky, J., Smith, D.A., McCallum., A.: Polylingual topic models. In: Proceedings of EMNLP, ACL (2009) 880–889