Optimizing Topic Distributions of Descriptions for Image Description Translation

Jian Tang Yu Hong[*] Mengyi Liu Jiashuo Zhang Jianmin Yao
Soochow University, Suzhou, Jiangsu, China
{johnnytang1120, tianxianer, mengyiliu22, jiasurezhang}@gmail.com, jyao@suda.edu.cn

Abstract. Image Description Translation (IDT) is a task to automatically translate the image captions (i.e., image descriptions) into the target language. Current statistical machine translation (SMT) cannot perform as well as usual in this task because there is lack of topic information provided for translation model generation. In this paper, we focus on acquiring the possible contexts of the captions so as to generate topic models with rich and reliable information. The image matching technique is utilized in acquiring the relevant Wikipedia texts to the captions, including the captions of similar Wikipedia images, the full articles that involve the images and the paragraphs that semantically correspond to the images. On the basis, we go further to approach topic modelling using the obtained contexts. Our experimental results show that the obtained topic information enhances the SMT of image caption, yielding a performance gain of no less than 1% BLUE score.

1 Introduction

The caption is defined as a short text which specially describes the image it attached with. Caption translation[6] is admittedly very practical for many current industrial applications, such as e-commerce, social media and computational advertising. By the rapid and effective caption translation (of course in a fashion of automatic processing), users can easily understand the meanings, intentions and subjects of the images posted (or released) in the foreign websites, medium or literatures.

For example, the image in Fig 1 shows different types of "*Bacalhau*" (in spanish *Bacalhau* means Gadus, a kind of fish living in Atlantic), and the inner captions in the photo detail the difference, including special long-island Gadus, pubertal island Gadus, etc. It is very difficult for a Chinese customer rapidly get deep insight into the difference among the piles of fishes, no matter from the perspective of the objects or the languages in the captions. In this case, theoretically speaking, caption translation can help the user figure out the difference and make a correct decision during the purchase process.

The closely related study to the caption translation is IDT[11], which is newly proposed in the inter-discipline of image processing and language processing. It follows a strict task definition, where the translator is required to translate a short caption into the target language, under the condition that there isn't any available reference data (contexts, topics and articles) except the corresponding image. In this paper, we limit our research to the IDT schema, including the task definition, corpus and evaluation metric.

[*] Corresponding author

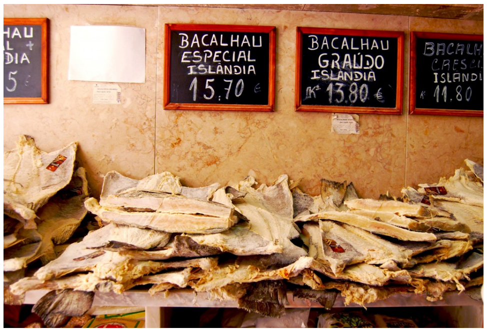

Fig. 1. An image and its corresponding captions

Topic information[25, 27, 14] is admittedly very important for SMT. The topics of a text span (clause or sentence) and the ones of its contexts are constructive to speculating about the correct translations of words, phrases or even the whole text span. For example, the English word "*mouse*" may mean a "*rat*" or a "*cursor controller*", and if provided with the topic about "*wild kingdom*", the translator could correctly translates it as a word of the meaning "*rat*" in the target language.

However, in IDT, there is absence of available contexts for detecting the related topics, and more seriously the caption is generally short and therefore involves fewer informative words for topic modelling, such as the caption "*white mouse, gray mouse and yellow mouse*". Thus when we directly employ an existing SMT system, the commonly-used linguistic knowledge and translation knowledge are actually task-independent. This will unavoidably reduce the translation quality.

In order to overcome the problem, we propose to use the image as a query to retrieve the related contexts to the captions[7]. And then, we use the contexts to reinforce topic modelling in the process of learning language and translation. In practice, we employ a deep convolutional neural network based image processing techniques to implement image-image matching, and further collect wikipedia web pages that contain the similar images. By using the articles in the web pages as the contexts, we re-build the topic model of the target caption and further re-train the language and translation models. In our experiments, our method improves the traditional SMT system[26] with a performance gain of no less than 1% BLUE score.

The rest of the paper is organized as below. In section 2, we detail our approach. Section 3 shows the experimental settings, results and analysis. In section 4, we briefly overview the related work. We draw our conclusion in section 5.

2 Methodology

In order to overcome the shortcomings of translating ambiguous words in captions, we acquire informative and related contexts of the captions from Wikipedia webpages. On the basis, we leverage the contexts to topic modelling for the captions. As what we will show in the upcoming sections, the key issue is to precisely extract the related contexts from the webpages. In this section, we first present the methodological framework and then detail the components respectively.

2.1 Framework

Our approach consists of four primary components, including 1) image search based relevant document acquisition, 2) related context extraction, 3) topic re-modelling and 4) SMT using informative topic models.

We build an image search engine to acquire the Wikipedia webpages which contain similar images to the image of the target caption. The so-called target caption is specified as the caption in the test data, i.e., the one we aim to translate into the target language. In this phase, we employ the dump file of Wikipedia to build a local multimodal database. The database regards the images as indexes. It also preserves the correspondence between the images and their captions and the documents where they occur. In section 2.2, we focus on the deep CNN based image-image matching, which is the most crucial step for acquiring the relevant documents.

When using the image of a target caption as the query to perform information retrieval, we obtain the most relevant document (webpage) to the caption in the database. Such a document is recalled because it contains an image which is visually-similar to the image of the target caption. In order to facilitate the reading of this paper, we simply name the visually-similar image as the mirror image (`Mir`) in Wikipedia webpage, and the document which holds the `Mir` as the container (`Cnt`).

The (`Cnt`) additionally contains a article and the caption of the `Mir`. This caption (`Mir`'s caption) can be jointly used with the target caption to provide richer information for topic modelling. But in this paper, we focus on the utilization of the article. Simply, the full text of the article can be directly used as the related context to the target caption. It is because the text is closely related to the `Mir`. But we argue that there are also many unrelated information in the text, which probably mislead the topic modelling of the target caption. For example, in a Wikipedia webpage about Gadus, there are many blocks corresponding to other name entities and images, such as Department of Fisheries, policies, persons, events and even a book titled with "*The Old Man and the Sea*". Therefore, we come up with the extraction of closely-related contexts from the article. In section 2.3, we will detail the context extraction approach, and in our experiments (section 3), we compare the effects of the caption of the `Mir`, full text of the article and the extracted contexts on the topic remodelling based caption translation.

Using the contexts extracted from the relevant articles, we remodel the topic of the target caption and put it into use for enhancing the translation model (Section 2.4).

2.2 Image Search

We regard the image of the target caption as a query, and search the relevant document (`Cnt`) in the multimodal database by image-image matching.

We encode the images as image embeddings and use Euclidean distance between embeddings to measure visual similarity. Each image embedding is consisted of a series of image block embeddings. An image block serves as the elementary visual semantics,

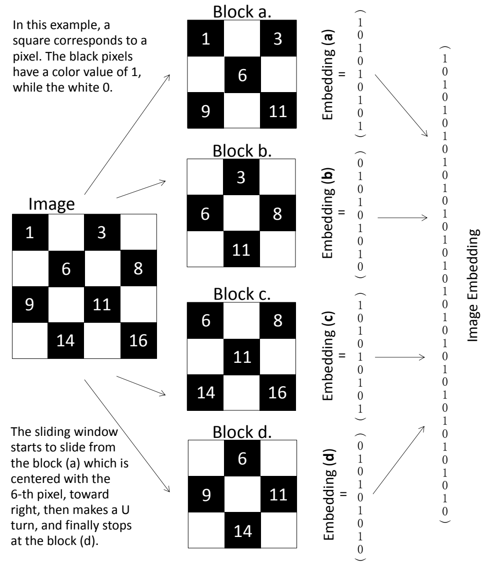
In this example, a square corresponds to a pixel. The black pixels have a color value of 1, while the white 0.
Image
1 3 6 8 9 11 14 16
Block a.
1 3 6 9 11
Block b.
3 6 8 11
Block c.
6 8 11 14 16
Block d.
6 9 11 14
The sliding window starts to slide from the block (a) which is centered with the 6-th pixel, toward right, then makes a U turn, and finally stops at the block (d).
Embedding (a) = (1 0 1 0 1 0 1 0 1)
Embedding (b) = (0 1 0 1 0 1 0 1 0)
Embedding (c) = (1 0 1 0 1 0 1 0 1)
Embedding (d) = (0 1 0 1 0 1 0 1 0)
Image Embedding
(1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0)

Fig. 2. Image embedding generation

as usual, occupying a region of 3 × 3 pixels (see Fig. 2.). To some extent, it plays a role of basic semantic unit just like that of word in linguistics. The block embedding is specified as a vector of real numbers. Each number corresponds to the color value of a pixel. When generating the whole embedding of the image, we first collect all possible blocks by successively moving a sliding window in the size of 3×3 pixels with a margin of 1 pixel, and then concatenate all the block embeddings to compose a uniform vector representation (see Fig. 2.).

Features of images in all datasets were extracted using the 16-layer version of VGGNet (VGG-16)[1][23], which is a Deep Convolutional Neural Network (CNN) pre-trained on 1,000 object categories of the classification/localization task of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [20]. We used Caffe to extract the fully-connected layer (FC7) of VGG-16 before applying the softmax function. The 4096 dimensional vectors from this layer provide the content of images.

2.3 Context Extraction

We select sentences from the article `Cnt` to form the contexts of the target caption. The sentences are constrained to be relevant to the caption of the mirror image (`Mir`). Different from the traditional bag-of-word based approach, we introduce sentence-level embeddings into the solution.

By tokenization, we transform a candidate sentence into a series of words. We follow [12] to represent a word with an embedding. Word embedding is a vector of real numbers [16] which is trained to be meaningful using massive semantic contexts. It is

[1] https://github.com/ry/tensorflow-vgg16

capable of representing an unique word sense and the local semantics of the surroundings. We generate the word embedding for each word in the candidate sentence, and concatenate all the embeddings to form a simple sentence-level embedding. Further, an average operation based convolution computing is used to generate a fixed-size sentence embedding.

In practice, we compute embeddings directly by word2vec[2] [11], which is pre-trained over the English articles in the latest 2015 wikipedia dump. We follow [4] to preprocess the articles, and perform pre-training using the skip-gram architecture [16]. We use a 100 dimensional vector as the word embedding. Euclidean distance is used to measure the similarity between the candidate sentences and the *Mir*'s caption. The candidates which have a similarity no less than the threshold 0.35 will be eventually selected as the relevant contexts.

The semantics based similarity computing generally causes the selection of the over-similar sentences from the *Cnt*. In our case, the contexts are expected to entail more different but related information. In order to meet this requirement, we expand the contexts by involving the sentences around the semantically-similar sentences. We set a sliding window to control the range of expansion. The maximum margins is set as the head and the tail of the paragraph which contains the sentences. In our experiments, we evaluate the effect of different sizes of contexts on the caption translation.

2.4 Topic Remodelling Based Translation Model Reinforcement

The purpose of this part is to integrate the topic distributions into the phrase table in the form of features, which aims at optimizing the translation model. Figure 3 depicts the basic structure of SMT system. The SMT system consists of three modules: translation model, language model and reordering model. We train translation model and reordering model by parrallel training corpus, but just use monolingual corpus of target language to train language model.

Euclidean distance between source image and images in multimodal database will be sorted from small to large, and we select Top10 images as similar images. The average vector of topic distributions of pseudo-documents is used to infer topic distribution of source caption. We use Latent Dirichlet Allocation (LDA)[3] [1] to learn topic distributions of these pseudo-documents and set the dimension of topic vectors to 100. The formula for acquiring the topic distribution is shown as follows:

$$P(T) = \frac{\sum\limits_{i=1}^{10} P_i(T)}{10} \tag{1}$$

$P_i(T)$ denotes the probability that document of i -th similar images belongs to topic T.

We use features extracted from topic distributions to optimize translation model. Features contain positive translation probability, negative translation probability and the topic sensitivity of the phrase pair.

[2] https://github.com/jhlau/doc2vec

[3] https://github.com/mrquincle/gibbs-lda

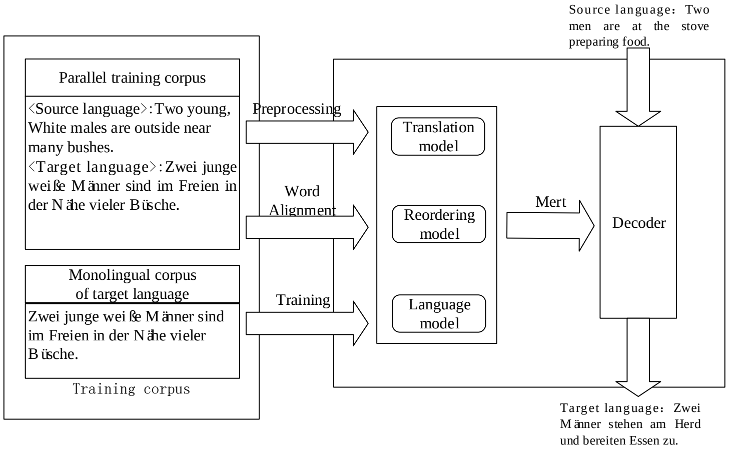
Source language：Two men are at the stove preparing food.
Parallel training corpus
<Source language>:Two young, White males are outside near many bushes.
<Target language>:Zwei junge weiße Männer sind im Freien in der Nähe vieler Büsche.
Monolingual corpus of target language
Zwei junge weiße Männer sind im Freien in der Nähe vieler Büsche.
Training corpus
Preprocessing
Word Alignment
Training
Translation model
Reordering model
Language model
Mert
Decoder
Target language：Zwei Männer stehen am Herd und bereiten Essen zu.

Fig. 3. Structure of SMT system

In the training corpus, *N* denotes the set of sentence pairs which contain phrase pairs (s,t). *M* denotes the set of sentence pairs which contain source phrase *s*. *K* denotes the set of sentence pairs which contains target phrase *t*. The formulas of positive translation probability are drawn as follows:

$$P(t|s,T)=\frac{\sum\limits_{i=1}^{n} Count_i(s,t)P_i(T)}{\sum\limits_{j=1}^{m}(\sum_{t'} Count_j(s,t'))P_j(T)} \tag{2}$$

$$P(t|s)=\sum_{i=1}^{100} P(t|s,T_i)P(T_i) \tag{3}$$

The formulas of negative translation probability are shown as follows:

$$P(s|t,T)=\frac{\sum\limits_{i=1}^{n} Count_i(s,t)P_i(T)}{\sum\limits_{j=1}^{k}(\sum_{s}' Count_j(s',t))P_j(T)} \tag{4}$$

$$P(s|t)=\sum_{i=1}^{100} P(s|t,T_i)P(T_i) \tag{5}$$

The sensitivity of phrase pairs is expressed in terms of cross-entropy and the formulas are constructed as follows:

$$C(T_j) = \sum_{i=1}^{n} Count_i(s,t) P_i(T_j) \tag{6}$$

$$Entropy(s,t) = -\sum_{i=1}^{100} C_i log(C_i) \tag{7}$$

where *s* denotes source phrase and *t* denotes target phrase. $P(t|s,T)$ denotes the probability of translating *s* to *t* in topic *T*. $Count_i(s,t)$ denotes the number of phrase pair (s,t) contained in the *i*-th sentence.

3 Experiments

3.1 Corpus

The data of local multimodal database is from Wikipedia. Training, Development and Test sets are from Multi30k dataset [5] which is an extension of the Flickr30k dataset [19]. The datasets of classification task in this paper are from MS COCO 2014 [15] which contains about 80,000 images with 5 English captions per image. Table 1 shows the distribution of experimental data:

Table 1. Experimental corpus

DATA	SOURCE	SCALE
Multimodal database	Wikipedia	About 300K
Training set	Multi30k	29K
Development set	Multi30k	1,014
Test set	Multi30k	1,000
Classification images	MS COCO	About 80K
Classification captions	MS COCO	About 400K

3.2 Settings

Our submissions are based on the NiuTrans SMT toolkit [26] to build phrase-based SMT models. They are constructed as follows: First, word alignments in both directions are calculated using GIZA++ [18]. The phrases and reordering tables are extracted using the default settings of the NiuTrans toolkit. The parameters of NiuTrans were tuned on the provided development set, using the MERT [17] algorithm. 3-gram back-off language were built using SRILM [24] and the target side of the parallel corpus. Training was performed using only the data provided by the task organizers, and so systems were built in the constrained setting. The results are evaluated by BLEU-4 value which is case-insensitive. In order to train the translation model incorporating topic information, this paper uses Gibbs sampling method to infer the parameters of LDA model, and uses GibbsLDA++ open source tool to infer the topic distributions. The topic model tool used in this paper uses the following parameters: the number of topics is 100, the super parameters are set to 0.05, and the number of iterations is set to 1000. The image feature is extracted by VGG-16.

In this paper, we conduct nine experiments based on the viewpoint that accurate topic distribution is good for the task of IDT. For Baseline, we conduct a phrase-based translation system that includes translation model, language model and reordering model. Other experiments contain different ways to acquire topic distributions of captions. We acquire topic distributions of source captions from three respects.

Captions from different sources are used to extract topic distribution of captions attached to source images by GPUDMM[13]. Experiments in Table 2 show different sources of captions.

Table 2. Acquiring topic distributions from captions

Methods	Source of Caption
DES_GDMM	Captions attached to source images
DES_GDMM_LIB	Captions attached to images in multimodal database

We use different matching methods to acquire articles. These articles are applied to extract topic distributions of captions attached to source images by LDA. Experiments in Table 3 show different matching methods.

Table 3. Different retrieval methods for topic distribution

Methods	Matching Methods
IR_ALLTEXT	image retrieval
TR_ALLTEXT	text retrieval

Related sentences are selected using different approaches. Pseudo-documents consisting of relevant sentences are applied to extract topic distributions of captions attached to source images by LDA. We use D2V to indicate extracting text features by Doc2vec. In contrast, SKIPS means extracting text features by Skip-Thoughts[4]. RF means using RandomForest to perform classification task. T-CNN[5] is a CNN model for text classification[10]. Experiments in Table 4 show different approaches for selecting relevant sentences.

[4] https://github.com/tensorflow/models/tree/master/skip_thoughts

[5] https://github.com/dennybritz/cnn-text-classification-tf

Table 4. Related sentences for Topic Distribution

Methods	Approaches
DES_D2V	Selecting sentences whose value of similarity to caption below 0.35.
DES_SKIPS	This method is different from DES_W2V in that text features are extracted by Skip-Thoughts.
RF_ITC	Using RF to Perform binary classification on text and image.
T-CNN_ITC_SKIPS	Using softmax to Perform binary classification on text and image.

3.3 Results and Analysis

Table 5. BLEU Results of Experiments

Method	BLEU(%)
Baseline	32.45
DES_GDMM	31.17
DES_GDMM_LIB	33.09
TR_ALLTEXT	32.65
IR_ALLTEXT	32.9
DES_D2V	33.45
DES_SKIP	33.13
RF_ITC	31.8
T-CNN_ITC_SKIPS	33.01
TSR-TXT	29.7
TSR-CNN	30.6
TSR-HCA	30.3

TSR-TXT, TSR-CNN and TSR-HCA are three approaches presented in [6]. The key idea is to perform image retrieval over database of images that are captioned in the target language, and use the captions of the most similar images for crosslingual reranking of translation outputs.

Table 5 show that the experiment DES_D2V achieves the best performance. For Baseline, TR_ALLTEXT achieves improvement of 0.25 BLEU point owing to introducing topic information into translation model. IR_ALLTEXT shows improvement of 0.2 BLEU point over TR_ALLTEXT due to rich information of images. The performance of DES_GDMM is poorer than Baseline lying in inaccurate topic information from short texts. DES_D2V improves 0.55 BLEU point and 1 BLEU point over IR_ALLTEXT and

Baseline resepectively. One explanation can clarify this behavior, sentences selected by text match are not only closely related to images, but also reduce interference information in whole document.

4 Related Work

In the previous year's competition [2], most of the systems were based on the phrase-base SMT in a monolingual setting [22]. [21] learned images information to rerank the results of translation system. [6] presents an approach to improve SMT of image captions by multimodal pivots defined in visual space. The key idea is to perform image retrieval over a database of images that are captioned in the target language, and use the captions of the most similar images for crosslingual reranking of translation outputs.

As the advances of deep learning, Neural Machine Translation (NMT) [9, 8] attracts research attention. [3] present a double-attentive multimodal machine translation model which learns to attend to source language and visual features as separate attention mechanisms.

Our work departs from the previous work based on SMT lying in incorporating images information into translation model.

5 Conclusion and Future Work

This paper presents an approach to effectively infer the topic distributions of source captions. By integrating the topic distributions into phrase table, it is proved that topic distribution is beneficial to improve the performance of IDT. The method in this paper selects some sentences from documents to form pseudo-documents, which not only guarantees the richness and diversity of text information, but also ensures text information is similar to captions of source images as far as possible.

From the perspective that the quality of topic distribution influences the performance of IDT. In the future work, we will try from the following two respects: First, extracting part of images associated with captions by neural networks to assist IDT; Second, extracting some sentences closely related to images to learn precise topic distributions.

Acknowledgement

This research work is supported by National Natural Science Foundation of China (Grants No.61373097, No.61672367, No.61672368, No.61331011, No.61773276), the Research Foundation of the Ministry of Education and China Mobile, MCM20150602 and the Science and Technology Plan of Jiangsu, SBK2015022101 and BK20151222. The authors would like to thank the anonymous reviewers for their insightful comments and suggestions. Yu Hong, Professor Associate in Soochow University, is the corresponding author of the paper, whose email address is tianxianer@gmail.com.

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of machine Learning research 3(Jan), 993–1022 (2003)
2. Bojar, O., Buck, C., Federmann, C., Haddow, B., Koehn, P., Leveling, J., Monz, C., Pecina, P., Post, M., Saint-Amand, H., et al.: Findings of the 2014 workshop on statistical machine translation. In: WMT@ ACL. pp. 12–58 (2014)
3. Calixto, I., Elliott, D., Frank, S.: Dcu-uva multimodal mt system report. In: WMT. pp. 634–638 (2016)
4. Dos Santos, C.N., Gatti, M.: Deep convolutional neural networks for sentiment analysis of short texts. In: COLING. pp. 69–78 (2014)
5. Elliott, D., Frank, S., Sima'an, K., Specia, L.: Multi30k: Multilingual english-german image descriptions. arXiv preprint arXiv:1605.00459 (2016)
6. Hitschler, J., Schamoni, S., Riezler, S.: Multimodal pivots for image caption translation. arXiv preprint arXiv:1601.03916 (2016)
7. Hong, Y., Yao, L., Liu, M., Zhang, T., Zhou, W., Yao, J., Ji, H.: Image-image search for comparable corpora construction. In: The 26th International Conference on Computational Linguistics (COLING 2016). p. 16 (2016)
8. Jean, S., Cho, K., Memisevic, R., Bengio, Y.: On using very large target vocabulary for neural machine translation. arXiv preprint arXiv:1412.2007 (2014)
9. Kalchbrenner, N., Blunsom, P.: Recurrent continuous translation models. In: EMNLP. vol. 3, p. 413 (2013)
10. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
11. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on Machine Learning (ICML-14). pp. 1188–1196 (2014)
12. Levy, O., Goldberg, Y.: Dependency-based word embeddings. In: ACL (2). pp. 302–308. Citeseer (2014)
13. Li, C., Wang, H., Zhang, Z., Sun, A., Ma, Z.: Topic modeling for short texts with auxiliary word embeddings. In: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. pp. 165–174. ACM (2016)
14. Li, S., Chua, T.S., Zhu, J., Miao, C.: Generative topic embedding: a continuous representation of documents. In: ACL (1) (2016)
15. Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollar, P.: Microsoft coco: Common objects in context. arXiv preprint arXiv:1405.0312 (2014)
16. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. pp. 3111–3119 (2013)
17. Och, F.J.: Minimum error rate training in statistical machine translation. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1. pp. 160–167. Association for Computational Linguistics (2003)
18. Och, F.J., Ney, H.: Improved statistical alignment models. In: Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. pp. 440–447. Association for Computational Linguistics (2000)
19. Plummer, B.A., Wang, L., Cervantes, C.M., Caicedo, J.C., Hockenmaier, J., Lazebnik, S.: Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models. In: Proceedings of the IEEE international conference on computer vision. pp. 2641–2649 (2015)

20. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. arXiv preprint arXiv:1409.0575 (2014)
21. Shah, K., Wang, J., Specia, L.: Shef-multimodal: Grounding machine translation on images. In: Proceedings of the First Conference on Machine Translation. vol. 2, pp. 660–665. ACL (2016)
22. Simard, M., Ueffing, N., Isabelle, P., Kuhn, R.: Rule-based translation with statistical phrase-based post-editing. In: Proceedings of the Second Workshop on Statistical Machine Translation. pp. 203–206. Association for Computational Linguistics (2007)
23. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
24. Stolcke, A., et al.: Srilm-an extensible language modeling toolkit. In: Interspeech. vol. 2002, p. 2002 (2002)
25. Su, J., Wu, H., Wang, H., Chen, Y., Shi, X., Dong, H., Liu, Q.: Translation model adaptation for statistical machine translation with monolingual topic information. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. pp. 459–468. Association for Computational Linguistics (2012)
26. Xiao, T., Zhu, J., Zhang, H., Li, Q.: Niutrans: an open source toolkit for phrase-based and syntax-based machine translation. In: Proceedings of the ACL 2012 System Demonstrations. pp. 19–24. Association for Computational Linguistics (2012)
27. Yang, W., Boyd-Graber, J.L., Resnik, P.: A discriminative topic model using document network structure. In: ACL (1) (2016)