Analysis on Transfer Learning Models and Applications in Natural Language Processing

Muzi Chen

Department of Statistics, University of California, Davis, California, 95616, U.S.

mzichen@ucdavis.edu

Abstract. Assumptions have been established that many machine learning algorithms expect the training data and the testing data to share the same feature space or distribution. Thus, transfer learning (TL) rises due to the tolerance of the different feature spaces and the distribution of data. It is an optimization to improve performance from task to task. This paper includes the basic knowledge of transfer learning and summarizes some relevant experimental results of popular applications using transfer learning in the natural language processing (NLP) field. The mathematical definition of TL is briefly mentioned. After that, basic knowledge including the different categories of TL, and the comparison between TL and traditional machine learning models is introduced. Then, some applications which mainly focus on question answering, cyberbullying detection, and sentiment analysis will be presented. Other applications will also be briefly introduced such as Named Entity Recognition (NER), Intent Classification, and Cross-Lingual Learning, etc. For each application, this study provides reference on transfer learning models for related researches.

Keywords: Transfer learning, cyberbully detection, question answering, sentiment analysis.

1. Introduction

In traditional machine learning setting, the unique models are required to perform the different tasks respectively, which infers that a mass of data are needed. Such construction of models and collection of data are usually costly in real-world applications. In contrast, transfer learning reuses the model trained on one dataset to perform other tasks. Among the tasks, there might be some related mechanisms that the model does not have to learn everything from scratch. After transferring, the trained model is expected to retain previously the learned knowledge to perform the different tasks [1]. In this situation, less data are in need of recollection [2] and thus the model achieve a more efficient performance. To introduce and to better understand the different types of TL, mathematical definition will be considered first.

2. Transfer Learning

2.1. Definition

The definition of TL is introduced in the study of Pan and Yang in detail. They described the notation of domain and task. They divided the domain and task further into source domain, target domain, source task, and target tasks. Task data are not observed but learned in training data. In the following categorization, all mathematical notations are explained in Pan and Yang's study.

2.2. Different Types of TL

If a model is forced to transfer without considering specific conditions or limitations, the transfer might be unsuccessful or even become counterproductive. There are different techniques to meet the requirement when transfer learning is applied. Choosing these techniques usually depends on the source and target domains and tasks.

2.2.1 Transductive TL

In the setting of transductive TL, the source and targettask are the same, while the domains are different. By considering the definition, $D = \{X, P(X)\}$, either $Xs \neq X_T$ or $Ps(X) \neq P_T(X)$ satisfies the context. Thus, two transductive TL can be further classified respectively as the following [3]:

● Cross-lingual Learning: The feature space between the source and task is different.For instance, the source is in English and the target is in Chinese. The situation often applied when one source language contains more resource than the other task language.

● Domain Adaption: The feature space is the same, while the marginal distribution of the source and task is different, which means $Ps(X) \neq P_T(X)$. For instance, in an NLP analysis, a review will be written on the performance of smartphones in the source domain while on the performance of television in the target domain.

2.2.2 Inductive TL

In the setting of inductive TL, the target and source tasksare different. Inductive TL can be further classified into two types, determined by the amount of labeled and unlabeled data.

● Multi-Tasks Learning: There are plenty of labeled data available in the source domain. Under such circumstance, training can be performed similarly to multitask learning.

● Sequential TL: There is no labeled datum in the source. In the circumstances, training is closer to self-taught learning. The process is slow compared to multi-task learning but feasible while not much data in the source domain are available.

2.3. Comparison of PretrainedModel and Other ML Models

Compared to the traditional machine learning models, transfer learning models have more advantages in memory preservation, data efficiency, and freedom to manipulate data [2, 4].

In the field of transfer learning, there exist techniques even more advanced than fine-tuning such as adapter modules. In the experiment of Houlsby's study, they invented the adapter modules which achieved within 0.4% of the execution compared to full fine-tuning by adding few parameters per task. It is worth noticing that fine-tuning itself can train 100% of the parameters per task. They also compared the adapter-based tuning with multi-task and continual learning. The adapter neither required simultaneous access to all tasks as multi-task learning nor loses memory of previous tasks as life-long continual learning. Thus, they achieved data efficiency and memory preservation in the study.

Although there are various studies on how to manage unlabeled data, these experiments require that the distribution of the data be the same. Transfer learning does not assume data to have the same domains, tasks, or distributions, which renders models the freedom to select data.

3. Different Application Fields

Transfer learning has been used to capture the nuances in various text forms.Applications in the natural language processing field take advantage of transfer learning to recognize opinions, sentiments, evaluations, appraisals, attitudes, and emotions. In this section, applications that are more general will be introduced. Challenges in some areas may be addressed as well.

3.1. Question Answering

Question answering becomes prevalent in recent years that it can save time and cost from receiving responses from human experts. It refers to the technique of how well a model can understand a story as humans do [5]. A story can be defined as a sequence of sentences or simply a phrase. It can also refer to system equations when the model is applied to algebra word problems. Listening test is also included as mentioning in the following experiment.

In the experiment of Chung's study, they improved models on both the TOEFL test and the MCTest by a TL technique from MovieQA.

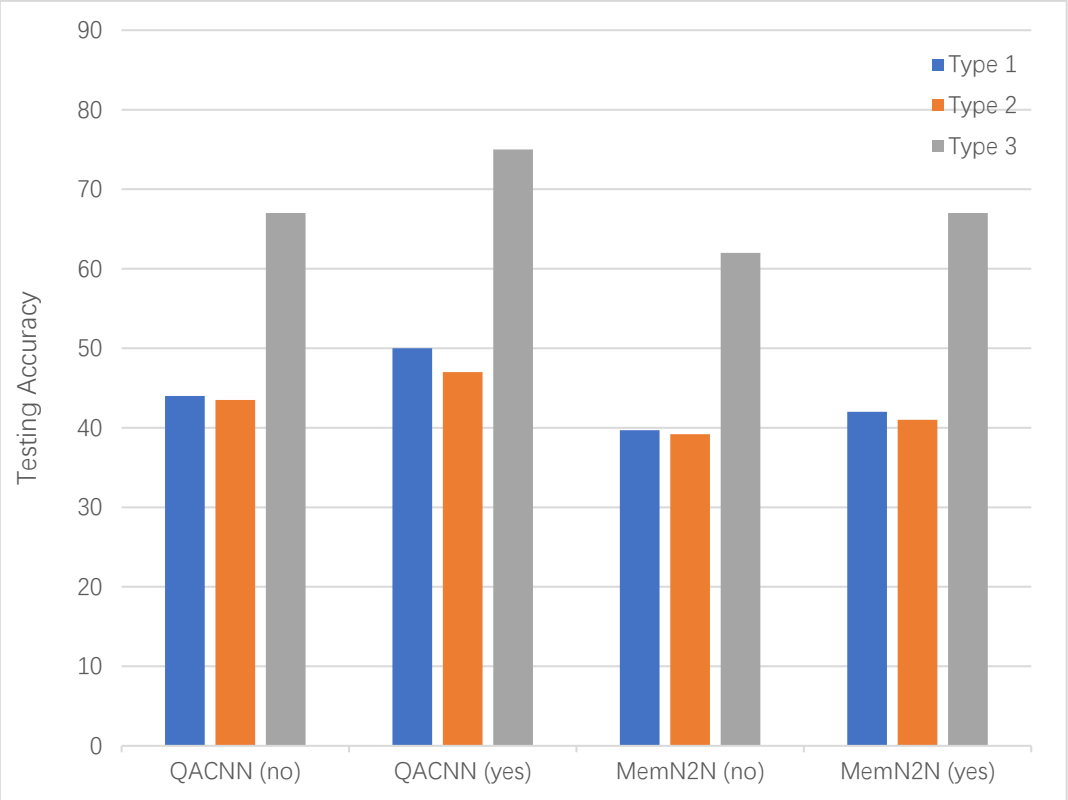

Figure 1. The performance of models on multiple choice question-answering on different types of questions in TOEFL (with and without pre-training). 'No' represents non-pretrained model, while 'Yes' representsthe pre-traineds [5]

One of the results shown by Chung indicates the different types of questions that benefit from transfer learning. Types 1, 2, and 3 represent the basic comprehension of the story, attitude, and utterance due to speakers'expression, and generalizations of stories respectively. The level of the questions upgrades stepwisely. Figure 1 shows that the performance on all three types of questions promotes after pretraining. They also showed that the transfer learning technique functions well in unsupervised QA experiments when target data are unavailable.

In another similar experiment conducted by Min [6], they obtained the same result: by applying transfer learning, their models improved greatly using two different sets of data set: WikiQA and SemEval-2016. Their models outperformed the previous best one by a significant amount. The result also shows that such question-answering tasks with sentence-level supervision benefit from transfer learning through span-level supervision.

3.2. Cyberbullying Detection

Cyberbullying detection mainly considers problems including flaming, harassment, denigration, impersonation, outing, boycotts, and cyberstalking. Such damage may cause mental health issues or affect the personality of a victim [7]. The challenge of cyberbullying focuses on recognizing the whole story and victims especially when there are variations in language on social media. It is sometimes hard to judge whether a statement is classified as cyberbully without sentiment analysis as well. Utilizing transfer learning, a model proposed by Uban [8] successfully analyzed aggressive and offensive language integrating sentiment. They made a comparison of predicted values from pretrained model and non-pretrained model. The result is shown in Table 4 in their original paper that the model produced correct classification after pretraining on some previously misclassified tweets. They also found similar features exist between cyberbullying and negative sentiment. What is more, they had shown that classifying aggressive language was easier than detecting offensive language.

Other than detecting some well-recognized cyberbullying from above, sometimes cyberbullying can be a subjective decision to a person whether a message is hurtful or not. Such ambiguity creates

more challenges for cyberbullying detection. As reported by Roy & Mali, some studies emphasized textual posts on this issue were found, but less effort was made by researchers on image-based cyberbullying. In their research, they provided a solution for image-based detection that they achieved the best accuracy of 89% using transfer learning among other methods. Three models were built to conduct the TL through different amount of epochs. A comprehensive result can be found in Figure 5 in the original paper while only plots with highest number of samples and most reasonable distributions are shown here.

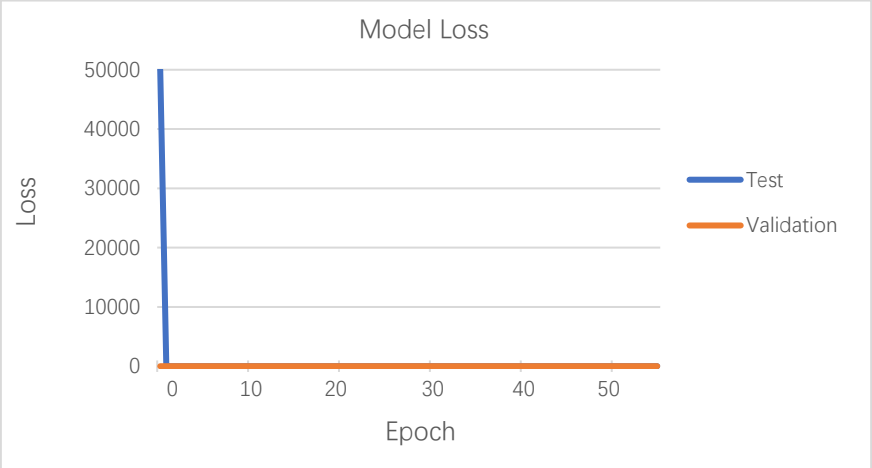

Figure 2. The loss Vs epoch observation of 2DCNN model [7]

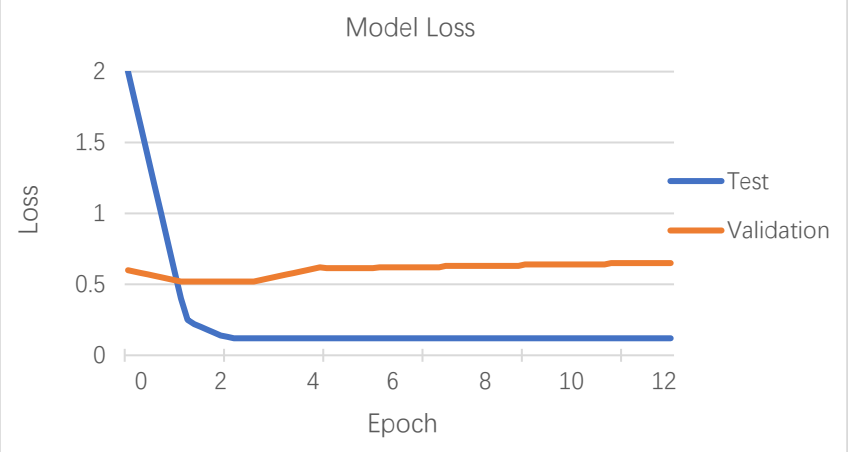

Figure 3. The loss Vs epoch observation of VGG16 model [7]

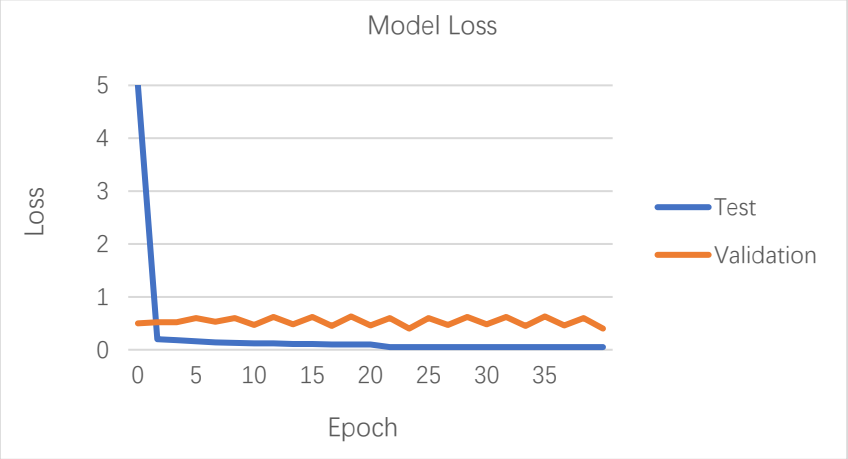

Figure 4. The loss versus epoch of Inception V3 model [7]

In Figure 2 to Figure 4, each one represent the lossversus epoch with3000 samples that was split in ratio 90:10 for training and test. It indicates that the loss is comparatively lower for VGG16 and Inception V3 models. As a result, these two models perform better while a large number of data is required.

3.3. Sentiment Analysis

Sentiment analysis has become increasingly popular these year. It is common to consider that people's behaviors are usually opinion-based which involves judgment and evaluation from others. Such a trend becomes more and more influential since gaining comments and evaluations on the web are more feasible these days. Sentiment analysis has a similar challenge as cyberbullying detection. It also requires understanding the context of long posts and articles [9].

In Zhang et al.'s research, they introduced how sentiment analysis is usually distributed by level: document, sentence, and aspect. At the document level, the message is simply put into general representation as it expresses positive or negative feelings. At the sentence level, the fundamental unit to be considered is one sentence in the document. Normally, sentence level analysis can be applied to the classification problem as judging a sentence to be negative, neutral, or positive. Many research examples of applying this technique can be found on social media. For example, researchers might want to find out if a person is under depression by analyzing his post. Aspect level analysis becomes more detailed on opinions extracting and summarizing. It is not only judging generally positive or negative feelings towards the whole story but separately of how a person considers each different aspect of the object. Specifically, both positive and negative comments can be given to a product so more fine-grained studies may need to be conducted.

Some specific transfer learning methods are applied when researchers conduct sentiment analysis. There are mainly three methods which are introduced below.

- Parameter-transfer Methods: The method involves the parameters of the source and target domain where model parameters in a large dataset can be transferred from the pretrained model to the target task.
- Instance-transfer Methods: Here, this method considers sharing the data of both source and target. The data are selected from the source domain by re-weighting first. Then, the target data is increased with the labeled source samples; however, the augmentation might cause a negative transfer which leads to misclassification.
- Feature-representation-transfer Methods: This method applies when the source and target share part of crossover characteristics. The data are transformed from the two domains into the same feature space through feature transformation and performed by the conventional ML method.

Another part of sentiment analysis focuses on negative sentiment. Same as cyberbullying, negative sentiment might also affect people's mental health. Research on negative sentiment analysis on social media had been proposed by Wang et al [10]. They found four main concerns of Covid-19 from people on a social media in China, which were the virus Origin, Symptoms, Production Activity, and Public Health Control. They found the keywords which led to the depression of people. The keywords can be found in Table 4 in the research which includes a list of the ordered keywords that cause negative sentiment. It was found that the words "Gamey Food"and "Bat", etc. occupied the most origin of negative sentiment. With constructive instruction on public health responses, this sharing information may help the public to relieve anxiety caused by Covid-19.

3.4. Other Applications

In this section, some other applications of transfer learning in the NLP field arebriefly summarized in Table 1. The table includes the usage of the different applications and what effect of transfer learning will reflect on each of them.

Table 1. Other applications in transfer learning field

Name	Usage
Named Entity Recognition (NER)	NER is used to extract fundamental entities (e.g. person, location, organization, etc.) from text and answer questions about where the entities are mentioned. It can be applied to many fields including sentiment analysis discussed above, semantic studies, machine translation, etc [11]. In Lee et al.'s research, they addressed the problem that labeled target data were usually hard to obtain in NER study. By applying TL techniques on artificial neural networks (ANNs), the models provided better fine-grained performance. Such transferring might be advantageous for the target data with fewer labels [12].
Intent Classification (IC)	Intent classification plays a major role in NLP. For instance, to better satisfy the specific needs of customers, it is more efficient to classify their intent first [13]. Researches focus mostly on utterances from each message. It also involves question answering that a system will try to figure out why the customers contact the organizations or what kind of help they need. By convention, intent classification is usually performed with proper classifiers such as bag-of-word or continuous bag-of-word. A research has been established by Kumar et al. that they successfully adapted the low availability of target data using TL techniques [14]. By increasing variability in training, they improved the performance of classification. They also reduced the bias through data augmentation.
Cross-Lingual Learning	Cross-Lingual Learning is used to understand opinions in different languages, which is partly mentioned in section 3.2. Application such as machine translation often occupies a decisive position in the field. A study has been done that TL techniques leverage large datasets available in one language to build multilingual models which can transfer to other languages [15]. It is worth noticing that translating some parts of an article independently, such as its premise and hypothesis, can reduce the lexical overlap by the model which improves performance.

4. Conclusion

Transfer learning allows reusing the existing labeled data among the related tasks and domains which is more efficient. The goal is to transfer as much information as feasible from the source to the target. In this paper, definition of transfer learning has been drawn attention, accompanied by the various types of transfer learning algorithms. Also, it is not as expensive as traditional machine learning algorithms. It attains memory preservation, data efficiency, and freedom to manipulate data. It is indicated that transfer learning can achieve a lot of state-of-the-art models using different techniques. There are also many applications that have been widely implemented in the real world. These NLP applications have become ubiquitous based on the advent of more and more advanced transfer learning solutions such as the adapter.

Some applications such as cyberbullying detection and sentiment analysis share similar features according to the preceding part of the research where abusive language is related to negative sentiment. Thus, the transfer of knowledge between these two areas might be more possible since they are under similar domains.

What is more, sometimes negative transfer might happen that one set of events could hurt performance on the related tasks. There is not as much research on negative transfer detection that is worth discovering in the future.

References

[1] Pan J. and Yang Q., "A Survey on Transfer Learning," in IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[2] Weiss, K., Khoshgoftaar, T.M. & Wang, D. A survey of transfer learning. J Big Data 3, 9 (2016). https://doi.org/10.1186/s40537-016-0043-6

[3] Zaid Alyafeai, Maged Saeed AlShaibani, & Irfan Ahmad. (2020). A Survey on Transfer Learning in Natural Language Processing.

[4] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M. & Gelly, S.. (2019). Parameter-Efficient Transfer Learning for NLP. Proceedings of the 36th International Conference on Machine Learning, in Proceedings of Machine Learning Research 97:2790-2799 Available from https://proceedings.mlr.press/v97/houlsby19a.html.

[5] Chung, Y. A., Lee, H. Y., & Glass, J. (2017). Supervised and unsupervised transfer learning for question answering. arXiv preprint arXiv: 1711.05345.

[6] Min, S., Seo, M., & Hajishirzi, H. (2017). Question answering through transfer learning from large fine-grained supervision data. arXiv preprint arXiv: 1702.02171.

[7] Roy, P. K., & Mali, F. U. (2022). Cyberbullying detection using deep transfer learning. Complex & Intelligent Systems, 1-19.

[8] Uban, AS., Dinu, L.P. (2019). On Transfer Learning for Detecting Abusive Language Online. In: Rojas, I., Joya, G., Catala, A. (eds) Advances in Computational Intelligence. IWANN 2019. Lecture Notes in Computer Science(), vol 11506. Springer, Cham. https://doi.org/10.1007/978-3-030-20521-8_57

[9] Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.

[10] Wang, T., Lu, K., Chow, K. P., & Zhu, Q. (2020). COVID-19 sensing: negative sentiment analysis on social media in China via BERT model. Ieee Access, 8, 138162-138169.

[11] S. D. A. Alzboun, S. K. Tawalbeh, M. Al-Smadi, and Y. Jararweh, ‘‘Using bidirectional long short-term memory and conditional random fields for labeling Arabic named entities: A comparative study,’’ in Proc. 5th Int. Conf. Social Netw. Anal., Manage. Secur. (SNAMS), Oct. 2018, pp. 135–140.

[12] Lee, J. Y., Dernoncourt, F., & Szolovits, P. (2017). Transfer learning for named-entity recognition with neural networks. arXiv preprint arXiv: 1705.06273.

[13] Schuurmans, J., & Frasincar, F. (2019). Intent classification for dialogue utterances. IEEE Intelligent Systems, 35(1), 82-88.

[14] Kumar, M., Kumar, V., Glaude, H., de Lichy, C., Alok, A., & Gupta, R. (2021, January). Protoda: Efficient transfer learning for few-shot intent classification. In 2021 IEEE Spoken Language Technology Workshop (SLT) (pp. 966-972). IEEE.

[15] Artetxe, M., Labaka, G., & Agirre, E. (2020). Translation artifacts in cross-lingual transfer learning. arXiv preprint arXiv: 2004.04721.