Transfer Learning in Knowledge-Intensive Tasks: A Test in Healthcare Text Analytics

Emergent Research Forum (ERF)

Long Xia  
Virginia Tech  
longxia1@vt.edu

David M Goldberg  
Virginia Tech  
goldberg@vt.edu

Sukhwa Hong  
Virginia Tech  
sukhwa@vt.edu

Patricia Garvey  
Virginia Tech  
pkgarvey@vt.edu

Abstract

Transfer learning has been widely utilized in many machine learning applications. However, our literature review suggests that transfer learning, especially relational knowledge transfer learning, has been underutilized in healthcare analytics tasks. In this study, we identify and incorporate three different types of knowledge transfer approaches into a deep learning-based healthcare text analytics model. Our evaluation results suggest that the improved performance can be achieved by incorporating the three relational knowledge transfer learning approaches individually and in combined. This study provides an excellent starting point for studying different types of transfer learning approaches in detail.

Keywords

Transfer learning, deep learning, healthcare text analytics, relational knowledge, ensemble framework.

Introduction

In the last few decades, artificial intelligence (AI) has made a remarkable progress in processing images, video, audio, and text (LeCun et al. 2015). Much of the progress has come from recent advances in the field of machine learning, especially in deep learning. Deep learning has achieved unprecedented performance gains in many domains (LeCun et al. 2015). It has also been extensively adopted in healthcare research. The literature shows machine learning has been utilized in three medical domains: medical image processing, medical text analytics (Cheng et al. 2016), and cancer diagnosis and staging (Esteva et al. 2017). Based on statistical learning theory, machine learning approaches tend to perform well when the dataset is sufficiently large and when the training and testing sets are homogenous (Vapnik 1999). However, in many real-world applications, especially in healthcare domain, these assumptions may not hold. In addition, healthcare is an extremely knowledge-extensive domain, in which solving healthcare related tasks often requires significant domain and expert knowledge (Shen et al. 2016).

Humans can learn patterns from just a few examples and have much better generalizability in applying what they learn to a wide spectrum. For example, humans may only need to see an object once in order to recognize the object in the future. In contrast, the best AI algorithms still require at least hundreds of examples to achieve similar performance. Thus, humans have much better learning and generalizing abilities than AI. One of the integral ingredients in human learning is transferring and integrating knowledge learned previously to solve new problems efficiently and effectively (Pan and Yang 2010).

Inspired by human learning abilities, many machine learning models now adopt transfer learning to improve performance (Li et al. 2012; Perlich et al. 2014). When the training data is not sufficient to build a high-performing model for a task, transfer learning can be utilized to transfer useful knowledge from related source domain to a target domain (Pan and Yang 2010). However, there have been very few uses of this technique in healthcare. In addition, as one important type of transfer learning, relational knowledge transfer learning has been largely overlooked in research. Given the fact that healthcare domain is extremely
knowledge-intensive, the application of transfer learning, especially the relational knowledge transfer learning could be significant. Fortunately, many knowledge databases and related datasets are available, making it possible to conduct an extensive study on the role of transfer learning in healthcare analytics.

Motivated by the research gaps we identified, we propose a deep learning-based framework for healthcare text analytics incorporating three different relational knowledge transfer learning approaches. Our results demonstrate that relational knowledge transfer learning, if it is designed carefully, can significantly improve learning outcomes in knowledge-intensive domains. Our findings suggest that different types of transfer learning can produce synergistic effects in improving the overall performance of the model.

Literature review

Transfer learning has been successfully applied to address many real-world challenges. For example, sometimes we want to solve a task in a target domain but lack voluminous data, and labeling is too time-consuming and often impossible. However, we may have sufficient data in another domain (source domain). In this case, transfer learning could potentially improve the performance of learning by transferring useful knowledge from the source domain to the target domain. Existing studies suggest that transfer learning can be classified to four settings based on what is being transferred: instance-based, feature-based, parameter-based, and relational knowledge-based transfer learning (Pan and Yang 2010).

Many existing studies utilize transfer learning as a standard procedure in building machine learning models. The studies before year 2010 mainly focus on the feature-based transfer learning, as feature extraction and engineering step is often one of the most important components in building machine learning models (Pan and Yang 2010). However, with the recent successes of deep learning in many domains and tasks, this situation has significantly changed. The key advantage of deep learning is that it can directly take raw data as input and automatically discover the features needed for the task (LeCun et al. 2015). Recent studies show that deep learning can learn transferable features which generalize well to new tasks (Long et al. 2015). Given the fact that parameter tuning is an important step in training deep learning models, parameter-based transfer learning has received extensive attention (Perrone et al. 2018).

Transfer learning has also been applied in information systems research. From the theoretical perspective, the mechanism of transfer learning can be best explained by two theoretical perspectives (Kang et al. 2017). The first perspective is from the nature of the tasks, also termed as environmental theory of transfer of learning, which mainly focuses on the characteristics of both source and target tasks. The transfer learning can only be effective if the source task and target task share some common elements and characteristics. The more similarities exist between the source task and target task, the better the transfer learning performance. The second perspective is that, given that similarities between the tasks exist, can they be identified and selected? This is known as the cognitive theory of transfer of learning, which “argues that the likelihood of learning transfer is determined by the ability of learners in retrieving relevant prior experience stored in memory” (Kang et al. 2017). Combining these two perspectives together, we can draw the conclusion that it is not sufficient to ensure that similarities exist between source and target tasks, as we also need to ensure that the similarities will be appropriately recognized and transferred.

Methodology

Despite the great successes achieved by transfer learning, very few healthcare studies have systematically utilized it to improve the model performance. In addition, relational knowledge transfer learning has been significantly overlooked. However, healthcare is a research area that domain-knowledge can be very useful and provide complementary information in addition to the features automatically learned from deep learning model. In this study, we try to fill in these research gaps by studying the applications of transfer learning, specifically relational knowledge transfer learning, in solving a healthcare text analytics task. Deep learning provides a natural platform for study on transfer learning and performance guarantee, so we utilize a deep learning model in our framework. More specifically, we use Long Short Term Memory (LSTM), a state-of-the-art deep learning model for natural language processing (Graves and Schmidhuber 2005). As we are focusing on text analytics tasks, using LSTM can maximize the generalizability of our model.

We operationalize relational knowledge transfer learning using three different approaches. First, word embeddings are utilized to capture the similarities between words. Word embeddings help learning
algorithms achieve better performance in many natural language processing tasks, as they have succeeded in capturing fine-grained semantic and syntactic regularities (Mikolov et al. 2013). The word relations captured by word embeddings have the potential to improve the performance in tackling various healthcare text mining tasks. In addition, the word embeddings pre-trained on in-domain text data can capture some domain-specific relations that are missed or downgraded by general corpus. Thus, we expect healthcare tasks can further benefit from medical in-domain word embeddings. The training of word embeddings is an unsupervised process, meaning almost no manual effort is needed, and can capture linguistic knowledge about a language, which serves as the basic analysis units for text analytics tasks. Thus, transferring general knowledge from a language perspective by word embeddings provides a basis to solve text analytics tasks.

Second, domain knowledge refers to the domain-specific knowledge or context in which the information system or application operates. Consider two sentences: (1) “tires may cause a blowout” and (2) “Metformin may cause a headache”. The first sentence reveals the possible defect in tires, while the second sentence indicates the possible defect (side effect) of the drug Metformin. Although they are referring to similar tasks, these two sentences come from different domains, and each has its own domain-specific knowledge. One possible way to reduce these domain discrepancies is to find the more generic hypernyms that group related elements into the same group. By applying WordNet hypernyms (Miller 1995), we find the hypernyms for “tire” are: “hoop”, “band”, “strip”, “artifact”, “whole”, “object”, “physical_entity”, and the hypernyms for “metformin” are: “antidiabetic”, “medicine”, “drug”, “agent”, “causal_agent”, “substance”, “physical_entity”. Similarly, the hypernyms for “blowout” are: “malfunction”, “failure”, “happening”, “event”, “psychological_feature”, and hypernyms for “headache” are: “ache”, “pain”, “symptom”, “evidence”, “information”, “cognition”, “psychological_feature”. As hypernyms can capture the common generic categories for terms from different domains, they could be used to reduce the domain discrepancies.

Lastly, to capture the important medical in-domain words and their functionalities, we utilized Unified Medical Language System (UMLS), which contains a list of key terminology and associated resources for developing “more effective and interoperable biomedical information systems and services” (Bodenreider 2004). A category label will be assigned to the word if it can be found in the UMLS dictionary.

We now present our deep learning-based relational knowledge transfer learning framework to perform healthcare text analytics tasks, which we display in Figure 1 below. The framework combines the (1) medical in-domain word embeddings, which can be used to capture important medical in-domain semantic and syntactic relationships, (2) relationship knowledge to reduce the domain discrepancy between the target and source tasks, and (3) UMLS medical terms to transfer the biomedical knowledge established by domain experts. Since each component addresses different aspects of the tasks, we expect they are complementary for one another rather than substitutive. Thus, we expect their combination could achieve the best performance. The detailed implementations will be demonstrated in the following section.

---

**Figure 1. Transfer learning framework for healthcare text analytics**

**Experimental design and preliminary results**

We then constructed an adverse drug event (ADE) dataset from drug review data. We identified 5,711 drug reviews and split them into a training set, a validation set, and a testing set. Each word in a review was labeled to one of three entity groups: drug, ADE, or other. Four graduate students were separated into two groups with each group annotating each word separately. A third rater was assigned to make a final decision whenever the two groups disagreed. A summary of the annotated distribution is provided in Table 1.
For evaluation and purposed, we built a LSTM model without utilizing any transfer learning techniques on the ADE dataset as baseline performance. Consistent with prior work, we used precision, recall, and F-measure as our performance measures. Precision refers to the proportion of classified positives that are true positives; recall refers to the proportion of all true positives classified as such; and F-measure is a weighted combination of precision and recall. In the first experiment, we evaluated the effectiveness of medical in-domain word embeddings. We applied the LSTM-based deep learning model on the ADE dataset, and we compared the effectiveness of the medical in-domain word embeddings against the model using regular word embeddings trained on general text corpus. In our second experiment, we evaluated the effectiveness of reducing domain discrepancy. The hypernyms are intended to reduce the discrepancies between different domain-specific knowledge. We evaluated the performance after incorporating these hypernym features on top of the baseline model. In our third experiment, we incorporated the UMLS medical terms, which can capture important medical in-domain relational knowledge, into our baseline model. Lastly, since we are interested in the relationships among different transfer learning settings, we also combined all the three components to deliver the proposed framework depicted in Figure 1. We display the results of these various models in Table 2. For each non-baseline model, we also show the p-value when testing whether that model’s F-measure significantly differs from the baseline model’s F-measure.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Entity classes</th>
<th>Number of records</th>
<th>Percentage of dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>Other</td>
<td>212,638</td>
<td>91.44%</td>
</tr>
<tr>
<td></td>
<td>Drug</td>
<td>7,472</td>
<td>3.21%</td>
</tr>
<tr>
<td></td>
<td>ADE</td>
<td>12,432</td>
<td>5.35%</td>
</tr>
<tr>
<td>Validation Set</td>
<td>Other</td>
<td>43,522</td>
<td>92.31%</td>
</tr>
<tr>
<td></td>
<td>Drug</td>
<td>1,396</td>
<td>2.96%</td>
</tr>
<tr>
<td></td>
<td>ADE</td>
<td>2,231</td>
<td>4.73%</td>
</tr>
<tr>
<td>Testing set</td>
<td>Other</td>
<td>37,934</td>
<td>90.69%</td>
</tr>
<tr>
<td></td>
<td>Drug</td>
<td>1,221</td>
<td>2.92%</td>
</tr>
<tr>
<td></td>
<td>ADE</td>
<td>2,674</td>
<td>6.39%</td>
</tr>
</tbody>
</table>

**Table 1. Label Distribution**

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
<th>P-value for F-measure (compared with baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline LSTM model</td>
<td>0.601</td>
<td>0.645</td>
<td>0.622</td>
<td></td>
</tr>
<tr>
<td>LSTM with medical in-domain word embeddings</td>
<td>0.781</td>
<td>0.758</td>
<td>0.769</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>LSTM with hypernyms</td>
<td>0.726</td>
<td>0.739</td>
<td>0.732</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>LSTM with UMLS terms</td>
<td>0.735</td>
<td>0.783</td>
<td>0.758</td>
<td>&lt; 0.0001***</td>
</tr>
<tr>
<td>LSTM with medical in-domain word embeddings, hypernyms, and UMLS terms</td>
<td>0.810</td>
<td>0.834</td>
<td><strong>0.822</strong></td>
<td>&lt; 0.0001***</td>
</tr>
</tbody>
</table>

**Table 2. Evaluation Results**

Based on the results, each individual transfer learning approach significantly improves the model performance (all p-values are <0.0001), indicating the transfer learning can indeed contribute to the performance improvements. One interesting pattern is that the medical in-domain word embeddings contribute more to improving the performance than the UMLS terms and hypernyms. We believe that this can be explained from two perspectives. From the general task perspective, word embeddings can capture precise syntactic and semantic word relations, which can provide useful initializations in the model training. From the domain-specific task perspective, our medical word embeddings can capture important medical in-domain knowledge to help improve adverse drug event extraction. In addition, after combining all three transfer learning approaches, we achieve the best performance. Contrary to existing studies, which suggest that knowledge in one type may interfere with or substitute for the accumulation of knowledge in another type, our preliminary results suggest that different relational knowledge transfer learning approaches very likely complement one another, thus they produce additional synergistic effects on performance. However, additional experiments need to be conducted to further corroborate this finding.
Conclusions and future steps

In this study, we research the application of relational knowledge transfer learning, a widely overlooked transfer learning approach, in solving a healthcare text analytics task. Our results suggest that each of the three knowledge transferring techniques can significantly improve the learning performance, while their combination can achieve the best overall performance. Proving the usefulness of relational knowledge transfer learning in text analytics task in healthcare domain, which is extremely knowledge extensive, contribute significantly to transfer learning and healthcare analytics literatures.

Although promising results were obtained, we are actively working on improving our proposed framework and expanding to more general scenarios. First, the current framework design provides an excellent starting point for studying transfer learning effects in detail. We are trying to further improve our designed artifact grounded by learning theories. In addition, we will extend our datasets to more domains and task types. We began with a task in healthcare domain as it is one of the most knowledge-intensive domains. We are trying to develop a more generic framework, so that it has the potential to be applied to a wider spectrum of domains and tasks. Lastly, we used word embeddings in our study, which are based on unsupervised learning algorithms and do not require a manual process. This suggests a great potential for applying unsupervised transfer learning. Unsupervised learning has been significantly overlooked because of the exceptional successes of supervised learning. However, it is recognized that human learning is largely unsupervised. Thus, we expect that unsupervised learning will become much more important in a long term. We believe that it would be very promising to develop a humanlike unsupervised transfer learning approach to improve current AI systems.

REFERENCES