Deep Multi-Instance Transfer Learning

Dimitrios Kotzias[1,2] **Misha Denil**[2] **Phil Blunsom**[2,4] **Nando de Freitas**[2,3,4]
[1]University of California, Irvine
[2]University of Oxford, United Kingdom
[3]Canadian Institute for Advanced Research (CIFAR)
[4]Google DeepMind
dkotzias@ics.uci.edu
{misha.denil,phil.blunsom,nando}@cs.ox.ac.uk

Abstract

We present a new approach for transferring knowledge from groups to individuals that comprise them. We evaluate our method in text, by inferring the ratings of individual sentences using full-review ratings. This approach combines ideas from transfer learning, deep learning and multi-instance learning, and reduces the need for laborious human labelling of fine-grained data when abundant labels are available at the group level.

1 Introduction

In many areas of human endeavour, such as marketing and voting, one encounters information at the group level. It might then be of interest to infer information about specific individuals in the groups [15]. As an illustrative example, assume we know the percentage of positive votes for each neighbourhood of a city on a public policy issue. In addition, assume we have features for the individual voters. This paper presents an approach for aggregating this information to estimate the probability that a specific individual, say Susan, voted positive. (If you're Susan, you should be concerned about the privacy of your vote.)

This application is probably of questionable ethical value (other than as a warning on privacy issues), but the same technology can be used to solve important problems arising in artificial intelligence. In this work, we present a novel objective function, for instance learning in an a multi-instance learning setting [10]. Our model, requires a similarity measure between instances in order to optimise the objective function. Deep Neural Networks have been very successful in creating representations of data, that capture their underlying characteristics [12]. This work capitalises on their success by using embeddings of data, as produced by a deep network, as instances for experiments.

In this paper we show that this idea can be used to infer ratings of sentences (individuals) from ratings of reviews (groups of sentences). This enables us to extract the most positive and negative sentences in a review. In applications where reviews are overwhelmingly positive, detecting negative comments is a key step toward improving costumer service.

Figure 1 presents an overview of our approach, which we refer to as deep multi-instance transfer learning. The first step in this approach involves creating a representation for sentences. We do that by training the supervised document convolutional neural network of Denil *et al.* [9] to predict review scores. As a result of this training, we obtain embeddings (vectors in a metric space) for words, sentences and reviews. These embeddings are the features for the individuals (sentences in this case). We chose this model, because it is the state of the art in convolutional networks, and the resulting sentence embeddings are not optimised for the problem we are attempting to solve. This adversarial scenario, illustrates the power of our model to work with multiple distributed representations.

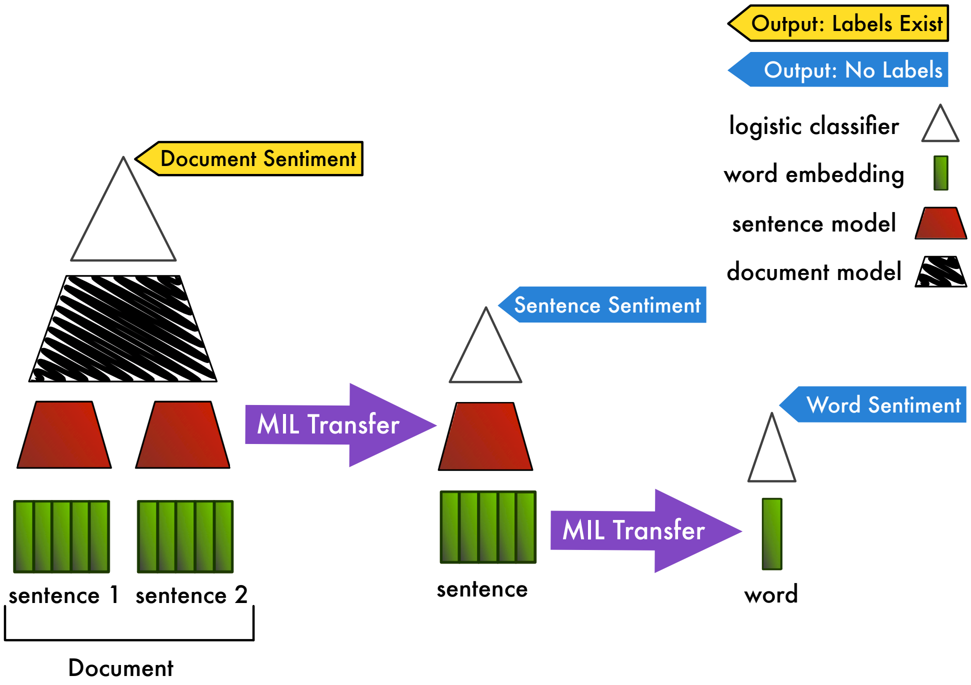


Figure 1: Deep multi-instance transfer learning approach for review data.

Using these features, we formulate a regularized manifold learning objective function to learn the labels of each sentence. That is, we transfer the labels from entire reviews to individual sentences. In doing so, we eliminate the high human cost of gathering labels for individual sentences.

2 Background

2.1 Deep Natural Language Processing

Following the sweeping success of deep learning in computer vision, researchers in deep learning have begun to focus their efforts on other tasks. In particular, applications of neural networks to natural language processing have received considerable attention.

Early work on applying neural networks to language dates back several decades [12, 3]. More recently interest in this area has been renewed by the demonstration that many low level NLP tasks can be solved effectively with convolutional neural networks [6] and also by the development of training methods for distributed representations for words [22, 21].

Moving beyond representations for words, neural network models have also been used to build representations for larger blocks of text. A notable example of this is the paragraph vector [16] which extends the earlier work of Mikolov *et al.* [22] to simultaneously build representations for words and paragraphs. Another recent move in this direction is the work of Denil *et al.* [9] which uses a convolutional neural network to build representations for words, sentences and documents simultaneously. We adopt this convolutional neural network in our work.

2.2 Multi-Instance Learning

Multi-instance Learning is a generalisation of supervised learning, in which labels are associated with *sets of instances*, often referred to as *bags* or *groups*, instead of individual instances. This powerful extension of supervised learning has been applied to a large variety of problems, including drug activity prediction [10], content based image retrieval and classification [20, 27], text categorization [1, 2], music retrieval [19], translation and object recognition [14, 5, 4, 7] and privacy [15, 13].

While there are many variations of multi-instance learning, the key property is that each data example is a bag, and not a single individual. While the goal of some works is to predict labels for new groups, others focus on predicting labels for individual instances in the training or test datasets.

The initial formulation of the multi-instance learning problem by Dietterich *et. al* [10] assumes that the label of each bag is binary, and that for a group to be positive, at least one the instances in the group must have a positive label. Weidmann *et. al* [25] consider a generalisation where the presence of a combination of instances determines the label of the bag. Xu *et. al* [26] assume that all instances contribute equally and independently to a bag's class label, and the bag label is determined by the expected value of the population in the bag. In this work, we use this assumption to derive a regulariser that transfers label information from groups to individuals.

Recent works have considered generalizations where each bag is described in terms of the expected proportion of elements of each class within the bag. Here, the goal is to predict the label of each individual within the bags [15, 23]. For a recent survey on multi-instance learning, we refer the reader to [11]. However, the literature on this topic is vast and that there is disagreement in the terminology. The closest works to ours are the ones of [14, 15, 23, 17].

3 Deep Multi-Instance Transfer Learning

In our formulation of deep multi-instance transfer learning, we are given a set of training instances

$$\mathcal{X} = \{\mathbf{x}_i\}_{i \in I}$$

Unlike in the standard supervised setting, we are not given labels for each training instance directly. Instead we are given labels assigned to groups of instances

$$\mathcal{D} = \{(\mathcal{G}_g, s_g)\}_{g=1,\ldots,G}$$

where $\mathcal{G}_g \subseteq \mathcal{X}$ is a mutli-set of instances from $\mathcal{X}$ and s_g is a label assigned to the group $\mathcal{G}_g$, which we assume to be an unknown function of the (unobserved) labels of the elements of $\mathcal{G}_g$. We are also equipped with a function $W(\mathbf{x}_i, \mathbf{x}_j) \in (0, 1)$ which measures the similarity between pairs of instances. An example illustrating how we construct this similarity measure will be presented in the next section.

Our goals here are twofold. Firstly, we would like to infer labels for each example by propagating information from the group labelling to the instances, essentially inverting the unknown label aggregation function on the training data. To do this we take advantage of the similarity measure to compute a label assignment that is compatible with the group structure of the data, and that assigns the same label to similar instances.

Our second goal is more ambitious. In addition to assigning labels to the training instances we also aim to produce a classifier $y(\mathbf{x}, \boldsymbol{\theta})$ which is capable of predicting labels for instances not found in the training set.

We achieve both of these goals by constructing a training objective for the classifier $y(\mathbf{x}, \boldsymbol{\theta})$ as follows:

$$J(\boldsymbol{\theta}) = \sum_{i,j \in I} W(\mathbf{x}_i, \mathbf{x}_j) \left(y(\mathbf{x}_i, \boldsymbol{\theta}) - y(\mathbf{x}_j, \boldsymbol{\theta})\right)^2 + \lambda \sum_{g=1}^{G} \left(\frac{1}{|\mathcal{G}_g|} \sum_{i \in \mathcal{G}_g} y(\mathbf{x}_i, \boldsymbol{\theta}) - s_g \right)^2 \tag{1}$$

Both terms in this objective can be seen as different forms of label propagation. The first term is a standard manifold-propagation term, which spreads label information over the data manifold in feature space. A similar term often appears in semi-supervised learning problems, where the goal is to make predictions using a partially labelled data set. In such a setting a label propagation term alone is sufficient; however, since we have labels only for groups of instances we require additional structure.

While we have adopted a weighted square-loss, any other weighted loss functions can be used as the first term of the objective function. It simply ensures that similar individual features $\mathbf{x}_i$ are assigned similar labels y.

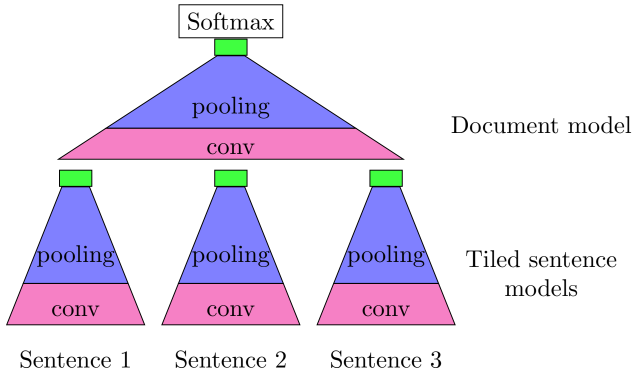


Figure 2: Model from Denil *et al.* [9]. The green squares indicate embedding vectors for sentences (atop the tiled sentence models) and for documents (atop the document model). This model is ideal for our setting because it produces sentence embeddings as an intermediate representation but requires only document level labels for training.

The second term parametrises the whole-part relationship between the groups and the instances they contain, which has the effect of propagating information from the group labels to the group instances. Here we have chosen the simplest possible parametrisation of the whole-part relationship, which says that the label of a group is obtained by averaging the labels of its elements.

Each individual term in the cost function by itself would not work well. This situation is not unlike what we find when we carry out kernel regression with ℓ_1 regularization, where the likelihood term often leads to pathological problems and the regularizer simply has the effect of shrinking the parameters to a common value (typically zero). However, when we combine the two competing terms, we are able to obtain reasonable results.

The parameter λ trades off between the two terms in this objective. The maximum theoretical value of the first term is $|I|^2$, since each summand falls in the interval $[0, 1]$. For the same reason, the second term is bounded by $|G|$. We therefore set $\lambda = \alpha\frac{|I|^2}{|G|}$ in order to trade off between their two contributions directly. Of course it may not be the case that both terms are equally important for performance, which is why we have left α as a parameter.

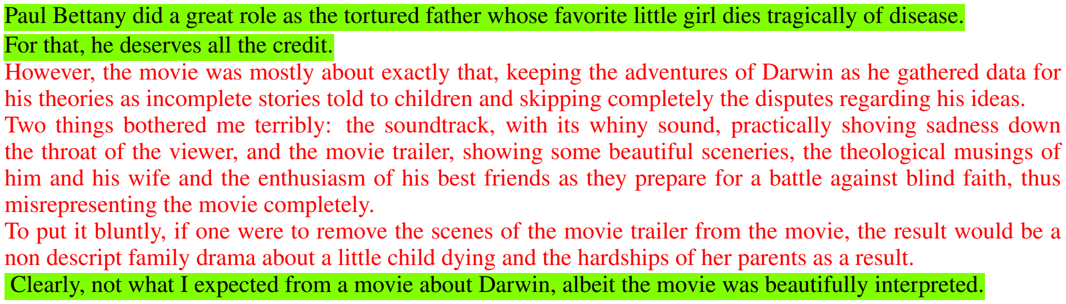
Optimising this objective produces a classifier $y(\mathbf{x}, \boldsymbol{\theta})$ which can assign labels to seen or unseen instances, despite having been trained using only group labels. This classifier simultaneously achieves both of our stated goals: we can apply the classifier to instances of $\mathcal{X}$ in order to obtain labels for the training instances, and we can also use it to make predictions for unseen testing instances.

The power of this formulation relies on having a good similarity measure. It would be simple to take the average score of each instance across groups, and minimise the second term of the objective. However, the presence of the first term pushes similar items to have similar labels.

4 From Review Sentiment to Sentence Sentiment

Sentiment Attribution refers to the problem of attributing the sentiment of a document to its sentences. Given a set of documents and a sentiment label for each one, we attempt to identify how much each sentence in each of these documents contributes positively or negatively towards its overall sentiment. This is a problem with an interesting set of applications, as it allows for a more efficient visualisation of documents, explores causality, and aids towards automatic summarisation.

We can express the sentiment attribution task as a deep multi instance learning learning problem by taking documents to be groups, and the sentences in each document to be individuals. Following the procedure outlined in the previous section, we parametrise the relationship between sentence labels and document labels by assuming that the sentiment of a document is simply the average sentiment of its constituent sentences.



Paul Bettany did a great role as the tortured father whose favorite little girl dies tragically of disease.
For that, he deserves all the credit.
However, the movie was mostly about exactly that, keeping the adventures of Darwin as he gathered data for his theories as incomplete stories told to children and skipping completely the disputes regarding his ideas.
Two things bothered me terribly: the soundtrack, with its whiny sound, practically shoving sadness down the throat of the viewer, and the movie trailer, showing some beautiful sceneries, the theological musings of him and his wife and the enthusiasm of his best friends as they prepare for a battle against blind faith, thus misrepresenting the movie completely.
To put it bluntly, if one were to remove the scenes of the movie trailer from the movie, the result would be a non descript family drama about a little child dying and the hardships of her parents as a result.
Clearly, not what I expected from a movie about Darwin, albeit the movie was beautifully interpreted.

Figure 3: For this review, our approach assigns positive sentiment to the first two and last sentences of the review. The remaining sentences are assigned negative sentiment.

In order to obtain a similarity measure for sentences we take advantage of recent work in learning distributed representations for text. Many works have shown how to capture the semantic relationships of words using the geometry of a continuous embedding space, and more recent works have extended this to learning representations of larger blocks of text [16, 9].

Given a distributed representation for each sentence in our training set we can create a similarity measure by setting

$$W(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2)$$

where $\mathbf{x})$ maps a sentence $\mathbf{x}$ to its distributed representation. If the distributed representations have been created correctly then we should expect nearby points in embedding space to correspond to semantically similar sentences, making the Euclidian norm an appropriate measure of closeness.

We obtain sentence embeddings using the convolutional neural network from Denil et al. [9], which is particularly well matched to our setting. This model is trained using only document level supervision but also produces word and sentence embeddings as an intermediate representation, which we can extract for our own purposes. The structure of this model is shown in Figure 2. We then use the embeddings of sentences as instances, and the binary score of the review as the group score and optimise our objective function with respect to our parameters $\boldsymbol{\theta}$.

4.1 Dataset and Experimental Setup

For evaluation and exploring the problem of sentiment attribution, we use the IMDB movie review sentiment dataset originally introduced by Maas *et al.* [18] as a benchmark for sentiment analysis. This dataset contains a total of 100,000 movie reviews posted on `imdb.com`. There are 50,000 unlabelled reviews and the remaining 50,000 are divided in a 25,000 review training set and a 25,000 review testing set. Each of the labelled reviews has a binary label, either positive or negative. In our experiments, we train only on the labelled part of the training set.

We use NLTK[1] to preprocess each review by first stripping the HTML markup, breaking it into sentences and then breaking each sentence into words. We also map numbers to a generic `NUMBER` token and any symbol that is not in `.?!` to `SYMBOL`. We replace all words that appear less than 5 times in the training set with `UNKNOWN`. This leaves us with a total of 29,493 words, 311,919 sentences in the training set and 305,929 sentences in the testing set.

We train the model of Denil *et al.* [9] to obtain embeddings, $\mathbf{x}_i) \in \mathbb{R}^{24}$, for sentences in the training and testing sets. This also results in word embeddings. In this work, we do not attempt to transfer sentiment to the words.

For these experiments we used a simple logistic regression,

$$y(\mathbf{x}_i, \boldsymbol{\theta}) = \sigma(\boldsymbol{\theta}^\top \mathbf{x}_i) = \frac{1}{1 + e^{-\boldsymbol{\theta}^\top \mathbf{x}_i}},$$

and set the regularisation coefficient in Equation 1 to $\lambda = 0.04\frac{|I^2|}{|G|}$.

[1] `http://www.nltk.org/`

	Precision	Recall
Socher *et al.* [24]	84.5%	100%
MIL Transfer	85.5%	100%
Socher *et al.* [24] (ignoring neutral class)	84.7%	76.2%
MIL Transfer (ignoring neutral class)	92.6%	76.2%

Table 1: Sentence-level classification performance.

We optimize the objective function with stochastic gradient descent (SGD) for 1050 iterations with a learning rate of $\alpha = 0.0001$. We used a batch size of 50 documents, and carried out 7 SGD iterations in each batch, for a total of 3 epochs. Different configurations showed very similar results to those reported.

The time required for training, is in the order of 3 minutes in a consumer laptop. Evaluation time is in the order of 0.1 seconds for all 305,929 sentences in the test set.

As a qualitative measure of the performance of our approach, Figure 3 illustrates the predicted sentiment for sentences in a review[2] from the test set. This is a particularly tricky example, as it contains both positive and negative sentences, which our model identifies correctly. Moreover, the largest part of this review is negative. Hence, the naive strategy of using a simple count of sentences to identify the total sentiment of review, would fail in this example, which accompanied a rating of $8/10$. Our approach on the other hand enables us to extract sentences that best reflect the sentiment of the entire review, and score them at the same time.

4.2 Evaluation

The purpose of our approach is to rely on supervision at the group level to obtain predictions for the individuals in the groups. This weak form of supervision is the most appealing feature of deep multi-instance transfer learning.

As a sanity check, we evaluate the performance of our model as a group (review) classifier. To accomplish this, we average the predicted scores for sentences in each review to classify the test set reviews.

The performance of the sentence score averaging classifier is comparable with the state-of-the art for review classification. The accuracy is 88.47% on the test set and 94.21% on the training set. We emphasize again, that the approach only has access to labels at the review level and must infer the labels of sentences even in the training set. The state-of-the-art on this data set is 92.58% [16].

The good performance of our naive review classifier provides good indication that we have been able to transfer the review labels to infer labels for the sentences. Furthermore it is an indication that we have trained our classifier y correctly.

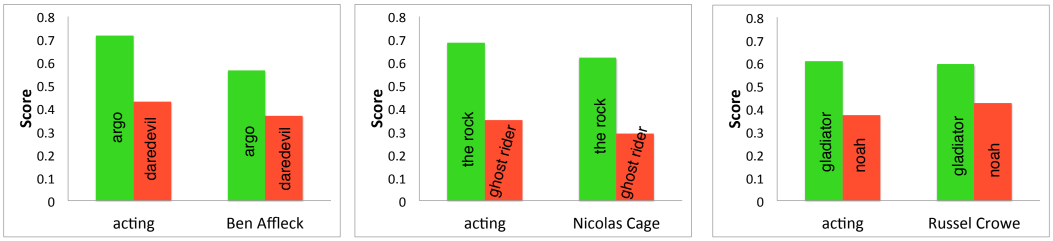
To further evaluate the sentence predictions, we manually labelled 2000 sentences from our dataset as either positive or negative[3]. We split this dataset in half, based on the split by Maas *et al.*, and report the results of sentences from the testing set.

We compared the performance of our approach on this dataset with the Sentiment Analysis tool described in Socher *et al.* [24]. This tool is made available online through a web interface[4] which we use predict labels for our test data. It must be emphasized, that this method is trained with supervision at the phrase-level, while we only require supervision at the review level. It is expensive to obtain labels at the phrase-level, but there exist millions, perhaps billions, of labelled reviews online.

[2] `http://www.imdb.com/title/tt0974014/reviews`

[3] `https://www.cs.ox.ac.uk/people/phil.blunsom/handlabelled_sentence_sentiment.zip`

[4] `http://nlp.stanford.edu/sentiment/` Accessed 20th of June 2014



(a) Ben Affleck

Argo, Daredevil

(b) Nicolas Cage

The Rock, Ghost Rider

(c) Russel Crowe

Gladiator, Noah

Figure 4: Scores associated with the embedding of the word *acting* and the protagonist names, when trained for different movies

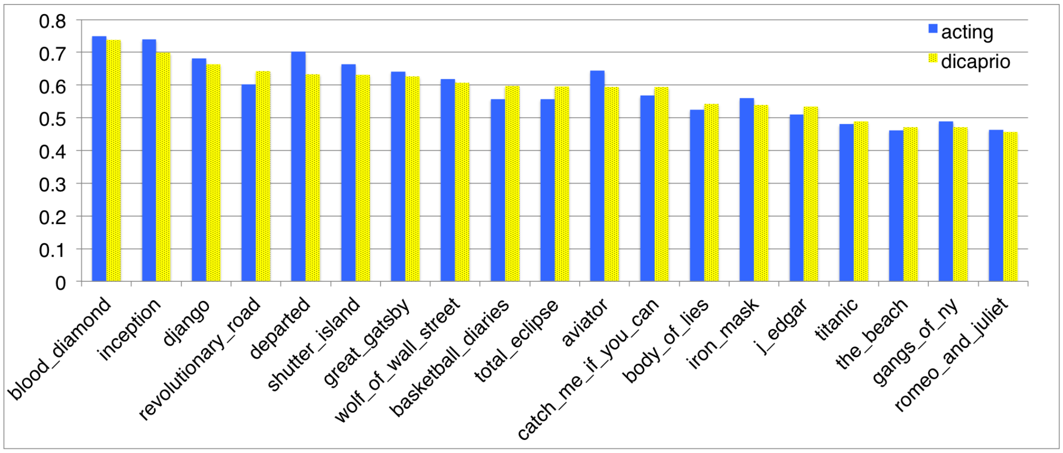


Figure 5: Movies by actor *Leonardo Di Caprio* sorted in order of the inferred sentiment for the embedding of his name, compared with the sentiment for the word *acting*.

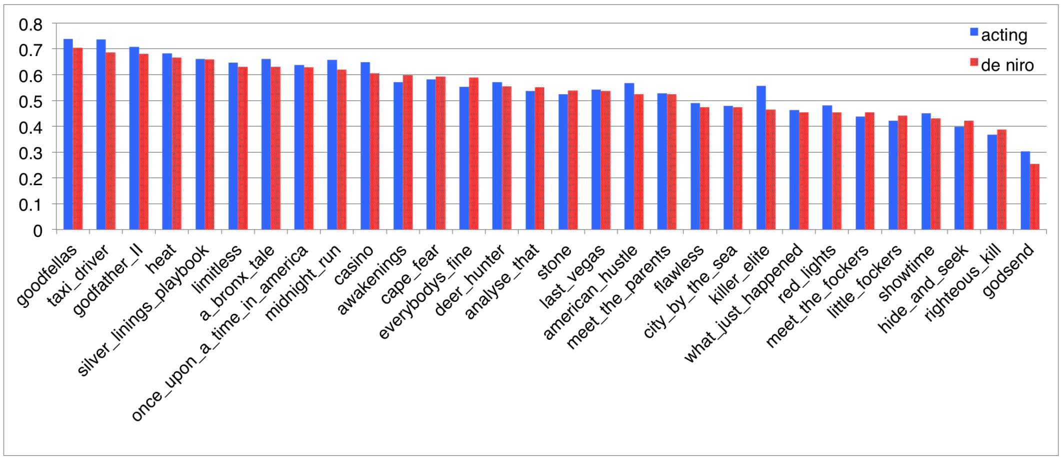


Figure 6: Movies by actor *Robert De Niro* sorted in order of the inferred sentiment for the embedding of his name, compared with the sentiment for the word *acting*.

The method of Socher *et al.* [24] outputs the probability of a sentence belonging to the following five classes: `[Very Negative, Negative, Neutral, Positive, Very Positive]`. Subsequently, it chooses the class of highest probability as the predicted class. To convert this output to a binary decision, we count both `Positive` and `Very Positive` labels as positive, and do the same for negative labels. To manage the `Neutral` class, we consider two strategies. First, we ignore sentences for which the prediction is `Neutral` in the test set, which has the effect of reducing recall. Second, when the label of highest probability is `Neutral`, we use the label of second highest probability to decide whether the review is positive or negative. We report results using both scoring strategies. As shown in Table 1, both strategies achieve similar precision.

Table 1 also shows that our deep multi-instance transfer learning approach achieves higher precision for 100% recall. In order to generate a neutral class with our approach, we introduce a boundary threshold b and label sentences whose positive probability falls in the range $(0.5 - b, 0.5 + b)$ as `Neutral`. We set $b = 0.048$ to calibrate to the recall level as Socher *et al.* [24] when sentences predicted as `Neutral` are ignored. For the same recall, deep multi-instance learning obtains much higher precision.

In spite of the fact that deep multi-instance transfer learning requires much less supervision, it is able to obtain better sentiment predictions for sentences than a state-of-the-art supervised learning approach.

Finally we show how our multi-instance learning approach can be used to obtain entity level sentiment in a specific context. For example, we can predict the sentiment associated with a particular entity (e.g., Leonardo di Caprio) in a chosen context (e.g. a move). To accomplish this we restrict our training data reviews of the chosen move, and train a multi-instance classifier on this restricted data. By restricting the training data to different movies, we force the model to predict sentiment within a specific context. We can then use the context-specific classifier to predict the sentiment associated with an actor's name to obtain sentiment about the actor's role in a specific movie.

Figure 4 illustrates the scores that the same actor achieved in two different movies. The total imdb movie scores agree with the ranking at each case, but more importantly this indicates how the same phrase, can have a completely different sentiment in a different context, which is desirable when ranking queries.

Figures 5 and 6 show this for a series of movies with the actors Leonardo di Caprio and Robert de Niro as the protagonist. The rankings are sorted based on the performance of the actor, and appear to be reasonable thus providing a visual indication that the approach is working well.

5 Concluding Remarks

This work capitalises on the advances and success of deep learning to create a model that considers similarity between embeddings to solve the multi-instance learning problem. In addition, it demonstrates the value of transferring embeddings learned in deep models to reduce the problem of having to label individual data items when group labels are available. Future work will focus on exploring different choices of classifiers, embedding models, other data modalities, as well as further development of applications of this idea.

References

[1] S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. In *Advances in Neural Information Processing Systems*, 2003.

[2] S. Andrews and T. Hofmann. Multiple instance learning via disjunctive programming boosting. In *Advances in Neural Information Processing Systems*, 2004.

[3] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin. A Neural Probabilistic Language Model. *Journal of Machine Learning Research*, 3:1137–1155, 2003.

[4] P. Carbonetto, G. Dork, C. Schmid, H. Kck, and N. de Freitas. Learning to recognize objects with little supervision. *International Journal of Computer Vision*, 77(1-3):219–237, 2008.

[5] P. Carbonetto, G. Dork, C. Schmid, H. Kck, and N. Freitas. A semi-supervised learning approach to object recognition with spatial integration of local features and segmentation cues. In J. Ponce, M. Hebert,

C. Schmid, and A. Zisserman, editors, *Toward Category-Level Object Recognition*, volume 4170 of *Lecture Notes in Computer Science*, pages 277–300. Springer Berlin Heidelberg, 2006.

[6] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. Natural language processing (almost) from scratch. *JMLR*, 12:2493–2537, 2011.

[7] T. Cour, B. Sapp, and B. Taskar. Learning from partial labels. *Journal of Machine Learning Research*, 12:1501–1536, 2011.

[8] G. E. Dahl, R. P. Adams, and H. Larochelle. Training Restricted Boltzmann Machines on Word Observations. In *International Conference on Machine Learning*, 2012.

[9] M. Denil, A. Demiraj, N. Kalchbrenner, P. Blunsom, and N. de Freitas. Modelling, visualising and summarising documents with a single convolutional neural network. Technical report, University of Oxford, 2014.

[10] T. G. Dietterich, R. H. Lathrop, T. Lozano-Perez, and A. Pharmaceutical. Solving the multiple-instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89:31–71, 1997.

[11] J. Foulds and E. Frank. A review of multi-instance learning assumptions. *The Knowledge Engineering Review*, 25(01):1–25, 2010.

[12] G. E. Hinton. Learning Distributed Representations of Concepts. In *Annual Conference of the Cognitive Science Society*, pages 1–12, 1986.

[13] D. Kifer. Attacks on privacy and de Finetti's theorem. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, pages 127–138, 2009.

[14] H. Kueck, P. Carbonetto, and N. Freitas. A constrained semi-supervised learning approach to data association. In *ECCV*, pages 1–12, 2004.

[15] H. Kueck and N. de Freitas. Learning about individuals from group statistics. In *Uncertainty in Artificial Intelligence*, pages 332–339, 2005.

[16] Q. Le and T. Mikolov. Distributed Representations of Sentences and Documents. In *International Conference on Machine Learning*, volume 32, 2014.

[17] P. Liang, M. I. Jordan, and D. Klein. Learning from measurements in exponential families. In *International Conference on Machine Learning*, pages 641–648, 2009.

[18] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150, 2011.

[19] M. I. Mandel and D. P. Ellis. Multiple-instance learning for music information retrieval. In *Proceedings of the 9th International Conference of Music Information Retrieval*, pages 577–582, 2008.

[20] O. Maron and A. L. Ratan. Multiple-instance learning for natural scene classification. In *International Conference on Machine Learning*, pages 341–349, 1998.

[21] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Distributed Representations of Words and Phrases and their Compositionality. In *Neural Information Processing Systems*, pages 1–9, 2013.

[22] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient Estimation of Word Representations in Vector Space. In *International Conference on Learning Representations*, pages 1–12, 2013.

[23] N. Quadrianto, A. J. Smola, T. S. Caetano, and Q. V. Le. Estimating labels from label proportions. *Journal of Machine Learning Research*, 10:2349–2374, 2009.

[24] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642, 2013.

[25] N. Weidmann, E. Frank, and B. Pfahringer. A two-level learning method for generalized multi-instance problems. In *European Conference on Machine Learning*, volume 2837, pages 468–479, 2003.

[26] X. Xu and E. Frank. Logistic regression and boosting for labeled bags of instances. In *PAKDD*, pages 272–281, 2004.

[27] C. Yang and T. Lozano-Perez. Image database retrieval with multiple-instance learning techniques. In *International Conference on Data Engineering*, pages 233–243, 2000.