Project Title

Optional Subtitle

Hermanni Hälvä [1]

MSc. Computational Statistics and Machine Learning

Supervisor: Prof. Bradley Love

Submission date: Day Month Year

[1]**Disclaimer:** This report is submitted as part requirement for the MSc CSML degree at UCL. It is substantially the result of my own work except where explicitly indicated in the text. The report may be freely copied and distributed provided the source is explicitly acknowledged

Abstract

Summarise your report concisely.

Contents

Chapter 1

Literature Review

Learning Hierarchical Visual Representations in DNNs Using Hierarchical Linguistic Labels - Peterson et al. 2018.

Overview

The study most closely related to ours is Peterson *et al.* [1] who investigate whether the use of hierarchical labels helps DNNs to learn better visual representations. As DNNs are typically trained against single labels, the features learned are biased by the arbitrary level of these labels; for instance, different breeds of dogs each have their own label in the widely used ImageNet data-set, but there is no information in the labels that informs the models that all the different breeds are part of a larger category of 'dogs' [1]. In order to control for these hierarchical relationships in categories, the authors define two levels of labels for each image, called baseline (top-level, e.g. 'dog') and subordinate level (original low-level label e.g. 'golden retriever'). Peterson *et al.* argue that this is much closer to how humans represent and learn object categories. To test this, they train four versions of the InceptionV3 [**?**] DNN architecture on the ImageNet Large Scale Visual Recognition Challenge 2012 data (ImageNet hereon): original pre-trained model that only uses subordinate-level labels, model pre-trained on the subordinate labels and fine-tuned on basic-level labels, model pre-trained on basic-level and fine tuned on subordinate lables, and finally a model trained purely on basic-level labels. Several interesting results were attained. First, the authors found that including the basic-level labels led to a much more clustered representation of the feature spaces (e.g. the features vectors of different breeds of dogs were now bundeled up together whilst if trained on just subordinate labels, then there was no clustering of similar categories). Additionally, dendrogram of hierarchical clustering of thelearned feature space with basic labels showed a clear separation between between nature related objectes 'natural images' and images of man-made objects 'artificial images', eventhough no such information about the two groups was given to the model *a priori*. The authors note that this is close to humans' mental representations. To further compare the model's learn representation to humans, the authors investigate how well the model is able to capture human similarity judgements. For the models, similarity between two images is measured as the inner product of their feature representations, and this was compared to human ratings (on scale 1 to 10) of similarity of the same images. Whilst the authors found that on aggregate including basic labels improved the explanatory power of the DNN, the R^2 was not very high (0.57) and as noted, similar results have in the past been attained using only the subordinate

labels. Finally, the paper also presents a generalization experiment in which the model is given just a few examples of either sub or basic level images and then told to find other images from the data set which it would predict to have the same label. The results of this few-shot generalization experiment show similar results as corresponding human studies [**?**] in that introducing basic-level in model training leads to basic-level bias (even if the example given is of subordinate level, the model will generalize by seeking matches in basic level).

Thoughts:

- This is probably the most relevant papers to ours, though I dont actually think it's too similar because they use only two levels of labels so focus on just basic/subordinate, so I dont really see this as hierarchical. It's quite different from how we are going to do with word2vec or some other embedding that allows a much richer relationship between the words rather than just a 'vertical' link between two categories.
- They approach quite strongly from psychology point of view and dont even talk about the accuracy of the model. I think I will have more ML focus than this and will definitely look at the models' performance
- I do like the psychology approach in that if we think of AI more generally then I suppose we wish to achieve human-like semantic understanding and one could argue that this paper perhaps has some of that going on
- I am tempted to also use InceptionV3 in case we do wish to repeat any of their experiments, and for the reason it was used here which is that it's near state-of-the-art and pretty quick to train
- Find it bit weird that they define the model's feature space to be just the final layer: '..we pose multi-level labeling problem simply as learning a set of independent softmax classifiers that are unconnected to each other and fully connected to the final representation layer of deep CNN while other alternative approaches exist for defining the network architecture and loss function, this approach provides a single embedding space for all images, which allows us to inspect the representations with classic psychological methods such as hierarchical clustering.' Not the biggest fan of this approach as would expect also hierarchy of representations through out the network (c.f. human visual cortex)
- further, with above in mind the authors only fine-tune the final layer: 'For fine-tuning models, we freeze all but the weights in the last block of the model to speed up training'. If we wish to look a representation across hierarchy of layers, we cant do this. Hopefully this wont be too much computation...

Bibliography

[1] J. C. Peterson, P. Soulos, A. Nematzadeh, and T. L. Griffiths, "Learning Hierarchical Visual Representations in Deep Neural Networks Using Hierarchical Linguistic Labels," 2018.