Beyond Labels: Knowledge Elicitation using Deep Metric Learning and Psychometric Testing

Lu Yin

Eindhoven University of Technology

l.yin@tue.nl

Abstract

Knowledge present in a domain is well expressed as relationships between corresponding concepts. For example, in zoology, animal species form complex hierarchies; in genomics, the different (parts of) molecules are organized in groups and subgroups based on their functions; plants, molecules, and astronomical objects all form complex taxonomies. Nevertheless, when applying supervised machine learning (ML) in such domains, we commonly reduce the complex and rich knowledge to a fixed set of labels. This oversimplifies and limits the potential impact that the ML solution can deliver. The main reason for such a reductionist approach is the difficulty in eliciting the domain knowledge from the experts. Developing a label structure with sufficient fidelity and providing comprehensive multi-label annotation can be exceedingly labor-intensive in many real-world applications. Here, we provide a method for efficient hierarchical knowledge elicitation (HKE) from experts working with high-dimensional data such as images or videos. Our method is based on psychometric testing and active deep metric learning. The developed models embed the high-dimensional data in a metric space where distances are semantically meaningful, and the data can be organized in a hierarchical structure.

1 Introduction

Supervised learning with a flat set of labels is the panacea of machine learning (ML) application to a wide range of problems in many fields. In practice, this can be an excitingly limiting the success of the application. One can imagine a scenario of a medical diagnosis involving images (for example diabetic retinopathy) and a set of diagnostic images (e.g. photography of the eye retina) and a set of labels denoting the severity of the disease (e.g. 'no disease', 'severity 1', ... 'severity 4'). From the very beginning, the domain expert is forced to project her comprehensive knowledge on this flat single axis. This, in turn, limits our model both from its performance as it cannot learn from the full depth of knowledge as well from its interpretability as projecting to a fixed set
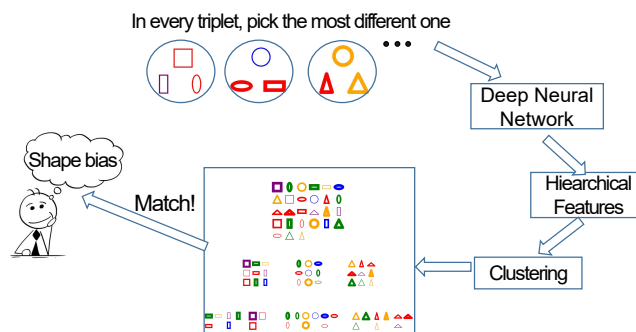

Figure 1: Proposed hierarchy knowledge elicitation framework, which is a example to elicitate and confirm the shape bias human have, that is shapes are more important than other properties of an object, such as colors when we aim to discriminate between them.

of points further contributes to the 'black-box-ness' of our model.

In contrast, adding a more rich hierarchical structure to the data adds a significantly larger understanding of the finer subtleties of the data stemming from the relationships between the concepts in the domain.

The reason why so many solutions are formulated as mapping of datapoints (commonly high dimensional) to a fixed set of labels is that eliciting the knowledge from the domain and forming a training dataset is a difficult task. A data point can have a rich set of properties associated with it and eliciting an expert to provide an exhaustive annotation for each datapoint is prohibitively unscalable.

In our thesis, we propose an approach which able to elicit individual knowledge in efficient, easy to carry way. Psychometric test [Gescheider, 2013] is introduced to capture the user's latent perception. It is to assess a participant's ability to perceive differences between different objects, which has been adapted to measure complex images and videos [Menkovski and Liotta, 2012; Feng et al., 2014; Hadizadeh and Bajić, 2013]. The captured differences is mapped to an embedding space by metric learning, where the em-bedded data are then structured in a hierarchical fashion that allows assigning properties and annotations to a large number of data points in a scalable way.
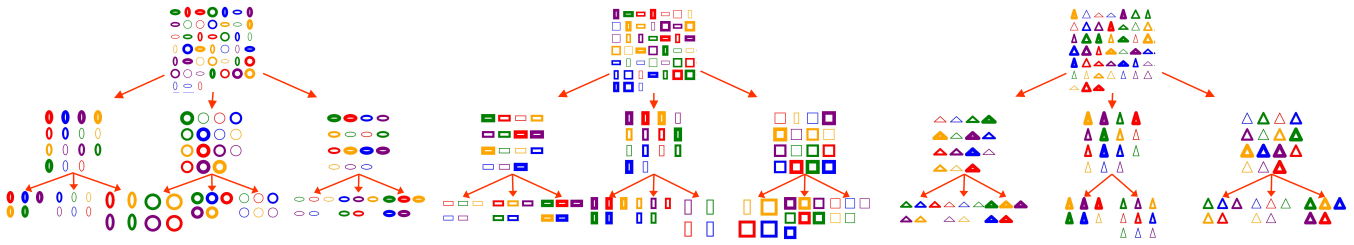
Figure 2: Extracted hierarchy of synthetic simple shape data-set, which empirical shows human have shape bias

2 Finished Work

We already tested our work in image data-sets. To elicitate hierarchical knowledge from image data sets, generally, there are two approaches. One is labeling each data point manually with multiple tags, then building a hierarchical structure with them. However, without pre-defined labels, annotators could give labels arbitrarily, and that could lead to a disorder and meaningless hierarchic structure. Even if we have the pre-provided labels, it is still a quite time-consuming and laborious method. On the other hand, some scholars try to group hierarchical images automatically based on language or visual features correlations. A common way to do these is to take advantage of the existing word semantic information (e.g.,WordNet) or the extracted images feathers. While this method is more efficient, it has some limitations: (1) for some data sets there is no prior word semantic information provided (e.g., in some expert domain) (2) The word semantic information or extracted images feathers can not represent user's perception, so the formed construct may not catering to the user's individual needs.

To cope with these limitations, we proposed an efficient hierarchical knowledge elicitation (HKE) framework as shown in is Fig. 1.

One of the main differences between our work and others is the way to extract the latent hierarchical knowledge from annotators' mind. Capturing it directly could be difficult, because annotators need the know the label classes in advance, and for one object the annotator has to give multiple hierarchical labels. Here, we take advantage of the psychometric discrimination test [Gescheider, 2013] to simplify the process. To be specific, a three-alternative-forced-choice test (3AFC) [Green et al., 1966] is employed in our work. As shown in Fig 3, the annotators are given 3 objects a_1, a_2, a_3, and are forced to choose the most dissimilar one among them. By carrying out this simple test, we can elicit the extract hierarchical knowledge from the data by simple ternary decision.

Psychometric testing offers many advantages. However, the number of all possible discriminative tests is k combination of n, where k is 3, and n is the number of datapoints in the dataset. Asking all possible questions is typically not feasible. To address this, we develop a question selection approach using a Bayesian-based active learning method that selects questions with high uncertainty and high utility.

Furthermore, We propose our dual-triplet loss with adaptive margin to construct a more clear hierarchical structure.

We provide an empirical evaluation on a synthetic data set of simple shapes. The results demonstrate that our method is able to uncover hierarchical structure on image data sets, and confirm the shape bias in the responses of a human participant as shown in Fig. 2.

3 Limits and Future Work

(1) we only studied single participator now, we are also interested in whether the model can elicit specified hierarchies when there are multiple users who do not fully agree with each other, and if there is a general hierarchic structure that captures the shared perception of all the users. (2) It would also be interesting to see other stimuli such as voice be researched using such a method. Classification of someone's voice whether they are angry or happy can also be a difficult task if we work with labels since they are not always very obvious. Therefore, using this method will potentially embed voice recordings in such a way that they become more interpretable. This can be very useful to any company having to deal with customers via the phone and able to detect quickly what the mood of someone is.

References

[Feng et al., 2014] Hsin-Chang Feng, Michael W Marcellin, and Ali Bilgin. A methodology for visually lossless jpeg2000 compression of monochrome stereo images. *IEEE Transactions on Image Processing*, 24(2):560–572, 2014.

[Gescheider, 2013] George A Gescheider. *Psychophysics: the fundamentals*. Psychology Press, 2013.

[Green et al., 1966] David Marvin Green, John A Swets, et al. *Signal detection theory and psychophysics*, volume 1. Wiley New York, 1966.

[Hadizadeh and Bajić, 2013] Hadi Hadizadeh and Ivan V Bajić. Saliency-aware video compression. *IEEE Transactions on Image Processing*, 23(1):19–33, 2013.

[Menkovski and Liotta, 2012] Vlado Menkovski and Antonio Liotta. Adaptive psychometric scaling for video quality assessment. *Signal Processing: Image Communication*, 27(8):788–799, 2012.

Question: which object is the most dissimilar to the other two objects?

| Option a_1 | Option a_2 | Option a_3 |

Figure 3: Three-alternative-forced choice interface