

Representation Learning @ Scale

Abstract:

Machine learning techniques are reaching or exceeding human level performances in tasks like image classification, translation, and text-to-speech. The success of these machine learning algorithms have been attributed to highly versatile representations learnt from data using deep networks or intricately designed Bayesian models. Representation learning has also provided hints in neuroscience, e.g. for understanding how humans might categorize objects. Despite these instances of success, many open questions remain.

Data come in all shapes and sizes: not just as images or text, but also as point clouds, sets, graphs, compressed, or even heterogeneous mixture of these data types. In this thesis, we want to develop representation learning algorithms for such unconventional data types by leveraging their structure and establishing new mathematical properties. Representations learned in this fashion were applied on diverse domains and found to be competitive with task specific state-of-the-art methods.

Once we have the representations, in various applications its interpretability is as crucial as its accuracy. Deep models often yield better accuracy, but require a large number of parameters, often notwithstanding the simplicity of the underlying data, rendering it uninterpretable which is highly undesirable in tasks like user modeling. On the other hand, Bayesian models produce sparse discrete representations, easily amenable to human interpretation. In this thesis, we want to explore methods that are capable of learning mixed representations retaining best of both the worlds. Our experimental evaluations show that the proposed techniques compare favorably with several state-of-the-art baselines.

Finally, one would want such interpretable representations to be inferred from large-scale data, however, often there is a mismatch between our computational resources and the statistical models. In this thesis, we want to bridge this gap by solutions based on a combination of modern computational techniques/data structures on one side and modified statistical inference algorithms on the other. We introduce new ways to parallelize, reduce look-ups, handle variable state space size, and escape saddle points. On latent variable models, like latent Dirichlet allocation (LDA), we find significant gains in performance.

To summarize, in this thesis, we want to explore three major aspects of representation learning --- diversity: being able to handle different types of data, interpretability: being accessible to and understandable by humans, and scalability: being able to process massive datasets in a reasonable time and budget.



Speaker:

Manzil Zaheer

Thesis Committee:

Barnabas Poczos, Co-Chair
Ruslan Salakhutdinov, Co-Chair
Alexander J Smola (Amazon)
Andrew McCallum (UMass Amherst)

Jan. 12, 2018

3:00pm

8102 GHC

Link to draft document:
<http://manzil.ml/proposal.pdf>

