Introduction

This workshop deals with the evaluation of general-purpose vector representations for linguistic units (morphemes, words, phrases, sentences, etc). What distinguishes these representations (or embeddings) is that they are not trained with a specific application in mind, but rather to capture broadly useful features of the represented units. Another way to view their usage is through the lens of transfer learning: The embeddings are trained with one objective, but applied on others.

Evaluating general-purpose representation learning systems is fundamentally difficult. They can be trained on a variety of objectives, making simple intrinsic evaluations useless as a means of comparing methods. They are also meant to be applied to a variety of downstream tasks, which will place different demands on them, making no single extrinsic evaluation definitive. The best techniques for evaluating embedding methods in downstream tasks often require investing considerable time and resources in retraining large neural network models, making broad suites of downstream evaluations impractical. In many cases, especially for word-level embeddings, these constraints have led to the rise of dedicated evaluation tasks like similarity and analogy which are not directly related either to training objectives or to downstream tasks. Tasks like these can serve a valuable role in principle, but in practice performance on these tasks has not been highly predictive of downstream task performance.

This workshop aims foster discussion of these issues, and to support the search for high-quality general purpose representation learning techniques for NLP. The workshop will accept submissions through two tracks: a proposal track will showcase submitted proposals for new evaluation techniques, and a shared task will accept submissions of new general purpose sentence representation systems – for which standard evaluations are notably absent – which will be evaluated on a sentence understanding task.