Weakly Supervised Learning for Processing Sequential Data

Summary

School of Information Technologies, Department of Software Sciences offers a 4-year PhD position in language technology.

Research field:	Information and Communication Technology
Supervisor:	Tanel Alumäe
Availability:	This position is available.
Offered by:	School of Information Technologies
	Department of Software Science
Application deadline:	Applications are accepted between June 01, 2020 00:00 and July 03, 2020 23:59 (Europe/Zurich)

Description

The goal of this PhD topic is to develop efficient methods for training AI models to process (segment, classify, transform, forecast) various types of sequential data, using weak or distant supervision.

Most commonly, sequential data is described as multidimensional data points occurring after fixed time intervals. Some examples of such types of data are speech, video, electricity consumption, heart rate and spatio-temporal user demand of transport services.

Developing accurate models that analyze sequential data is usually based on hand-annotated training data. For example, training models for automatic video annotation requires a large hand-labeled corpus which is costly to produce.

However, for many types of data, partial annotations or metadata are often readily available. For example, videos in news portals are often accompanied by news stories. While distant supervision is not as straightforward to use as supervised data, it can nonetheless provide valuable information that can drastically improve the performance of AI models when no or only small amount of fully supervised data is available.

The proposed topic addresses the problem of developing techniques (model architectures, learning methods) for efficient learning from weakly supervised data. The methods will be evaluated on tasks from different domains, such as speech and video. Using these tasks, we will establish the data efficiency of the developed weakly supervised methods, assessing, for each technique, what is the equivalent amount of labeled data that would result in similar performance of the model.

To get more information or to apply online, visit https://taltech.glowbase.com/positions/80 or scan the the code on the left with your smartphone.