

Graduiertenkolloquium Angewandte Informatik

Multi-view Representation Learning for Unifying Languages, Knowledge and Vision

M. Sc. Aditya Mogadala, AIFB

The growth of heterogeneous content on the web has raised varied challenges, yet also provided numerous opportunities. Content is either represented with textual form existing in different languages or appear as a visual embodiment like images and videos. In many occasions, different amalgamation of them co-exist to complement each other or to provide consensus. For solving challenges that requires intelligent content processing and has dependency between diverse domains, leveraging such multiple view data instances with data driven learning will be beneficial. Despite having availability of such content. However, many data driven learning (i.e. machine learning) approaches still solve tasks separately from varied computer science domains such as computer vision (CV), natural language processing (NLP) and semantic web. Similar endeavor is not shown for the tasks which require input either from all of those domains or subset of them.

In this dissertation, we develop models and techniques that can leverage prevailing multiple views of data instances for assisting several challenges that require connection between aforementioned domains. In particular, first we develop models that can jointly model diverse representations arising from two views of data instances by learning their common space representation. Specifically, modeling such common space representations are helpful to retrieve, recommend and classify cross-view content. Second, we develop models that can cater more than two views to generate one view from another. Lastly, we describe a model that can handle missing views, we demonstrate that this model also can generate one view from another by utilizing auxiliary data. We argue that techniques the models leverage internally provide many practical benefits and lot of immediate value applications. From the modeling perspective, our contributed model design in this thesis can be summarized under the phrase Multi-view Representation Learning. These models are variations and extensions of shallow statistical and deep neural networks approaches that can jointly optimize and exploit all views of the input data arising from different independent representations. We show that our models advance the state of the art on tasks such as cross-modal retrieval, cross-channel recommendations, cross-language text classification, image-caption generation in multiple languages and caption generation for images containing novel visual object categories. In general, they also provide assistance in unifying languages, knowledge and vision.

Termin: Freitag, 17. November 2017, 14.00 Uhr

Ort: Kaiserstr. 89, 76133 Karlsruhe

Kollegiengebäude am Kronenplatz (Geb. 05.20), 1. OG, Raum 1C-04

(Hinweise für Besucher: www.aifb.kit.edu/web/Kontakt)

Veranstalter: Institut AIFB, Forschungsgruppe Web Science

Zu diesem Vortrag lädt das Institut für Angewandte Informatik und Formale Beschreibungsverfahren alle Interessierten herzlich ein.

A. Oberweis, H. Sack, H. Schmeck, Y. Sure-Vetter (Org.), J. M. Zöllner