

Masterarbeit in Informatik/Wirtschaftsinformatik

Ausgangssituation

Kaum ein System kann heute noch ohne Informations- und Kommunikationstechnologien (IKT) existieren; der Austausch von Daten ist für alle Geschäftsprozesse essenziell. Jedoch existiert selten für eine Anwendungsdomäne ein einziges Datenformat, vielmehr liegt fast überall eine Fragmentierung in verschiedene Formate unterschiedlicher Hersteller, die sich teilweise sogar innerhalb verschiedener Versionen unterscheiden, vor. Um diese Systeme miteinander kommunizieren zu lassen, muss entweder ein gemeinsames Datenaustauschformat vorliegen, oder ein Übersetzer existieren, der die Datenfelder eines Formats auf das andere abbildet (Mapping). Dies schließt nicht nur die reine Syntax ein, sondern häufig auch semantische Eigenarten der jeweiligen Formate. Die Erstellung solcher Mapper ist zeitaufwändig; ein wirksamer Automatismus existiert nicht.

Aufgabenstellung

Künstliche neuronale Netze haben bisher in vielen Disziplinen erstaunliche Erfolge erreicht; die automatische Übersetzung von Texten ist eine davon. Das Training solcher Netze ist zeitintensiv, da viele semantische Feinheiten beachtet werden müssen: Sprachverständnis ist nicht nur für Rechtschreibung und Grammatik, sondern auch vom Kontext abhängig. Gleichzeitig ist der Trainingsprozess auch datenintensiv. Durch Transfer Learning wurde die Trainingszeit hier deutlich reduziert (Kocmi und Bojar 2018). Transfer Learning beschreibt die Methodik, existierende, trainierte Netze auf eine andere, aber verwandte Problemstellung zu übertragen; dabei wird lediglich die Eingabeschicht bzw. die unmittelbar folgenden Schichten ersetzt und nachtrainiert. Diese Taktik soll in der hier ausgeschriebenen Thesis auf elektronische Datenaustauschformate (electronic data interchange, EDI (Nöcker 2002)) angewandt werden: Basierend auf existierenden Daten soll ein künstliches neuronales Netz mittels Transfer Learning trainiert werden, um den manuell erstellten Mapper zu ersetzen und die Übersetzung von einem EDI-Format auf ein anderes zu automatisieren.

An wen richtet sich die Themenstellung?

Die Arbeit richtet sich an Studierende der Informatik oder Wirtschaftsinformatik, die Interesse an aktuellen Technologien des Deep Learning haben und sie auf eine Problemstellung im EDI-Umfeld anwenden möchten. Ein Besuch der Vorlesung „Computational Intelligence“ ist wünschenswert.

Betreuung

Dr.-Ing. Eric MSP Veith | eric.veith@offis.de

Prof. Dr. Sebastian Lehnhoff | lehnhoff@offis.de

Literaturverzeichnis

Kocmi, T., and O. Bojar. "Trivial Transfer Learning for Low-Resource Neural Machine Translation." *arXiv*. September 2018. <https://arxiv.org/abs/1809.00357> (accessed Mai 15, 2019).

Nöcker, Gregor. *Die beleglose Spedition*. LIT Verlag, 2002.

Surafel M. Lakew, Aliia Erofeeva, Matteo Negri, Marcello Federico, Marco Turchi. "Transfer Learning in Multilingual Neural Machine Translation with Dynamic Vocabulary." *arXiv*. November 3, 2018. <https://arxiv.org/abs/1811.01137> (accessed Mai 15, 2019).

