

GENTIO

Generative Learning Networks for Text and Impact Optimization

Programm / Ausschreibung	IKT der Zukunft, IKT der Zukunft - 7. Ausschreibung (2018)	Status	abgeschlossen
Projektstart	01.01.2020	Projektende	30.06.2023
Zeitraum	2020 - 2023	Projektlaufzeit	42 Monate
Keywords	Generative Networks, Deep Learning, Natural Language Processing, Knowledge Extraction		

Projektbeschreibung

Die automatische Extraktion von faktischem, affektivem und kontextuellem Wissen aus digitalen Inhalten ist ein aktives Forschungsfeld mit wesentlichen Entwicklungen und Fortschritten in den vergangenen Jahren. Aufbauend auf den neuesten Fortschritten zielt GENTIO darauf ab, den Zugang zu Herstellung, Anreicherung und Analyse von digitalen Inhalten grundlegend zu revolutionieren. Eine zentrale Innovation des Projektes ist dabei die Entwicklung eines Multi-Task Learning (MTL) Ansatzes – aufbauend auf sogenannten Generative Learning Networks – zur Integration dreierlei verschiedener Ebenen von Textinterpretation: der strukturellen, der inhaltlichen, und der kontextuellen Textinterpretation. Das Projekt zielt einerseits darauf ab, die Fähigkeit von Computerlinguistiksystemen zur Verarbeitung und Kontextualisierung digitaler Inhalte zu verbessern, andererseits auch darauf, die hohen Erstellungskosten von Trainingsdatensätzen zu reduzieren. Dadurch wird eine kosteneffiziente Entwicklung von intelligenten semantischen Systemen möglich. Interaktive Visualisierungen der extrahierten Features geben dem Benutzer zusätzlich Einblick und Verständnis der erlernten Modelle, was bei bestehenden Ansätzen von Maschinellem Lernen oftmals vernachlässigt wird.

Die Verwendung mehrsprachiger und zugleich zuverlässig skalierbarer Wissensgraphen-Technologien macht den geplanten Ansatz in multiplen Domänen und Regionen anwendbar. Zur Evaluation der vorgeschlagenen Methode sowie zur Demonstration deren Vielfältigkeit wurden zwei unterschiedliche Use Case Szenarien ausgewählt. Das erste Szenario betrifft die Marketing-Domäne. Hier wird mit neuartigen Methoden experimentiert, die Kommunikationsexperten in die Lage versetzen sollen, die Wirkung ihrer datengetriebenen Publikationstätigkeit zu maximieren. Das zweite Szenario ist im Nachrichtensektor verortet und zielt auf die automatisierte Korrektur und Klassifikation fehlerbehafteter Ausgaben optischer Zeichenerkennungssysteme (OCR) ab. Themenfelder aus der öffentlichen Debatte auf anderen Mikroblogging-Kanälen wird dabei genutzt, die notwendige Kontextinformation zu extrahieren.

Beide Anwendungsszenarien ermöglichen es GENTIO sowohl die Produktion als auch die Analyse digitaler Inhalte zu beleuchten, angetrieben durch zwei führende Anwendungspartner mit sehr unterschiedlichen Anforderungen. Ketchum Publico als österreichischer Vertreter eines globalen Kommunikationsdienstleisters, und OBSERVER als etabliertes, österreichisches KMU im Bereich Media Intelligence mit mehr als 100 Jahren Erfahrung. Als Teil der Planung der Verwertungsstrategie wird GENTIO in der zweiten Projekthälfte andere Anwendungsfelder analysieren, um das Potential seines MTL-Konzepts in unterschiedlichen Domänen zu demonstrieren, unter anderem in der Rundfunkbranche (semantische

Suche für Videosuche), im Einzelhandel und Endverbrauchermarketing (Reputation Management), in der Telekommunikationsbranche (Helpdesk und Support), im Consulting und Auditing (Evaluation und Annotation von Gesetzestexten), sowie im Mobilitätssektor (Crowd-basierte Feedbacksysteme für autonome Fahrzeuge).

Abstract

Recent years have shown major advances in the automated extraction of factual, affective and contextual knowledge from digital content streams. GENTIO builds on these advances to change the way we produce, enrich and analyse digital content. The project will develop a flexible Multi-Task Learning (MTL) approach based on Generative Learning Networks to unify the understanding of text at three fundamental levels: structure, content and context. Thereby it aims to boost the context processing capabilities of Natural Language Processing (NLP) frameworks, reduce the high cost of developing training data, and support the cost-effective development of intelligent semantic systems. By offering interactive visualizations to explore the extracted features, the project will also put special emphasis on increasing the transparency of the underlying computational processes, which is a typical shortcoming of Artificial Intelligence-based systems.

Supported by multilingual and highly scalable knowledge graph technology, the envisioned approach will be applicable across numerous domains and regions. To demonstrate its versatility, two distinct domains have been chosen. The first use case targets the marketing domain. It will experiment with new methods for communication experts to maximize the impact of data-driven publishing. The second use case targets the news media sector, automatically correcting and classifying noisy output from Optical Character Recognition (OCR) systems – using topics extracted from the public debate on other microblogging sites to obtain the required context information.

The two use cases allow GENTIO to investigate the production as well as the analysis of digital content, driven by leading use case partners in their respective fields – Ketchum Publico as the Austrian representative of a global communications consultancy versus the OBSERVER as an established Austrian media intelligence SME with a history of more than 100 years. As part of the exploitation planning in the second half of the project, GENTIO will clearly define the potential of using its MTL capabilities in a variety of other domains including broadcasting (semantic search for video retrieval), retailing and consumer brands (reputation management), telecommunications (helpdesk and support), consulting and auditing (legal text annotation and evaluation) and mobility (crowd-based feedback systems for autonomous driving applications).

Projektkoordinator

- webLyzard technology gmbh

Projektpartner

- "Observer" Gesellschaft m.b.H.
- Ketchum GmbH
- Modul Technology GmbH
- Wirtschaftsuniversität Wien Forschungsinstitut für rechenintensive Methoden