Including a contemporary NLP application within an introductory course: an example with student feedback from a University of Applied Sciences

Saurabh Kumar and Alessandra Zarcone
FORVIA Clean Mobility, Augsburg, Germany; Technische Hochschule Augsburg, Augsburg, Germany

FORVIA ·faurecia | THA Technische Hochschule Augsburg | HIGHTECH Agenda Bayern

Why a contemporary NLP application for an introductory course?

➢ Heightened curiosity about recent NLP applications among students, in particular in Universities of Applied Sciences

➢ It is not trivial to find **contemporary topics**
  which are suitable for an introductory course in NLP
  ➢ Theoretical background
  ➢ Computational requirements
  ➢ Datasets

> A **University of Applied Sciences** (Hochschule in the German system)
> ➢ focuses on industrial application
> ➢ maintains strong ties with local industrial partners
>   (technology transfer, applied research projects)
> ➢ integrates practical skills and applied research topics in its teaching

Semantic Textual Similarity (STS, Agirre et al., 2012) as a contemporary application

➢ **Proposal:** use **Semantic Textual Similarity**
  to introduce the challenges of sentence embeddings
  as well as their relevance for real-world use cases
  ➢ focus on popular semantic IR systems using
    sentence embeddings (Reimers and Gurevych, 2019)
    and vector indices (Johnson et al., 2021)

> The **Intro to NLP course**
> ➢ Elective course for the Bachelor and Master programs
>   of the Faculty of Computer Science (TH Augsburg, summer 2023)
> ➢ Alternation of frontal teaching, practical exercises and student presentations
> ➢ STS application presented at Unit 6 of 12
>   (after n-gram language modeling, text classification and word embeddings)

➢ **Goals:**
  ➢ showing the significance of the previous units (FNNs for language modelling and classification)
  ➢ stimulate the students' curiosity for coming units
  ➢ present a topic that the students could explore on their own
    (code shared for this purpose at https://github.com/saurabhkumar/lecture1_semantic_similarity)
 ➢ provide visualizations for an intuitive understanding of the results and discuss possible solutions

➢ **Example data:** Wikipedia texts on countries, capitals, language, economics, demographics, and cuisine
  ➢ examples at increasing levels of complexity, to gradually show the contribution of different techniques

Example Sentences, Methods and Visualizations

Sentence Set 1
S1: *Paris is the capital of France*
S2: *Berlin is the capital of Germany*
S3: *French is a Romance language of the Indo-European family*
S4: *German is an Indo-European language which belongs to the West Germanic group of Germanic languages*
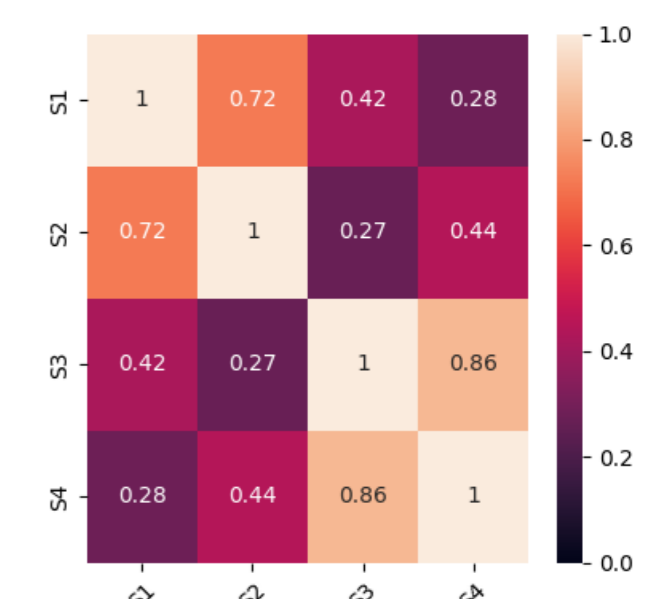
Method 1: cosine similarities between sentence vectors
(average of all normalized word vectors)
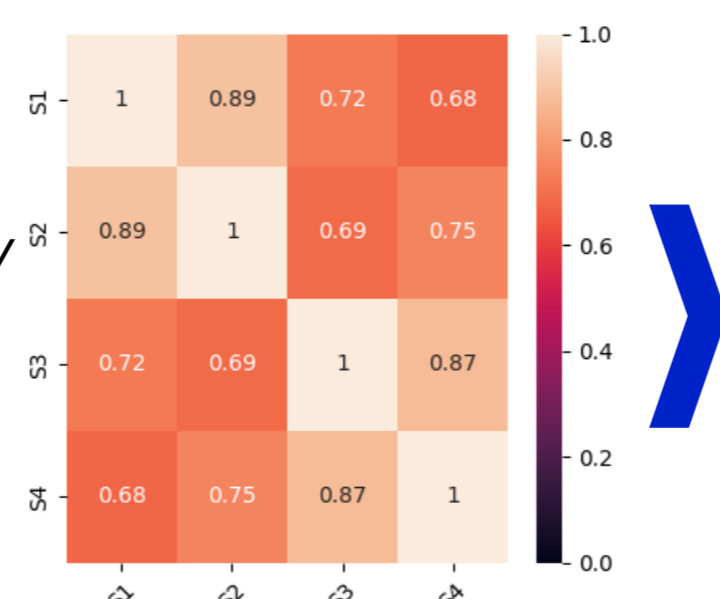
Sentence Set 2
S1: *Paris capital France*
S2: *Berlin capital Germany*
S3: *French language*
S4: *German language*

Method 2: "just the important words"

Sentence Set 3
S1: *France has a developed high-income mixed economy characterised by sizeable government involvement economic diversity a skilled labour force and high innovation. For roughly two centuries the French economy has consistently ranked among the ten largest globally*
S2: *Germany is a federal, parliamentary, representative democratic republic. Federal legislative power is vested in the parliament consisting of the Bundestag (Federal Diet) and Bundesrat (Federal Council), which together form the legislative body.*
S3: *With a population of 80.2 million according to the 2011 German Census, rising to 83.7 million as of 2022, Germany is the most populous country in the European Union, the second most populous country in Europe after Russia, and the nineteenth-most populous country in the world.*
S4: *Each region of France has traditional specialties: cassoulet in the Southwest, choucroute in Alsace, quiche in the Lorraine region, beef bourguignon in Burgundy, provençal tapenade, etc.*

Method 3: "stopword removal"

Sentence Set 4
S1 to S4 (same as Set 3)
S5: *Each region of France has traditional specialties: cassoulet in the Southwest, choucroute in Alsace, quiche in the Lorraine region, beef bourgui... Burgundy, provençal tapenade, etc.*
S6: *A typical French Christmas dish is turkey with chestnuts.*

Method 4: better embeddings with Sentence Transformers (*Reimers and Gurevych, 2019*) – as a sneak peek into the future units
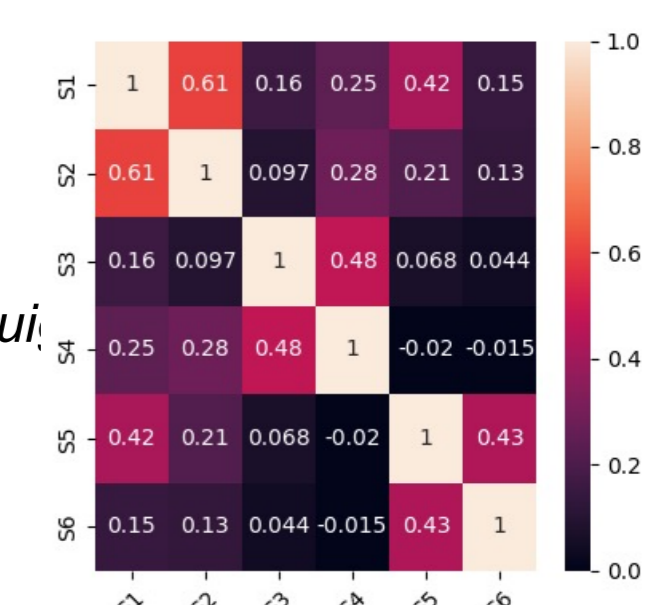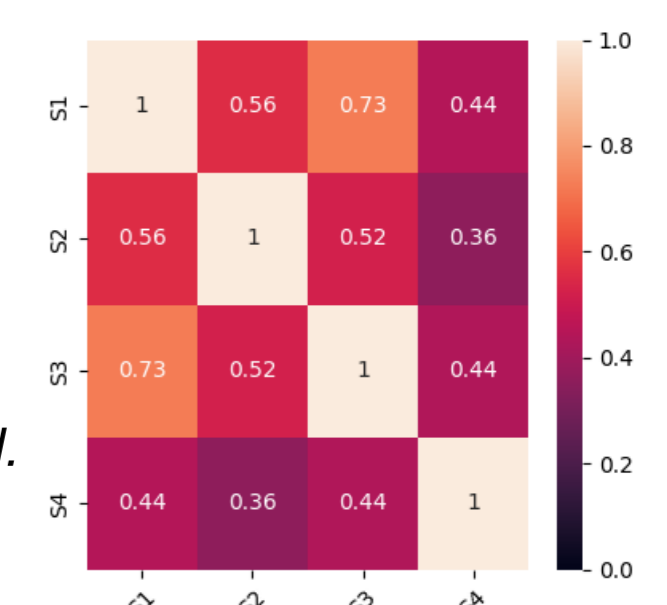
Evaluation and Outlook

Anonymized feedback from 11 students on the session
➢ wish for more application-oriented presentations
➢ appreciation for the availability of the script
➢ session rated as relevant and interesting
➢ 3 students also explicitly mentioned the session in the general course evaluation

Follow-up Session on Instruction Finetuning for LLMs
➢ use-case: QA on information from the website of TH Augsburg
➢ demo within the existing constraints on computational power
➢ code and slides for the second session available at:
  https://github.com/saurabhkumar/instruction_tuned_llm