ASSIGNMENT 1: EXPLORING WORD EMBEDDINGS

Juhao Liang, Shunian Chen, Xidong Wang, Zhengyang Tang, Benyou Wang[*]
The teaching team of Natural Language Processing
The Chinese University of Hong Kong, Shenzhen
Shenzhen, China
{zhengyangtang, xidongwang1}@link.cuhk.edu.cn, wangbenyou@cuhk.edu.cn

1 INTRODUCTION

Assignment 1 is structured into three main tasks, aimed at deepening your practical understanding of word embeddings:

- **Task 1: Train Word Embeddings with Word2Vec** (50 points)
- **Task 2: Explore Word Embeddings** (30 points)
- **Task 3: Utilize Word Embeddings** (20 points)

In **Task 1**, you will engage in the hands-on implementation and training of Word2Vec models to generate word embeddings. **Task 2-3** will guide you on how to use word embeddings effectively.

2 ASSIGNMENT REQUIREMENTS

This assignment requires using a Google Colab notebook provided at the following link:

```
https://colab.research.google.com/drive/1sxPeegJ_-aQNM-RO_
_FutZ5WLZUdb0uu?usp=sharing
```

YOU ARE REQUIRED TO:

1. **Make a copy of the provided Google Colab file.** First, you need to make a copy of the provided file into your own Google Drive. To accomplish this, open the Colab file link, navigate to File → Save a copy in Drive.
2. **Execute the notebook to generate results.** You can click on "Connect to GPU" to apply for a free T4 GPU. Then, you can press the large play button to run a code cell.
3. **Complete the Necessary Parts.** Some sections of the code are incomplete and require your input, especially pay attention to the parts marked with **[task]**. These sections are critical for scoring the assignment.

For more detailed instructions, refer to Working with Google Colab.

3 SUBMISSION GUIDELINES

Follow these steps to submit your assignment:

1. **Export the Notebook:** Navigate to File → Download .ipynb to download your notebook.
2. **Upload Your File:** Access the Blackboard system and upload your .ipynb file.

4 DEADLINE

Important: The assignment must be submitted by 11:59 pm October 10th, 2025. Late submissions may be subject to penalties.

---

[*]Benyou Wang is the instructor.