

Практичне завдання №5: Тестування промптів та класифікація тексту

Мета: Навчитись порівнювати ефективність Zero-Shot та Few-Shot промптів на практичних завданнях; отримати навички створення спеціалізованих текстових класифікаторів за допомогою техніки Few-Shot.

Ключові концепції: A/B тестування, Zero-Shot prompting, Few-Shot prompting, консистентність, класифікація тексту.

Завдання 1: A/B Тест (вилучення теми з новин)

Ідея: Наочно порівняти точність та стабільність (консистентність) результатів двох підходів на одному наборі даних.

Завдання:

1. Виконайте завдання з вилучення ключової теми для кожного з наведених заголовків двома способами: спочатку за допомогою простого Zero-Shot запиту, а потім — з використанням Few-Shot запиту з прикладами.
2. Занотуйте результати в таблицю для подальшого порівняння.

Вхідні дані (заголовки новин):

1. *"В Україні зібрали рекордний врожай зернових"*
2. *"SpaceX успішно запустила нову партію супутників Starlink"*
3. *"Нацбанк знову підвищив облікову ставку"*
4. *"Збірна України з футболу перемогла у відбірковому матчі"*
5. *"У прокат виходить новий фільм відомого українського режисера"*
6. *"Парламент ухвалив закон про цифровізацію державних послуг"*

Крок А (Zero-Shot промпт):

Визнач ключову тему (одним-двома словами) для цього заголовка:

"[вставити заголовок]"

(Запустіть цей промпт для кожного з 6 заголовків і запишіть відповіді моделі.)

Крок В (Few-Shot промпт):

Ти – аналізатор новин. Твоє завдання – визначити ключову тему заголовка. Ось кілька прикладів:

Заголовок: "Apple представила новий iPhone 17"

Тема: Технології

Заголовок: "У Києві відбувся благодійний марафон"

Тема: Спорт

Заголовок: "Ціни на нерухомість у Львові зросли на 15%"

Тема: Економіка

Тепер визнач тему для цього заголовка: "[вставити заголовок]"

(Запустіть цей промпт для кожного з 6 заголовків і також запишіть відповіді.)

Оцініть:

1. **Точність:** Який підхід дав більш релевантні та точні теми?
2. **Консистентність:** Наскільки однаковими були формулювання тем у Few-Shot підході порівняно з Zero-Shot? (напр., чи завжди модель використовувала слово "Економіка", а не "Фінанси" чи "Бізнес" для схожих тем?).
3. **Неоднозначність:** Чи були заголовки, де Zero-Shot підхід помилився, а Few-Shot впорався? Чому так сталося?

Завдання 2: Створення власного Few-Shot класифікатора

Ідея: Сконструювати Few-Shot промпт для вирішення практичної задачі класифікації на основі власних прикладів.

Задача: Розробити промпт-класифікатор для сортування вхідних звернень від студентів до деканату.

Завдання:

1. Придумайте та напишіть по 2 приклади запитів для кожної з трьох категорій:
 - Фінансові питання (оплата, стипендія, знижки)
 - Навчальні питання (розклад, предмети, перездача)
 - Адміністративні питання (довідки, гуртожиток, документи)
2. Сконструйте один великий Few-Shot промпт, який включає ваші приклади та інструкцію для моделі.
3. Перевірте роботу класифікатора на трьох нових тестових запитах.

Промпт (приклад):

Твоє завдання – класифікувати звернення студентів на три категорії: "Фінансові питання", "Навчальні питання", "Адміністративні питання".

Ось приклади:

[Ваш приклад 1 для фінансових питань] -> Фінансові питання

[Ваш приклад 2 для фінансових питань] -> Фінансові питання

[Ваш приклад 1 для навчальних питань] -> Навчальні питання

[Ваш приклад 2 для навчальних питань] -> Навчальні питання

[Ваш приклад 1 для адміністративних питань] -> Адміністративні питання

[Ваш приклад 2 для адміністративних питань] -> Адміністративні питання

Тепер класифікуй це звернення:

"[Тестове звернення]"

Тестові запити (для перевірки):

1. "Добрий день, де я можу отримати довідку про те, що я є студентом?"
2. "Підкажіть, будь ласка, до якого числа потрібно сплатити за навчання за наступний семестр?"
3. "Я пропустив контрольну роботу через хворобу. Як я можу її перездати?"

Оцініть:

1. Чи правильно модель класифікувала всі три тестові запити?
2. Наскільки важливим є якісний підбір прикладів для навчання моделі?
Що було б, якби приклади були невдалими або неоднозначними?

Завдання 3 (для СРС): Розширення та дослідження «складних» випадків

Ідея: Дослідити, як Few-Shot промпти справляються з додаванням нових, більш тонких та неоднозначних категорій. Розвинути навички ітеративного покращення промпту.

Завдання:

1. Візьміть за основу класифікатор відгуків на заклади харчування, який може визначати тональність (Позитивна, Негативна, Нейтральна). Його можна знайти в матеріалах лекції (слайд 16, «Few-Shot (для більш складного завдання)»).

2. Завдання — розширити його, додавши дві нові, більш складні категорії:
 - Порівняння (коли у відгуку порівнюють цей заклад з іншим, напр., "Тут смачніше, ніж у закладі X, але дорожче").
 - Запитання (коли відгук насправді є запитанням до закладу, напр., "А у вас є доставка?").
3. Створіть новий Few-Shot промпт:
 - Напишіть по 2-3 якісних приклади для кожної з 5 категорій (Позитивна, Негативна, Нейтральна, Порівняння, Запитання).
 - Сконструйте промпт, що включає всі ці приклади.
4. Проведіть тестування:
 - Придумайте 5 нових тестових відгуків (по одному на кожну категорію).
 - Перевірте, як ваш класифікатор справляється з ними.

Оцініть і напишіть у звіт:

1. Наведіть ваш фінальний промпт.
2. Наведіть тестові дані та результати, які ви отримали.
3. Проаналізуйте:
 - З якими труднощами ви зіткнулися при "навчанні" моделі новим категоріям?
 - Які приклади виявилися найефективнішими?
 - Чи довелося вам кілька разів переписувати промпт та приклади для досягнення бажаного результату? Опишіть цей ітеративний процес.