Comparative Analysis of Zero-Shot Learning and Supervised Learning for Spam Detection

A Study Using GPT-4 and LSTM Models

@LLMsLab

Agenda

- Introduction
- Dataset Analysis
- Zero-Shot In-Context Learning
- Few-Shot In-Context Learning
- Supervised Learning
- Comparative Analysis
- Improvement Strategies
- Conclusion
- Questions and Discussions

Introduction

- Our task revolves around the SMS Spam Collection Dataset, a public set of SMS labeled messages collected for mobile phone spam research.
- The aim is to create models that can accurately classify new, unseen messages into these categories.
- We face the issue of class imbalance, the number of 'ham' messages significantly outweighs the number of 'spam' messages.
- Our aim is to compare the efficacy of zero-shot learning, few-shot learning, and traditional supervised learning techniques in solving this text classification problem.

Dataset Analysis

- Dataset: Consists of SMS messages, manually labeled as either 'spam' or 'ham'.
- Balance: Dataset is imbalanced with 'ham' messages significantly outnumbering 'spam' messages.
- Performance Metrics: Precision, Recall, and F1 Score.

```
|label    |message                                         |
|---------|------------------------------------------------|
|ham      |Go until jurong point, crazy.. Available only ...     |
```

Zero-Shot In-Context Learning Using GPT-4 Model

- Concept: Model makes predictions on unseen data categories without explicit training.
- Implementation: Utilized GPT-4 model with specific prompts based on SMS content.
- Balancing Strategy: Randomly sampled equal number of 'spam' and 'ham' messages for the test set.

	Precision	**Recall**	**F1 Score**
GPT-4	0.92	0.97	0.95

Zero-Shot In-Context Learning (Prompt Example)

For the GPT-4 model, we used a specific structure for the prompt. The prompt is as follows:

```
prompt = (
    "Question: Given the following content, is it of type A or type B?\n"
    f"Content: {content}\n"
    "Answer Choices: (A) Type A, (B) Type B."
)
```

In our case, 'Type A' corresponds to 'spam' and 'Type B' corresponds to 'ham'. `{content}` represents the SMS message we want the model to classify.

Few-Shot In-Context Learning with GPT-4

- Concept: Model makes predictions after seeing only a few examples.
- Implementation: Utilized GPT-4 model with specific prompts and a few examples of the task in the conversation history.
- Balancing Strategy: Similar to zero-shot learning, equal number of 'spam' and 'ham' messages sampled for the test set.

	Precision	Recall	F1 Score
Result	-	-	-

Few-Shot In-Context Learning with GPT-4 (Prompt Example)

For the GPT-4 model in the few-shot in-context learning approach, we used the following prompt structure:

```
prompt = (
    "I am a model trained to identify spam and non-spam emails. Here are some examples of my training:\n"
    "Example 1:\n"
    "Question: Given the following email content, is the email spam or ham?\n"
    "Email Content: 'You have won a lottery! Claim your prize now.'\n"
    "Answer: Spam\n"
    "...\n"
    "Now, a new example to classify:\n"
    "Question: Given the following email content, is the email spam or ham?\n"
    f"Email Content: {msg}\n"
)
```

We provide a few examples of training instances, where each example consists of a question asking whether the given email content is spam or ham, along with the corresponding answer.

Supervised Learning using LSTM

- LSTM: A type of recurrent neural network, useful for sequence prediction problems.
- Preprocessing: Text data converted to sequences of numerical values.
- Architecture: Includes an Embedding layer, an LSTM layer, and a Dense output layer.
- Class Imbalance: Despite the class imbalance in the dataset, the LSTM model achieved high precision, recall, and F1 Score.

	Precision	Recall	F1 Score
LSTM	0.97	0.96	0.97

Comparative Analysis

- In-Context Learning with GPT-4: Simplifies the process, but model is a black-box with limited interpretability.
- Supervised Learning using LSTM: More control over the model's parameters, but requires more effort in preprocessing the text data and designing the model architecture.
- Class Imbalance: LSTM model performed well despite the class imbalance in the dataset.

Model Type	Precision	Recall	F1 Score
GPT-4 (Zero-Shot)	0.92	0.97	0.95
LSTM	0.97	0.96	0.97

Improvement Strategies

- Choice of method depends on specific requirements and constraints, such as the availability of labeled data and computational resources.
- For models affected by class imbalance, strategies such as resampling the data, assigning class weights, or using different evaluation metrics can be applied.

Conclusion

- Both in-context learning with GPT-4 and supervised learning with LSTM have shown to be effective for the text classification task.
- LSTM model demonstrated slightly superior performance, highlighting the value of targeted, supervised learning.
- GPT-4 performed remarkably well without explicit training, underscoring the power of zero-shot learning.

Questions and Discussions

- Open for questions and further discussions on the topic.