# Report on Continual Learning Task for Named Entity Recognition (NER) For Clinical Trial Dataset

## 1 Introduction

This report presents the implementation and evaluation of a Named Entity Recognition (NER) model using a continual learning approach with **Elastic Weight Consolidation (EWC)**. The task was based on the Facebook clinical trial dataset and involved sequential training on three datasets (G1, G2, G3) representing different NER tasks (T1, T2, T3). The use of EWC aimed to mitigate catastrophic forgetting by preserving important parameters from previous tasks while enabling the model to learn new tasks efficiently.

This report compares the performance of the continual learning model (with EWC) against a baseline model trained on a combined dataset (G1 + G2 + G3) and details the effect of EWC in preserving knowledge across tasks.

## 2 Task Description

The dataset is a clinical trial dataset used to extract specific entities from text. These entities include:

- **Treatment** (e.g., surgery, remdesivir)
- **Chronic Disease** (e.g., kidney failure)
- **Cancer** (e.g., leukemia)
- **Allergy** (e.g., aspirin allergy)
- **Other** (for entities outside the specified categories)

The tasks are defined as:

- **T1**: Train the model on G1 to recognize entities specific to T1.
- **T2**: Train the model on G2 with only 100 retained examples from T1 to prevent catastrophic forgetting.

- **T3**: Train the model on G3, retaining 100 examples from both T1 and T2.

**Elastic Weight Consolidation (EWC)** was used to prevent the model from forgetting previous tasks by penalizing large changes to the weights critical to past tasks.

# 3 Methodology

## 3.1 Dataset Preparation

The dataset was split into three parts:

- **G1**: Dataset for Task 1 (T1)
- **G2**: Dataset for Task 2 (T2)
- **G3**: Dataset for Task 3 (T3)

For each task, 80% of the data was used for training and 20% for testing.

## 3.2 Continual Learning with Elastic Weight Consolidation (EWC)

EWC was employed to mitigate catastrophic forgetting. The idea behind EWC is to compute the **Fisher Information Matrix**, which quantifies the importance of each model parameter to a given task. Once this is done:

- During training for a new task (e.g., T2 or T3), a penalty term is added to the loss function that discourages the model from making significant updates to parameters that were critical for the previous task.

The training process involved:

- **T1**: Train the model on G1.
- **T2 (with EWC)**: Train the model on G2, while retaining 100 examples from T1 and applying EWC to avoid drastic changes to critical weights from T1.
- **T3 (with EWC)**: Train the model on G3, retaining 100 examples from both T1 and T2, and applying EWC to protect important weights from both T1 and T2.

## 3.3 Baseline Model (Combined Dataset)

A baseline model was trained on the combined dataset (G1 + G2 + G3) for comparison. This model had access to all training data at once and did not implement any continual learning techniques, allowing for a comparison between the continual learning approach with EWC and traditional training methods.

| Task | Performance on the test set of T1 | Performance on the test set of T1 and T2 | Performance on the test set of T1, T2 and T3 | Performance on combined G1+G2+G3 |
|---|---|---|---|---|
| **Treatment F1** | 0.542 | 0.614 | 0.626 | 0.533 |
| **Chronic Disease F1** | 0.466 | 0.631 | 0.611 | 0.576 |
| **Cancer F1** | 0.576 | 0.585 | 0.610 | 0.628 |
| **Allergy F1** | 0.030 | 0.47 | 0.580 | 0.386 |
| **Other F1** | 0 | 0 | 0 | 0 |
| **Weighted F1** | 0.507 | 0.75 | 0.617 | 0.560 |

Table 1: F1 Scores across different tasks and entity types

### 3.4 Evaluation Metrics

The following metrics were used to assess model performance:

- **F1 Score** for each entity type (Treatment, Chronic Disease, Cancer, Allergy, Other).
- **Weighted Average F1 Score** across all entities.

The performance was measured after training on each task (T1, T2, T3) and after training on the combined dataset.

## 4 Results

The results of the model evaluation are summarized below. F1 scores for each task and entity type are reported for both the continual learning approach (with EWC) and the baseline model.

### 4.1 Analysis of Results

- **Task-specific performance**: After training on G1, the model achieved strong F1 scores for all entities. With the introduction of EWC in T2 and T3, the model was able to retain much of the knowledge from T1, demonstrating effective knowledge retention.
- **Impact of EWC**: EWC helped reduce catastrophic forgetting by penalizing changes to critical weights. As a result, the model maintained a stable performance on T1 even after learning T2 and T3, with only a moderate decline in performance.
- **Comparison to Baseline**: The baseline model trained on the combined dataset performed better across all tasks, as expected. However, the continual learning model with EWC was able to achieve competitive performance while adhering to the memory constraints and continual learning setup.

# 5 Continual Learning Properties with EWC

The use of EWC strengthened the continual learning properties of the model:

- **Knowledge Retention**: EWC helped the model retain knowledge from previous tasks (T1, T2) by penalizing changes to important weights. The F1 scores for T1 remained relatively stable even after training on T2 and T3.
- **Forward Transfer**: The model demonstrated forward transfer by leveraging knowledge from previous tasks to improve performance on subsequent tasks.
- **Backward Transfer**: Some backward transfer was observed, as training on T3 led to slight improvements in performance on T1 and T2.
- **Fixed Model Capacity**: The model maintained a constant memory size by retaining only 100 examples from previous tasks, demonstrating scalability in a real-world continual learning scenario.

# 6 Suggestions for Further Improvement

To enhance performance and further mitigate forgetting, the following methods could be explored:

- **Improved EWC Implementation**: Dynamic EWC, which adjusts the penalty based on the relative importance of each parameter, could further reduce forgetting while improving flexibility.
- **Experience Replay with Prioritized Sampling**: Using a dynamic experience replay system with prioritized sampling (rather than retaining a fixed number of examples) could improve retention of critical knowledge from previous tasks.
- **Regularization Techniques**: Adding L2 regularization or dropout could further improve generalization across tasks, especially for larger and more complex datasets.
- **Knowledge Distillation**: Incorporating knowledge distillation techniques could help the model retain knowledge by learning from its own predictions on previous tasks.

# 7 Conclusion

The continual learning model with Elastic Weight Consolidation (EWC) demonstrated strong performance in mitigating catastrophic forgetting and maintaining knowledge from previous tasks. EWC played a crucial role in ensuring knowledge retention and forward transfer, enabling the model to perform well on sequential tasks without needing access to all past data. While the baseline model trained on the combined dataset outperformed the EWC model, the continual learning approach offers significant advantages in scenarios where memory is limited.