**Comparison of Structural SVM (SSVM), Conditional Random Fields (CRF), and Maximum Margin Markov Networks (MMMN) performance on OCR dataset**

Maria Mercedes Balderas

COGS 185

Spring 2025

Abstract

We evaluated several structure prediction models on an OCR sequence dataset using structured learning techniques such as sliding windows on Structural SVM, Conditional Random Fields, and Maximum Margin Markov Networks. The data consists of sequences of characters that have been grouped into words. We evaluate and compare both character-level accuracy and word-level accuracy. Results demonstrate that the sliding window technique greatly aids in prediction accuracy, especially at the character level prediction.

**Introduction**

Optical Character Recognition is a well studied task in computer vision and language processing, especially when characters are a part of a sequence (words). Sequential structures are able to provide contextual cues that improve classification accuracy

The purpose of this report is to explore and compare the performance of different algorithms, in this case Structural SVM (SSVM), Conditional Random Fields (CRF), and Maximum Margin Markov Networks (MMMN) in a structured prediction task on an OCR (Optical Character Recognition) dataset.

**Method**

To benchmark these three algorithms, they will all be applied to the same OCR dataset and compared using the same window size and training/test size split.

Due to how the data was formatted (characters of each word separated), there was some pre-processing needed to be done on the dataset to group each character with the rest of the characters to which word they belonged to, as well as a need to sort them in order to restore the full word. The dataset was then truncated to 5000 words to be used as training and testing data. In order to compare the three different algorithms, I tested them across many different window sizes and training/test splits. The training/test splits being: 1000/4000, 2000/2000, 2500/2500, 3000/2000, 4000/1000; within each of these splits each algorithm was also tested with the window size L which could be of 2, 3, 5, and 7.

A sliding window of size L was used for feature extraction, where each window was centered around each character to generate a sort of local context based on its L surrounding characters. This method was used for all three algorithms- SSVM, CRF, and MMMN.

To evaluate the accuracy of each model we did both a character level accuracy test, and a word level accuracy test. At the character level accuracy it is determined by how many characters in the word are predicted right based on the context received by the sliding window, at the word level accuracy it is predicted by how many entire words get predicted right (in other words how often the model gets an entire sequence (in this case a word) right).

**Experiment**

The results are divided for each of the training splits into tables, where among the columns are the varying window sizes and across the rows are the varying algorithms. Within each result the top number is the character level accuracy and the bottom one the word level accuracy.

1000/4000 training/test split

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Window size | 2 | 3 | 5 | 7 |
| SSM | 0.57  0.089 | 0.57  0.089 | 0.61  0.1 | 0.64  0.14 |
| CRF | 0.78  0.40 | 0.78  0.40 | 0.83  0.52 | 0.85  0.57 |
| MMMN | 0.78  0.35 | 0.78  0.35 | 0.79  0.38 | 0.81  0.44 |

2000/3000 training/test split

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Window size | 2 | 3 | 5 | 7 |
| SSM | 0.60  0.11 | 0.60  0.11 | 0.65  0.16 | 0.69  0.21 |
| CRF | 0.81  0.47 | 0.81  0.47 | 0.85  0.56 | 0.88  0.64 |
| MMMN | 0.80  0.40 | 0.80  0.40 | 0.83  0.45 | 0.85  0.52 |

2500/2500 training/test split

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Window size | 2 | 3 | 5 | 7 |
| SSM | 0.61  0.13 | 0.61  0.13 | 0.66  0.17 | 0.70  0.225 |
| CRF | 0.83  0.50 | 0.83  0.50 | 0.86  0.59 | 0.88  0.65 |
| MMMN | 0.80  0.40 | 0.82  0.42 | 0.84  0.47 | 0.86  0.54 |

3000/2000 training/test split

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Window size | 2 | 3 | 5 | 7 |
| SSM | 0.62  0.13 | 0.62  0.13 | 0.67  0.18 | 0.71  0.24 |
| CRF | 0.83  0.51 | 0.83  0.51 | 0.87  0.61 | 0.89  0.67 |
| MMMN | 0.82  0.43 | 0.82  0.43 | 0.84  0.48 | 0.86  0.55 |

4000/1000 training/test split

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Window size | 2 | 3 | 5 | 7 |
| SSM | 0.64  0.15 | 0.64  0.15 | 0.67  0.20 | 0.71  0.27 |
| CRF | 0.84  0.52 | 0.84  0.53 | 0.87  0.61 | 0.89  0.69 |
| MMMN | 0.83  0.45 | 0.83  0.45 | 0.85  0.50 | 0.87  0.56 |

**Conclusion**

As we can see, Conditional Random Fields was the algorithm with the highest accuracy test, both at the character and word level; followed by the Maximum Margin Markov Network, and lastly by the Structural SVM. This regardless of window size and test training splits

Project github:

**References**