




# Contents

<b>Certificate</b>	<b>ii</b>
<b>Acknowledgments</b>	<b>v</b>
<b>Preface</b>	<b>vii</b>
<b>Contents</b>	<b>xi</b>
<b>List of Figures</b>	<b>xiv</b>
<b>List of Tables</b>	<b>xvi</b>
<b>Abbreviations</b>	<b>xviii</b>
<b>Symbols</b>	<b>xx</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Research Objectives . . . . .	4
1.2 Scope . . . . .	5
1.3 Contributions . . . . .	6
1.4 Chapterization . . . . .	7
<b>2 Backgrounds and Literature survey</b>	<b>9</b>
2.1 Low Resource Languages . . . . .	9
2.2 Traditional Learning Techniques in NLP . . . . .	11
2.2.1 Trigrams ‘n’ Tags . . . . .	11
2.2.2 Maximum Entropy Markov Model . . . . .	11
2.2.3 Conditional Random Fields . . . . .	12
2.2.4 Structured Support Vector Machine . . . . .	13
2.3 Deep Learning Models in NLP . . . . .	14
2.3.1 Convolutional Neural Network . . . . .	14
2.3.2 Recurrent Neural Network . . . . .	15
2.3.3 Attention Mechanisms . . . . .	19

2.4	Transfer Learning . . . . .	22
2.5	Distributional Vector Representation . . . . .	24
2.5.1	Context Independent Embedding . . . . .	25
2.5.2	Context Dependent Embedding . . . . .	28
2.6	Hand Crafted Features in Neural Network . . . . .	31
2.7	Literature Survey on Sequence Labeling . . . . .	31
2.7.1	Sequence Labeling Work on Indian Languages . . . . .	32
2.7.2	Sequence Labeling Models with Feature Engineering . . . . .	37
<b>3</b>	<b>Baseline Experiments on Low Resource Languages for Sequence Labeling</b>	<b>41</b>
3.1	Introduction . . . . .	41
3.2	Contributions of Chapter . . . . .	44
3.3	Methodology . . . . .	45
3.4	Experiment-I: POS Tagging and Chunking . . . . .	48
3.4.1	Dataset Description . . . . .	48
3.4.2	Machine Learning Strategy . . . . .	49
3.4.3	Deep Learning Strategy . . . . .	50
3.4.3.1	Feature transfer . . . . .	50
3.4.4	Results and Analysis . . . . .	51
3.4.4.1	Error analysis . . . . .	58
3.5	Experiment-II: NER . . . . .	63
3.5.1	Dataset Description . . . . .	63
3.5.2	Experimental Settings . . . . .	65
3.5.2.1	CRF model training . . . . .	65
3.5.2.2	Deep learning model training . . . . .	66
3.5.3	Results and Analysis . . . . .	66
3.5.3.1	Effect of epoch in LSTM-CNNs-CRF model . . . . .	70
3.5.3.2	Error analysis . . . . .	70
3.6	Summary . . . . .	74
<b>4</b>	<b>Leveraging Contextual Information for Sequential Labeling</b>	<b>75</b>
4.1	Introduction . . . . .	75
4.2	Contributions of Chapter . . . . .	76
4.3	Methodology . . . . .	77
4.3.1	Input Encoder . . . . .	77
4.3.2	Contextual Encoder . . . . .	81
4.3.3	Disambiguator . . . . .	83
4.4	Inclusion of Morphological Features . . . . .	84
4.5	Contrastive Training Component . . . . .	86
4.6	Experiment-I: Feature Integration in MCCGS . . . . .	88
4.6.1	Dataset Description . . . . .	90
4.6.2	Experimental Settings . . . . .	91
4.6.3	Results and Analysis . . . . .	92

4.6.3.1	Feature inclusion . . . . .	93
4.6.3.2	Effect of chunk size . . . . .	96
4.6.3.3	Experiments on Indo-Aryan languages . . . . .	97
4.6.3.4	Analysis . . . . .	97
4.7	Experiment-II: Contrastive MCCGS . . . . .	99
4.7.1	Domain Adaptation . . . . .	101
4.7.2	Dataset Description . . . . .	102
4.7.3	Experimental Settings . . . . .	103
4.7.4	Results and Analysis . . . . .	103
4.7.4.1	Single-source domain adaptation . . . . .	105
4.7.4.2	Multi-source domain adaptation . . . . .	107
4.7.4.3	Effect of layers freezing in domain adaptation . . . . .	107
4.8	Summary . . . . .	109
<b>5</b>	<b>Multilingual Learning based Sequence Labeling</b>	<b>111</b>
5.1	Introduction . . . . .	111
5.2	Contribution of Chapter . . . . .	113
5.3	Methodology . . . . .	113
5.3.1	Residual BiLSTM-CRF Model . . . . .	113
5.3.2	Meta Learning Algorithms . . . . .	116
5.3.2.1	Model agnostic meta learning . . . . .	117
5.3.2.2	Reptile . . . . .	118
5.4	Dataset Description . . . . .	119
5.5	Experimental Details . . . . .	119
5.6	Results and Discussion . . . . .	121
5.6.1	Results for Few Shot POS Tagging . . . . .	121
5.6.2	Results for Low Resource POS Tagging . . . . .	122
5.6.3	Ablation Study . . . . .	124
5.6.4	Analysis . . . . .	124
5.7	Summary . . . . .	126
<b>6</b>	<b>Conclusions and Future Scopes</b>	<b>127</b>
6.1	Proposed Methods and Findings . . . . .	127
6.2	Future Scopes . . . . .	129
<b>A</b>	<b>List of Publications</b>	<b>132</b>
	<b>Bibliography</b>	<b>134</b>

# List of Figures

2.1	CNN over the text [171]	15
2.2	Folded RNN with feedback loop to a input sequence	16
2.3	Unfolded RNN to a input sequence ( $x$ ) along with its time steps ( $t$ )	16
2.4	LSTM block at $t^{th}$ time step [98]	17
2.5	GRU block at $t^{th}$ time step	18
2.6	Seq2seq model with encoder and decoder to translate the input sequence $ABC$ to $WXYZ$	19
2.7	Bahdanau’s attention model to a encoder-decoder based model	21
2.8	Vaswani et al. [242] transformer architecture which comprise the encoder-decoder	22
2.9	Transfer learning taxonomy for NLP	23
2.10	The continuous skip-gram model	27
2.11	The continuous bag-of-words model	27
2.12	ELMo model architecture for pre-training representation	29
2.13	BERT model architecture for pre-training representation	30
3.1	SAHBiLC model architecture for POS Tagging	46
3.2	F-score result of POS tagging for (a) Bhojpuri, (b) Maithili and (c) Magahi	53
3.3	F-score results of Chunk tagging for (a) Bhojpuri and (b) Maithili	54
3.4	F-scores of POS tagging after applying SAHBiLC model and Fine-SAHBiLC model on (a) Bhojpuri and (b) Maithili	56
3.5	Most affected POS tags after applying SAHBiLC model and Fine-SAHBiLC model on (a) Bhojpuri, (b) Maithili and (c) Magahi	56
3.6	The F-scores of chunk tagging after applying SAHBiLC model and Fine-SAHBiLC model on (a) Bhojpuri and (b) Maithili	57
3.7	Effect of epochs on LSTM-CNNs-CRF model’s accuracy on Bhojpuri, Maithili and Magahi.	70
3.8	Confusion matrix for Bhojpuri for the LSTM-CNNs-CRF (above) and CRF (below) models; The  refers to correctly prediction	71
3.9	Confusion matrix for Maithili for the LSTM-CNNs-CRF (above) and CRF (below) models; The  refers to correctly prediction	72
3.10	Confusion matrix for Magahi for the LSTM-CNNs-CRF (above) and CRF (below) models; The  refers to correctly prediction	73
4.1	Overview of the MCCGS architecture with its components	78
4.2	Character level CNN model architecture to generate word vector	80

---

4.3	MCCGS model architecture with symbolic feature inclusion . . . . .	86
4.4	Impact of morphological features and longer contextual dependencies . . . . .	88
4.5	Comparison of mean accuracy scores . . . . .	96
4.6	Example of models prediction to POS tagging . . . . .	99
4.7	Domain adaptation . . . . .	101
4.8	Accuracy comparison of SOTA and CMCCGS models . . . . .	104
4.9	Accuracy comparison of the CMCCGS model with (part(i) and part(ii)) and without (none) freezing layers' . . . . .	110
5.1	Residual BiLSTM-CRF model for POS tagging . . . . .	114
5.2	Meta learning based training for POS tagging . . . . .	116
5.3	Sanskrit POS tagger's performance as a function of the number of Hindi training examples used in the meta pretraining stage . . . . .	125
5.4	Marathi POS tagger's performance as a function of the number of Hindi training examples used in the meta pretraining stage . . . . .	125

# List of Tables

3.1	Annotation statistics of POS tagging datasets . . . . .	48
3.2	Annotation statistics of Chunking datasets . . . . .	48
3.3	Feature set used for POS Tagging and Chunking in machine learning techniques and contextual boundary value is upto 3, represented by $j$ . . . . .	49
3.4	Parameters and Hyper-parameters employed during training the SAHBiLC model . . . . .	51
3.5	Results of traditional machine learning techniques (%) in terms of Accuracy, Precision, Recall and F-score for POS Tagging . . . . .	52
3.6	Results for traditional machine learning techniques in terms of Accuracy, Precision, Recall and F-score for Chunking . . . . .	53
3.7	Results for Deep Learning techniques in terms of Accuracy, Precision, Recall and F-score for POS Tagging . . . . .	55
3.8	Results for Deep Learning techniques in terms of Accuracy, Precision, Recall and F-score for Chunking . . . . .	55
3.9	Language-wise dataset statistics used for annotation of named entities . . . . .	63
3.10	The statistics of annotated for the three broad categories for Bhojpurī, Maithilī and Magahī . . . . .	64
3.11	The statistics of annotated hierarchical entities for <b>Bhojpurī</b> , <b>Maithilī</b> and <b>Magahī</b> . The ENAMEX, NUMEX and TIMEX categories contained 11, 4 and 7 hierarchical named entities. ‘Other’ denotes regular words or tokens which are not named entities. . . . .	64
3.12	The dataset sizes for each language. The OOV percentage is calculated by token-type differences between test data and the training data. . . . .	65
3.13	The value of (hyper-)parameters used for training of the LSTM-CNNs-CRF model . . . . .	66
3.14	NER tag-wise scores obtained by CRF and LSTM-CNNs-CRF for Bhojpurī. The metrics, which are <b>Precision</b> , <b>Recall</b> and <b>F<sub>1</sub>-score</b> . . . . .	67
3.15	NER tag-wise scores obtained for CRF and LSTM-CNNs-CRF for Maithilī . . . . .	68
3.16	NER tag-wise scores obtained for CRF and LSTM-CNNs-CRF for Magahī . . . . .	69
4.1	The languages with their statistics, obtained from the UD treebank. Some languages have more than one treebank; hence the related treebank information is mentioned after - in the language name. . . . .	91
4.2	Obtained results using the MCCGS model . . . . .	93

---

4.3	Obtained scores after inclusion of morphological features in the MCCGS model. Here, IE, CE and D stand for MCCGS-IE, MCCGS-CE and MCCGS-D models respectively. . . . .	94
4.4	Effect of the chunk size on the model performance in terms of mean accuracy	96
4.5	Obtained scores on the Indo-Aryan languages of UD treebank and HUTB dataset . . . . .	98
4.6	Dataset statistics for each domain . . . . .	103
4.7	Results obtained on each domain of HT dataset . . . . .	105
4.8	Results obtained on PTB, TweepBank and ARK datasets . . . . .	105
4.9	Domain adaptation on the HT dataset . . . . .	106
4.10	Domain adaptation on ARK and TweepBank datasets . . . . .	106
4.11	Results of Multi-source domain adaptation by CMCCGS model . . . . .	108
4.12	Results of the domain adaptation based on freezing layers' by CMCCGS model . . . . .	109
5.1	Data statistics . . . . .	119
5.2	Results for few shot POS tagging . . . . .	122
5.3	Results for low resource POS tagging . . . . .	122
6.1	Mapping of chapters with their associated research questions . . . . .	128