

Table of Content

1	Introduction	1
1.1	Problem Statement	2
1.2	Research Hypothesis	3
1.3	Objectives	3
1.3.1	General Objectives	3
1.3.2	Specific Objectives	4
1.4	Methodology	4
1.5	Research Outcome	5
1.6	Outline	5
2	Background and Related Work	7
2.1	Scientific Disciplines	7
2.1.1	Artificial Intelligence	7
2.1.2	Machine Learning	8
2.1.3	Deep Learning and Feedforward Neural Network	9
2.1.4	Natural Language Processing	10
2.1.5	Incremental and Streaming Learning	12
2.1.6	Instance and Batch Incremental Learning	13
2.2	Word Representation	14
2.2.1	One Hot Representation	14
2.2.2	Distributional Hypothesis and Distributional Representations	15

2.2.3	Word Context Matrices	15
2.2.4	Distributed Representation or Word Embeddings	20
2.2.5	Other methods	26
2.3	Intrinsic NLP Tasks	27
2.4	Streaming in Word Embedding models	28
2.5	Related Work	29
2.5.1	Incremental Word Embedding Models	29
2.5.2	Stream Machine Learning Libraries	30
2.5.3	Intrinsic Evaluation	31
3	RiverText Foundations	32
3.1	Misra Greis Algorithm	33
3.2	Incremental Learning Approaches	34
3.3	Periodic Evaluation	35
3.4	Implemented Methods	37
3.4.1	Incremental Word Context Matrix	37
3.4.2	Incremental Word2Vec	38
4	Experiments and Results	41
4.1	Data	41
4.2	Experimental setup	41
4.2.1	Hyperparameter settings	42
4.2.2	Results and discussion	43
5	RiverText Library	49
5.1	Motivation	50
5.2	Components	50
5.2.1	Word Embedding Model	50
5.2.2	Periodic Evaluation	58

5.2.3	Utils	58
5.2.4	Training Process	60
6	Conclusion and Future Work	64
6.1	Conclusion	64
6.2	Future Work	65
	Bibliography	72
	Annexes	73
	Annex A Experiment Results by Task and Model	73
	Annex B Time serie plots	82