Appendix A: Hyperparameters

In this appendix are the range of hyperparameters that were optimized, along with the used hyperparameters for all the provided models.

Table 6. All hyperparameters and their ranges during optimization of LCR-Rot-hop-ont++.

Hyperparameter	Range
Learning Rate (LR)	$\in \{0.005, 0.001, 0.01, 0.02, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1\}$
Dropout Rate (DR)	$\in \{0.3, 0.4, 0.5, 0.6, 0.7\}$
Momentum (Mom.)	$\in \{0.85, 0.9, 0.95, 0.99\}$
Weight Decay (WD)	$\in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
Rotatory hops (RH)	$\in \{3\}$
γ	$\in \{-1.5, -0.9, -0.8, \dots, 0.9, 1.5\}$

Table 7. All hyperparameters and their ranges during optimization of LCR-Rot-hop-Kfont++.

Hyperparameter	Range
Learning Rate (LR)	$\in \{0.01, 0.02, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1\}$
Dropout Rate (DR)	$\in \{0.3, 0.4, 0.5, 0.6, 0.7\}$
Momentum (Mom.)	$\in \{0.85, 0.9, 0.95, 0.99\}$
Weight Decay (WD)	$\in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
Rotatory Hops (RH)	$\in \{3\}$
γ	$\in \{-1.5, -0.9, -0.8, \dots, 0.9, 1.5\}$

Table 8. Hyperparameters used in training LCR-Rot-hop-ont++ with different knowledge injections for the SemEval 2014 restaurant dataset.

Knowledge	Regime	Max. evals	Hyperparameters					
			LR	DR	Mom.	WD	RH	γ
None	-	10	0.06	0.4	0.9	10^{-5}	3	-
0-hop	Training	10	0.1	0.5	0.85	10^{-5}	3	0.8
	Testing	10	0.02	0.5	0.9	10^{-4}	3	-0.4
	Both	10	0.07	0.4	0.95	10^{-4}	3	1.5
1-hop	Training	10	0.02	0.5	0.9	0.01	3	1.1
	Testing	10	0.05	0.7	0.9	10^{-4}	3	1.0
	Both	10	0.06	0.5	0.95	10^{-5}	3	1.1
2-hop	Training	10	0.07	0.4	0.95	10^{-4}	3	0.3
	Testing	10	0.005	0.3	0.85	10^{-5}	3	0.2
	Both	10	0.05	0.6	0.9	10^{-4}	3	1.2

Table 9. Hyperparameters used in training LCR-Rot-hop-ont++ with different knowledge injections for the SemEval 2015 restaurant dataset.

Knowledge	Regime	Max. evals	Hyperparameters					
			LR	DR	Mom.	WD	RH	γ
None	-	10	0.01	0.4	0.85	10^{-5}	3	-
0-hop	Training	10	0.07	0.7	0.95	10^{-5}	3	-1.5
	Testing	10	0.1	0.3	0.95	10^{-3}	3	0.5
	Both	10	0.06	0.4	0.95	10^{-5}	3	1.0
1-hop	Training	10	0.001	0.3	0.99	10^{-3}	3	0.9
	Testing	10	0.005	0.4	0.99	10^{-3}	3	-1.3
	Both	10	0.07	0.3	0.95	10^{-5}	3	1.4
2-hop	Training	10	0.09	0.7	0.95	10^{-4}	3	-0.7
	Testing	10	0.05	0.4	0.85	10^{-4}	3	-0.8
	Both	10	0.02	0.7	0.95	10^{-3}	3	-1.3

Table 10. Hyperparameters used in training LCR-Rot-hop-ont++ with different knowledge injections for the SemEval 2016 Restaurant dataset.

Knowledge	Regime	Max. evals	Hyperparameters					
			LR	DR	Mom.	WD	RH	γ
None	-	10	0.05	0.5	0.85	10^{-5}	3	-
0-hop	Training	10	0.02	0.3	0.9	10^{-4}	3	0.7
	Testing	10	0.08	0.6	0.85	10^{-4}	3	-0.5
	Both	10	0.08	0.6	0.85	10^{-3}	3	0.0
1-hop	Training	10	0.09	0.6	0.99	10^{-4}	3	-0.1
	Testing	10	0.08	0.4	0.85	10^{-2}	3	-0.7
	Both	10	0.01	0.6	0.9	10^{-5}	3	-1.4
2-hop	Training	10	0.06	0.5	0.9	10^{-2}	3	1.4
	Testing	10	0.01	0.7	0.95	10^{-2}	3	0.3
	Both	10	0.10	0.6	0.85	10^{-4}	3	-1.0

Table 11. Hyperparameters used in training LCR-Rot-hop-ont++ with different knowledge injections for the SemEval 2014 laptop dataset.

Knowledge	Regime	Max. evals	Hyperparameters					
			LR	DR	Mom.	WD	RH	γ
None	-	10	0.05	0.5	0.95	10^{-3}	3	-
0-hop	Training	10	0.07	0.7	0.95	10^{-5}	3	-1.5
	Testing	10	0.2	0.3	0.95	10^{-3}	3	0.5
	Both	10	0.06	0.4	0.95	10^{-5}	3	1.0
1-hop	Training	10	10^{-3}	0.3	0.99	10^{-3}	3	0.9
	Testing	10	0.005	0.4	0.99	10^{-3}	3	-1.3
	Both	10	0.07	0.3	0.95	10^{-5}	3	1.4
2-hop	Training	10	0.09	0.7	0.95	10^{-4}	3	-0.7
	Testing	10	0.05	0.4	0.85	10^{-4}	3	-0.8
	Both	10	0.02	0.7	0.95	10^{-3}	3	-1.3

For LCR-Rot-hop-Kfont++, for 0-hop, γ is not included in the parameter search range. Some models were trained multiple times, and the following tables record the parameter for the first time.

Table 12. Hyperparameters used in training LCR-Rot-hop-Kfont++ with different knowledge injections for the SemEval 2015 restaurant dataset.

Knowledge	Regime	Max. evals	Hyperparameters					
			LR	DR	Mom.	WD	RH	γ
None	-	14	0.05	0.4	0.95	0.01	3	-
0-hop	Training	14	0.05	0.7	0.85	0.001	3	-
	Testing	14	0.1	0.7	0.9	0.0001	3	-
	Both	14	0.05	0.6	0.99	0.001	3	-
1-hop	Training	20	0.05	0.6	0.99	0.0001	3	-1.0
	Testing	20	0.1	0.5	0.95	0.00001	3	-0.5
	Both	20	0.09	0.5	0.95	0.00001	3	-0.5
2-hop	Training	20	0.07	0.4	0.9	0.01	3	1.5
	Testing	20	0.1	0.7	0.85	0.01	3	-1.5
	Both	20	0.07	0.6	0.99	0.01	3	-1.5

Table 13. Hyperparameters used in training LCR-Rot-hop-Kfont++ with different knowledge injections for the SemEval 2014 laptop dataset.

Knowledge	Regime	Max. evals	Hyperparameters					
			LR	DR	Mom.	WD	RH	γ
None	-	14	0.1	0.4	0.99	0.01	3	-
0-hop	Training	14	0.07	0.4	0.99	0.0001	3	-
	Testing	20	0.07	0.5	0.99	0.001	3	-
	Both	14	0.05	0.6	0.9	0.001	3	-
1-hop	Training	20	0.1	0.6	0.9	0.0001	3	1.5
	Testing	20	0.08	0.4	0.99	0.001	3	-1.0
	Both	20	0.1	0.6	0.95	0.0001	3	-0.5
2-hop	Training	20	0.09	0.3	0.95	1e-05	3	1.0
	Testing	16	0.07	0.4	0.95	0.01	3	-1.0
	Both	20	0.06	0.3	0.99	0.001	3	-1.0

Table 14. Hyperparameters used in training LCR-Rot-hop-Kfont++ with different knowledge injections for the SemEval 2016 restaurant dataset.

Knowledge	Regime	Max. evals	Hyperparameters					
			LR	DR	Mom.	WD	RH	γ
None	-	20	0.07	0.4	0.9	0.01	3	-
0-hop	Training	20	0.02	0.4	0.95	0.001	3	-
	Testing	20	0.06	0.5	0.99	0.0001	3	-
	Both	20	0.06	0.7	0.99	0.001	3	-
1-hop	Training	20	0.05	0.3	0.95	1e-05	3	1.5
	Testing	27	0.1	0.5	0.95	0.01	3	0.0
	Both	20	0.01	0.7	0.85	0.0001	3	-1.5
2-hop	Training	20	0.08	0.5	0.9	0.001	3	0.0
	Testing	20	0.08	0.5	0.95	0.01	3	1.0
	Both	20	0.06	0.5	0.85	0.0001	3	-1.5

Table 15. Hyperparameters used in training LCR-Rot-hop-Kfont++ with different knowledge injections for the SemEval 2014 restaurant dataset.

Knowledge	Regime	Max. evals	Hyperparameters					
			LR	DR	Mom.	WD	RH	γ
None	-	20	0.08	0.3	0.9	0.01	3.0	-
0-hop	Training	20	0.05	0.4	0.99	0.01	3	-
	Testing	15	0.05	0.7	0.99	0.0001	3	-
	Both	20	0.1	0.4	0.9	0.0001	3	-
1-hop	Training	20	0.05	0.6	0.99	0.01	3	0.0
	Testing	23	0.1	0.7	0.99	0.001	3	-1.0
	Both	20	0.07	0.7	0.9	0.001	3	0.0
2-hop	Training	20	0.08	0.7	0.99	0.0001	3	0.5
	Testing	20	0.1	0.6	0.99	0.01	3	-0.5
	Both	20	0.05	0.7	0.95	0.0001	3	-1.0

Note: To decrease computational demands, only a randomly selected 50% subset of the SemEval-2014 restaurant training dataset is used to optimize hyperparameters.