Recursive Neural Structural Correspondence Network for Cross-domain Aspect and Opinion Co-Extraction

2019.4.10

Problem definition: Fine-grained opinion analysis

- Extract aspect and opinion terms from each sentence for opinion summarization.

"*They offer good appetizers*"

opinion terms   aspect terms

- Denote a sentence by a sequence of tokens: $\mathbf{x} = (w_1, w_2, ..., w_n)$

- The output is a sequence of token-level labels:

$$\mathbf{y} = (y_1, y_2, ..., y_n) \qquad y_i \in \{\mathrm{BA}, \mathrm{IA}, \mathrm{BO}, \mathrm{IO}, \mathrm{N}\}$$

Problem definition: Cross-domain Fine-grained opinion analysis

- One active line focuses on projecting original feature spaces of two domains into the same low-dimensional space to reduce domain shift using pivot features as a bridge.

- Another line learns domain-invariant features via auto-encoders.

Motivation: Cross-domain Fine-grained opinion analysis

- The dependency relations.

  "*They offer <u>good</u> <u>appetizers</u>*"

  "*The laptop has a <u>nice</u> <u>screen</u>*"

  "OPINION-amod-ASPECT"

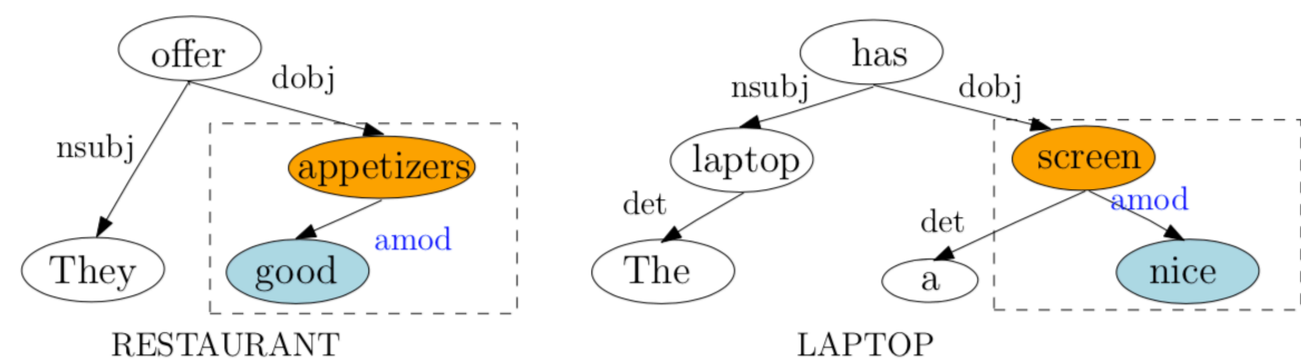


Figure 1: An example of two reviews with similar syntactic patterns.

- Structural learning : the success of target task depends on the ability of finding good predictive structures learned from other related tasks.

Model: RNSCN-GRU

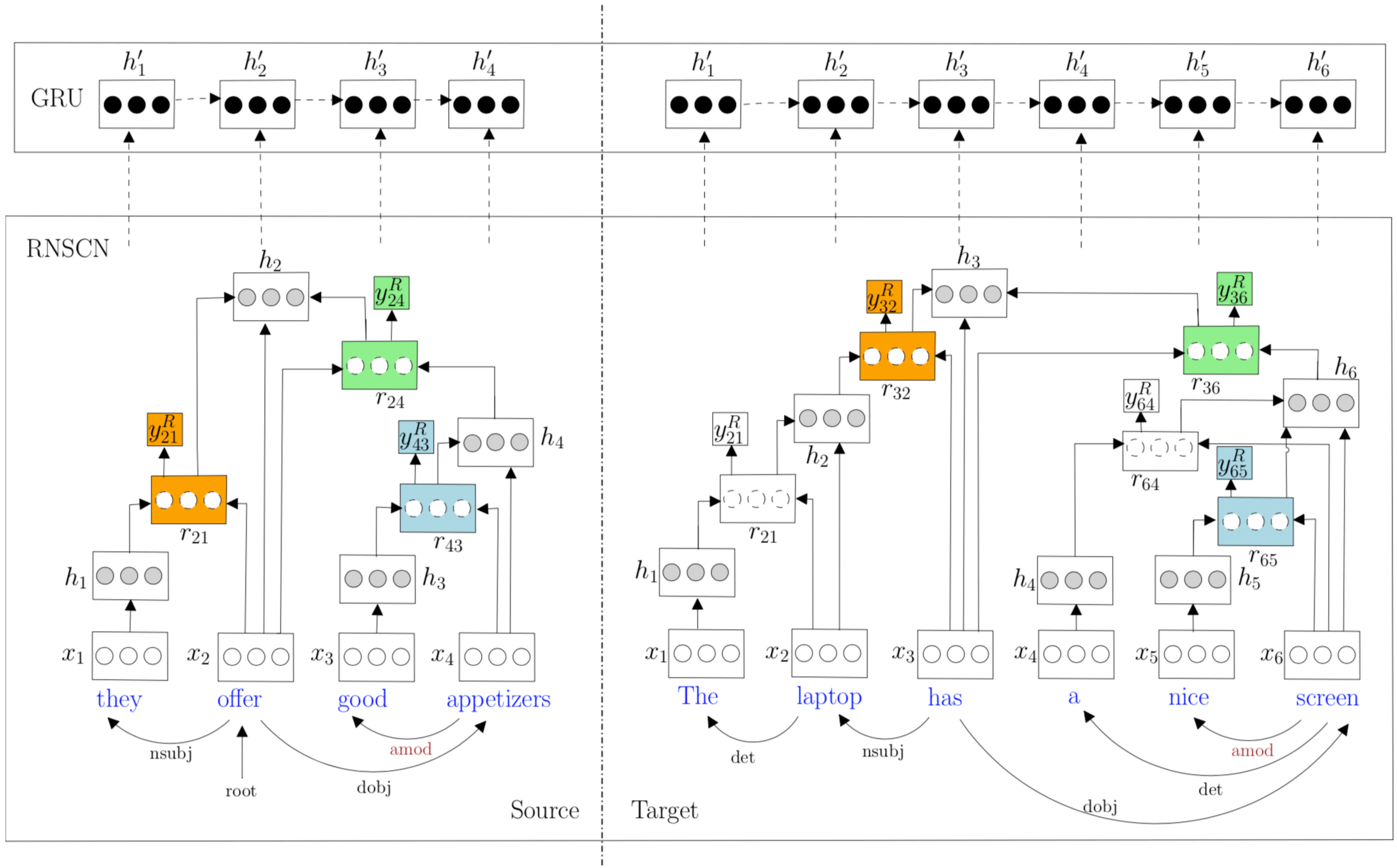
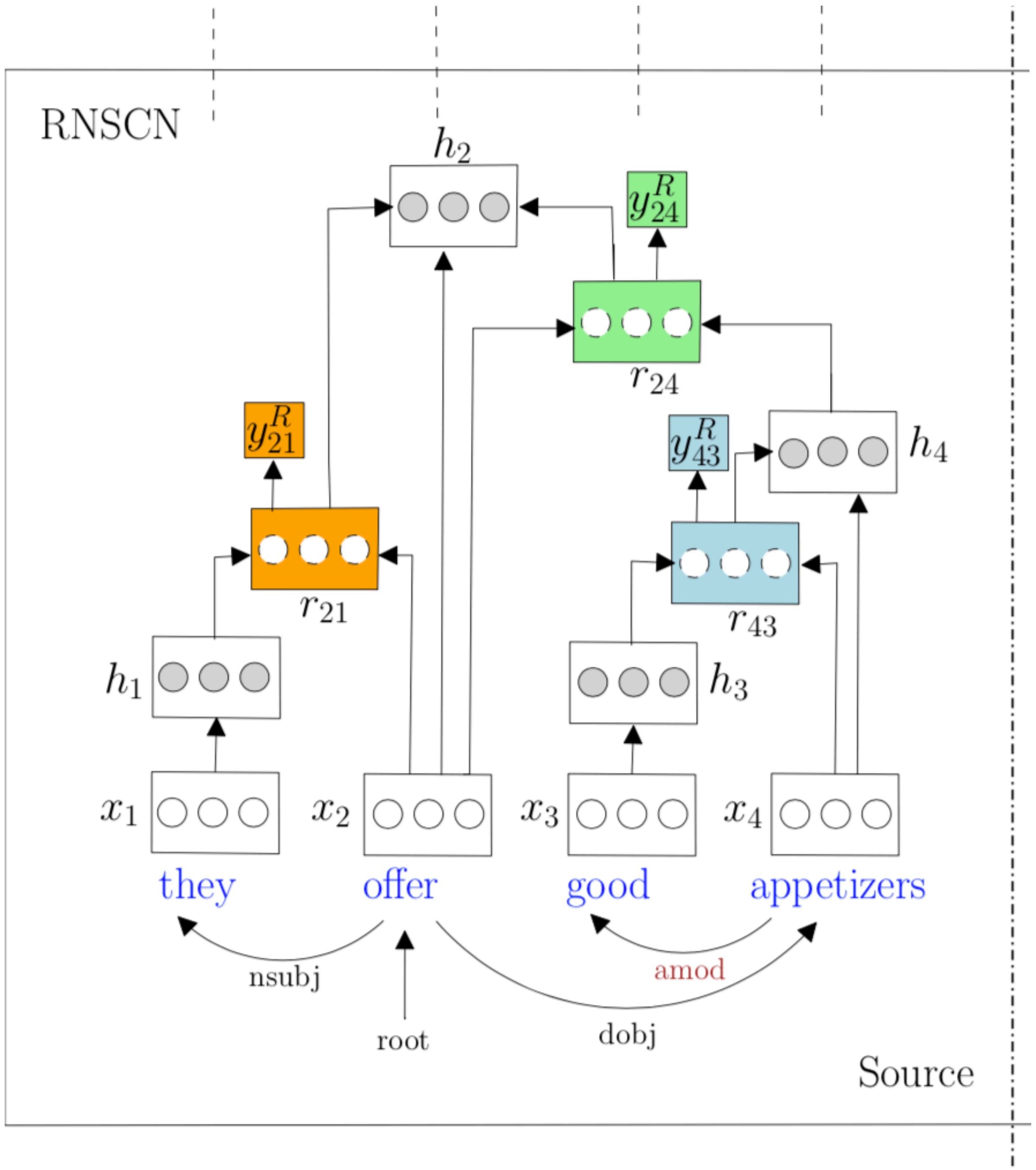


Figure 2: The architecture of RNSCN-GRU.

Model: Recursive Neural Structural Correspondence Network (RNSCN)



$$\mathbf{h}_1 = \tanh(\mathbf{W}_x \mathbf{x}_1 + \mathbf{b}), \quad \mathbf{h}_3 = \tanh(\mathbf{W}_x \mathbf{x}_3 + \mathbf{b})$$

$$\mathbf{x}_4 \text{ (appetizers)} \xrightarrow{\text{amod}} \mathbf{x}_3 \text{ (good)}$$

$$\mathbf{r}_{43} = \tanh(\mathbf{W}_h \mathbf{h}_3 + \mathbf{W}_x \mathbf{x}_4)$$

$$\hat{\mathbf{y}}^R_{43} = \text{softmax}(\mathbf{W}_R \mathbf{r}_{43} + \mathbf{b}_R)$$

$$\mathbf{h}_4 = \tanh(\mathbf{W}_{\text{amod}} \mathbf{r}_{43} + \mathbf{W}_x \mathbf{x}_4 + \mathbf{b})$$

$$\mathbf{h}_n = \tanh\Big(\sum_{m \in \mathcal{M}_n} \mathbf{W}_{R_{nm}} \mathbf{r}_{nm} + \mathbf{W}_x \mathbf{x}_n + \mathbf{b}\Big)$$

Model: Reduce Label Noise with Auto-encoders

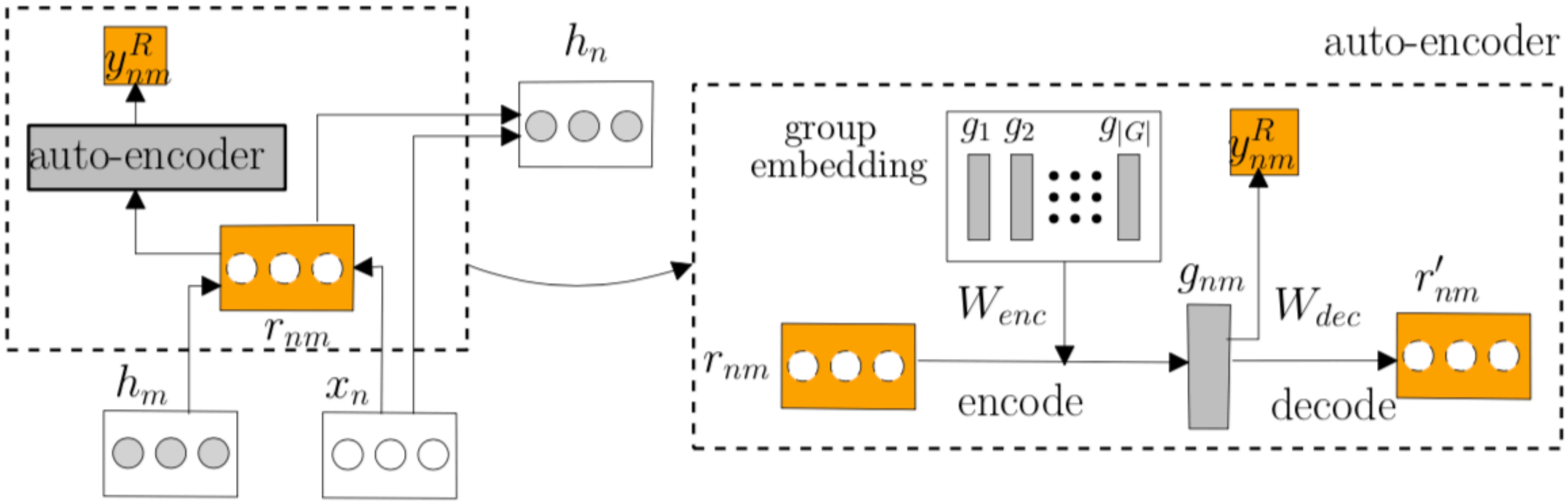


Figure 3: An autoencoder for relation grouping.

$$p(G_{nm} = i|\mathbf{r}_{nm}) = \frac{\exp(\mathbf{r}_{nm}^\top \mathbf{W}_{enc}\mathbf{g}_i)}{\sum_{j \in G} \exp(\mathbf{r}_{nm}^\top \mathbf{W}_{enc}\mathbf{g}_j)}$$

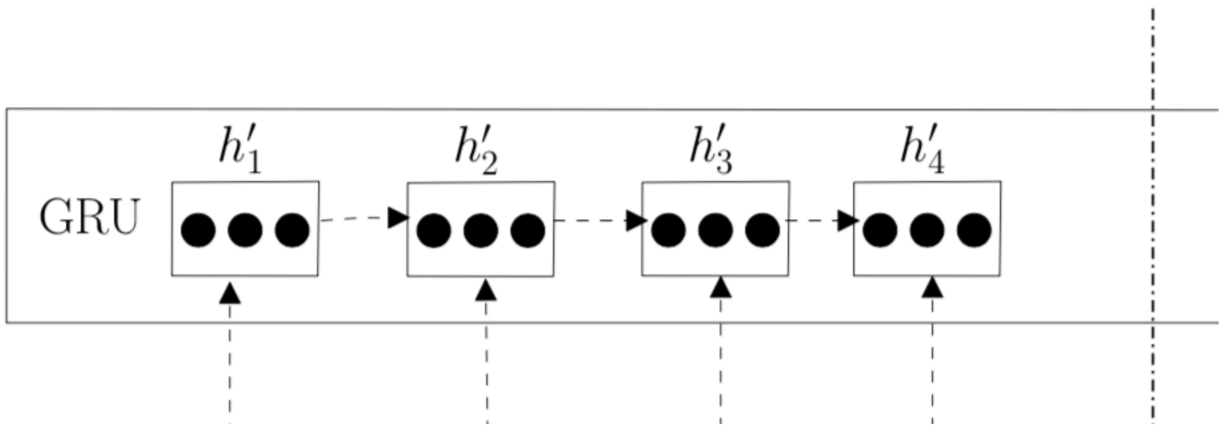
$$\hat{\mathbf{y}}^R_{nm} = \text{softmax}(\mathbf{W}_R \cdot \mathbf{r}_{nm} + \mathbf{b}_R)$$

$$\mathbf{g}_{nm} = \sum_{i=1}^{|G|} p(G_{nm} = i|\mathbf{r}_{nm})\mathbf{g}_i$$

$$\hat{\mathbf{y}}^R_{nm} = \text{softmax}(\mathbf{W}_R \mathbf{g}_{nm} + \mathbf{b}_R)$$

Model: RNSCN+GRU / RNSCN+CRF

GRU:



$$\mathbf{h}'_n = f_{GRU}(\mathbf{h}'_{n-1}, \mathbf{h}_n; \mathbf{\Theta})$$

$$\hat{\mathbf{y}}_n = \text{softmax}(\mathbf{W}_l \cdot \mathbf{h}'_n + \mathbf{b}_l)$$

CRF:

$$\begin{aligned}\hat{\mathbf{y}} &= \arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{H}) \\ &= \arg\max_{\mathbf{y}} \frac{1}{Z(\mathbf{H})} \prod_{c \in C} \exp\langle \mathbf{W}_c, g(\mathbf{H}, \mathbf{y}_c)\rangle\end{aligned}$$

$$\ell_S(\mathbf{y}_{S_i}, \hat{\mathbf{y}}_{S_i}) = -\log(\mathbf{y}_{S_i}|\mathbf{h}_{S_i}).$$

Model: Train

RNSCN+ Loss：

$$\ell_R = \ell_{R_1} + \alpha \ell_{R_2} + \beta \ell_{R_3},$$

where

$$\ell_{R_1} = \left\|\mathbf{r}_{nm} - \mathbf{W}_{dec}\mathbf{g}_{nm}\right\|_2^2,$$

Auto-encoder loss

$$\ell_{R_2} = \sum_{k=1}^{K} -\mathbf{y}_{nm[k]}^{R} \log \hat{\mathbf{y}}_{nm[k]}^{R},$$

Relation classification loss

$$\ell_{R_3} = \left\|\mathbf{I} - \bar{\mathbf{G}}^{\top}\bar{\mathbf{G}}\right\|_F^2.$$

Enforces orthogonality：$\mathbf{g}_i$ and $\mathbf{g}_j$ for $i \neq j$

RNSCN+-GRU Loss：

$$\mathcal{L} = \sum_{\mathcal{D}_S} \ell_S(\mathbf{y}_{S_i}, \hat{\mathbf{y}}_{S_i}) + \gamma \sum_{\mathcal{D}_R} \ell_R(\mathbf{r}_j, \mathbf{y}_j^R)$$

Experiments:

Models	R→L		R→D		L→R		L→D		D→R		D→L	
	AS	OP	AS	OP	AS	OP	AS	OP	AS	OP	AS	OP
CrossCRF	19.72 (1.82)	59.20 (1.34)	21.07 (0.44)	52.05 (1.67)	28.19 (0.58)	65.52 (0.89)	29.96 (1.69)	56.17 (1.49)	6.59 (0.49)	39.38 (3.06)	24.22 (2.54)	46.67 (2.43)
RAP	25.92 (2.75)	62.72 (0.49)	22.63 (0.52)	54.44 (2.20)	46.90 (1.64)	67.98 (1.05)	34.54 (0.64)	54.25 (1.65)	45.44 (1.61)	60.67 (2.15)	28.22 (2.42)	59.79 (4.18)
Hier-Joint	33.66 (1.47)	- -	33.20 (0.52)	- -	48.10 (1.45)	- -	31.25 (0.49)	- -	47.97 (0.46)	- -	34.74 (2.27)	-
RNCRF	24.26 (3.97)	60.86 (3.35)	24.31 (2.57)	51.28 (1.78)	40.88 (2.09)	66.50 (1.48)	31.52 (1.40)	55.85 (1.09)	34.59 (1.34)	63.89 (1.59)	40.59 (0.80)	60.17 (1.20)
RNGRU	24.23 (2.41)	60.65 (1.04)	20.49 (2.68)	52.28 (2.69)	39.78 (0.61)	62.99 (0.95)	32.51 (1.12)	52.24 (2.37)	38.15 (2.82)	64.21 (1.11)	39.44 (2.79)	60.85 (1.25)
RNSCN-CRF	35.26 (1.31)	61.67 (1.35)	32.00 (1.48)	52.81 (1.29)	**53.38** (1.49)	67.60 (0.99)	34.63 (1.38)	56.22 (1.10)	48.13 (0.71)	65.06 (0.66)	46.71 (1.16)	61.88 (1.52)
RNSCN-GRU	37.77 (0.45)	62.35 (1.85)	33.02 (0.58)	57.54 (1.27)	53.18 (0.75)	71.44 (0.97)	35.65 (0.77)	60.02 (0.80)	**49.62** (0.34)	69.42 (2.27)	45.92 (1.14)	63.85 (1.97)
RNSCN^{+}-GRU	**40.43** (0.96)	**65.85** (1.50)	**35.10** (0.62)	**60.17** (0.75)	52.91 (1.82)	**72.51** (1.03)	**40.42** (0.70)	**61.15** (0.60)	48.36 (1.14)	**73.75** (1.76)	**51.14** (1.68)	**71.18** (1.58)

Table 2: Comparisons with different baselines.

Experiments: Auto-encoder

Dataset	Description	# Sentences	Training	Testing
R	Restaurant	5,841	4,381	1,460
L	Laptop	3,845	2,884	961
D	Device	3,836	2,877	959

Table 1: Data statistics with number of sentences.

Models	R→L		R→D		L→R		L→D		D→R		D→L	
	AS	OP	AS	OP	AS	OP	AS	OP	AS	OP	AS	OP
RNSCN-GRU	37.77	62.35	33.02	57.54	**53.18**	71.44	35.65	60.02	**49.62**	69.42	45.92	63.85
RNSCN-GRU (r)	32.97	50.18	26.21	53.58	35.88	65.73	32.87	57.57	40.03	67.34	40.06	59.18
RNSCN^{+}-GRU	**40.43**	**65.85**	**35.10**	**60.17**	52.91	**72.51**	**40.42**	**61.15**	48.36	**73.75**	**51.14**	**71.18**
RNSCN^{+}-GRU (r)	39.27	59.41	33.42	57.24	45.79	69.96	38.21	59.12	45.36	72.84	50.45	68.05

Table 3: Comparisons with different variants of the proposed model.

Experiments: the effect of different components

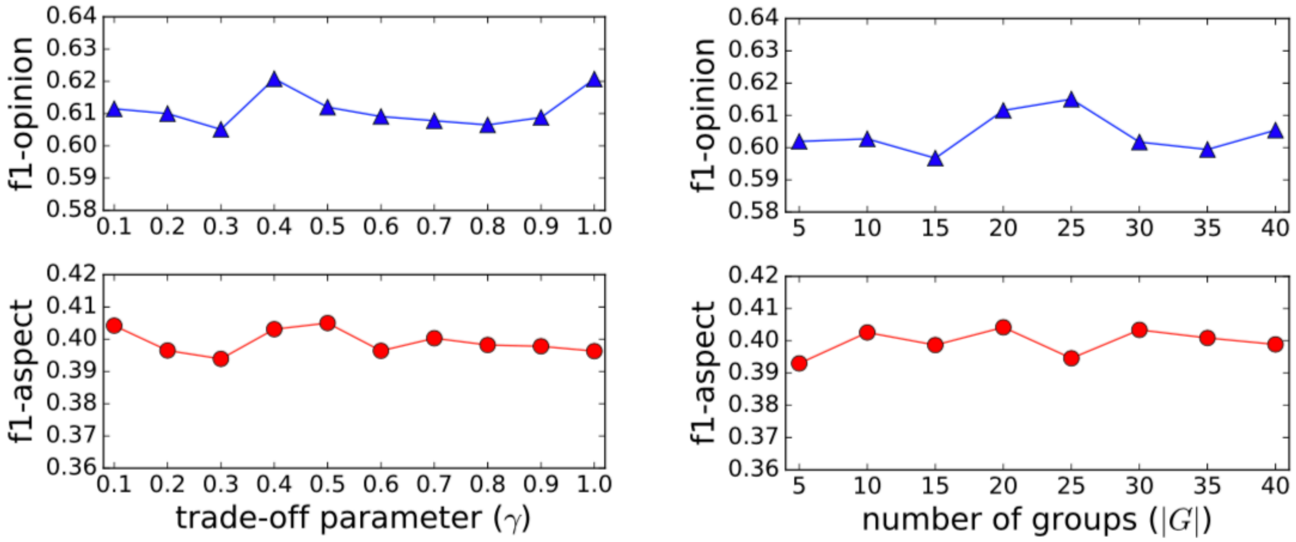
		R→L		R→D		L→R		L→D		D→R		D→L	
		AS	OP	AS	OP	AS	OP	AS	OP	AS	OP	AS	OP
OUT	Hier-Joint	33.66	-	33.20	-	48.10	-	31.25	-	47.97	-	34.74	-
	RNSCN^{+}-GRU*	39.06	-	34.07	-	47.98	-	38.51	-	47.49	-	48.49	-
	RNSCN^{+}	31.60	**65.89**	24.37	60.01	39.58	71.03	34.40	60.47	41.02	71.23	45.54	69.00
	RNSCN^{+}-GRU	**40.43**	65.85	**35.10**	**60.17**	**52.91**	**72.51**	**40.42**	**61.15**	**48.36**	**73.75**	**51.14**	**71.18**
IN	Hier-Joint	32.41	-	29.79	-	47.04	-	31.26	-	**47.41**	-	33.80	-
	RNSCN^{+}-GRU*	40.34	-	30.75	-	48.69	-	37.40	-	46.49	-	48.50	-
	RNSCN^{+}	30.76	63.65	22.48	59.24	39.54	70.25	35.32	60.00	37.75	70.64	43.72	68.27
	RNSCN^{+}-GRU	**41.27**	**65.44**	**33.58**	**60.28**	**52.48**	**72.10**	**39.73**	**60.18**	47.10	**72.19**	**50.23**	**70.21**

Table 4: Comparisons with different transfer setting.

OUT: Inductive setting

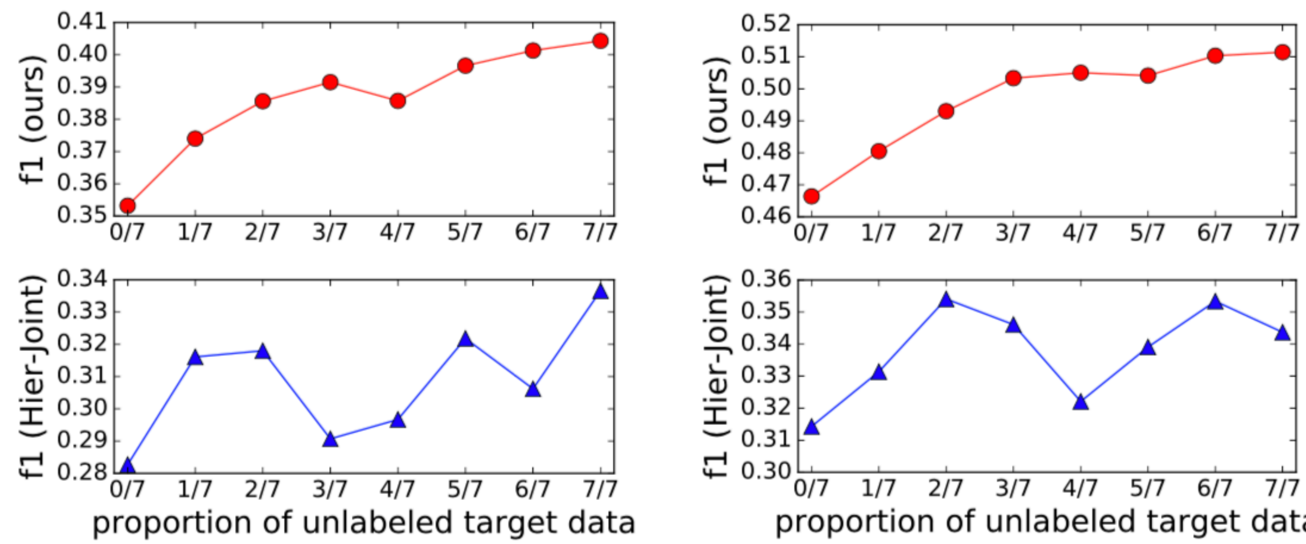
IN: Transductive setting

Experiments: robustness and capability



(a) On trade-off parameter. (b) On number of groups.

Figure 4: Sensitivity studies for L→D.



(a) F1-aspect on R→L (b) F1-aspect on D→L

Figure 5: F1 vs proportion of unlabeled target data.