Profiling News Discourse Structure Using Explicit Subtopic Structures Guided Critics

Prafulla Kumar Choubey and Ruihong Huang
Department of Computer Science and Engineering
Texas A&M University
(prafulla.choubey, huangrh)@tamu.edu

EMNLP-2021

https://github.com/prafulla77/Discoure_Profiling_RL_EMNLP21Findings

Leibniz-Institut
für Sozialwissenschaften

Reported by Dongdong Hu

Introduction

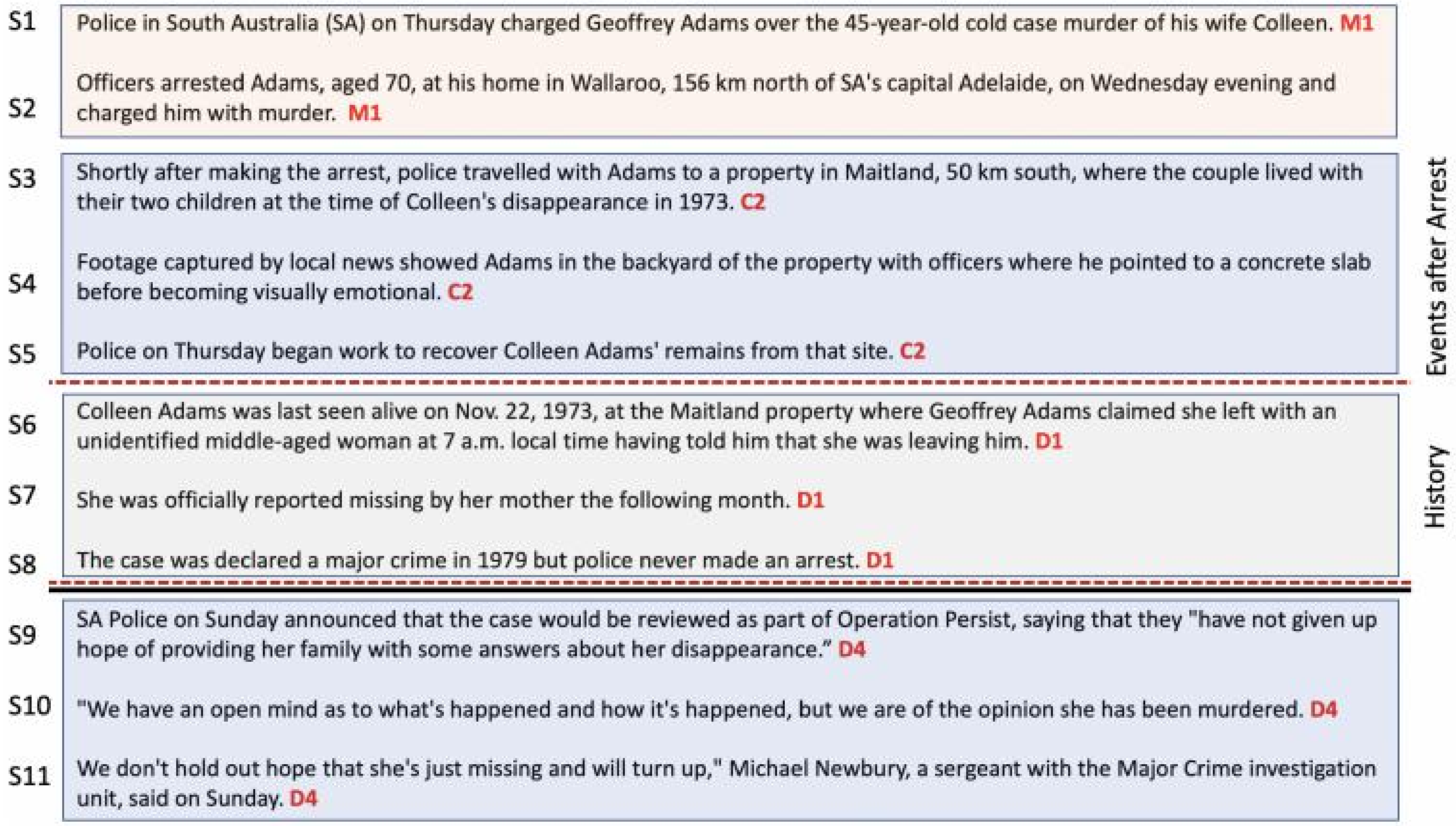

Figure 1: An example document annotated with three different subtopic structures. The first is based on TextTiling (Hearst, 1997) and is shown with the black-solid line ([S1-S8],[S9-S11]). The second structure is based on locally inverted pyramid structure (discussed in § 5.2) and is shown through red-dashed lines ([S1-S5],[S6-S8],[S9-S11]). The third, shown by colored boxes, segments document based on the temporal position where the first segment (S1, S2) focuses on the main event, second segment (S3, S4, S5) describes events following the main event, third segment (S6, S7, S8) describes historical events and the last segment (S9, S10, S11) again covers current context.

The hierarchical model, provides a mechanism for capturing both global and local dependencies among sentences and the main topic.
However, the model is completely unaware of the underlying content organization structures that are used while producing news reports.

Main Event (M1)

Consequence (M2)

Previous Events (C1)

Current Context (C2)

Historical Event (D1)

Anecdotal Event (D2)

Evaluation (D3)

Expectation (D4)

Method

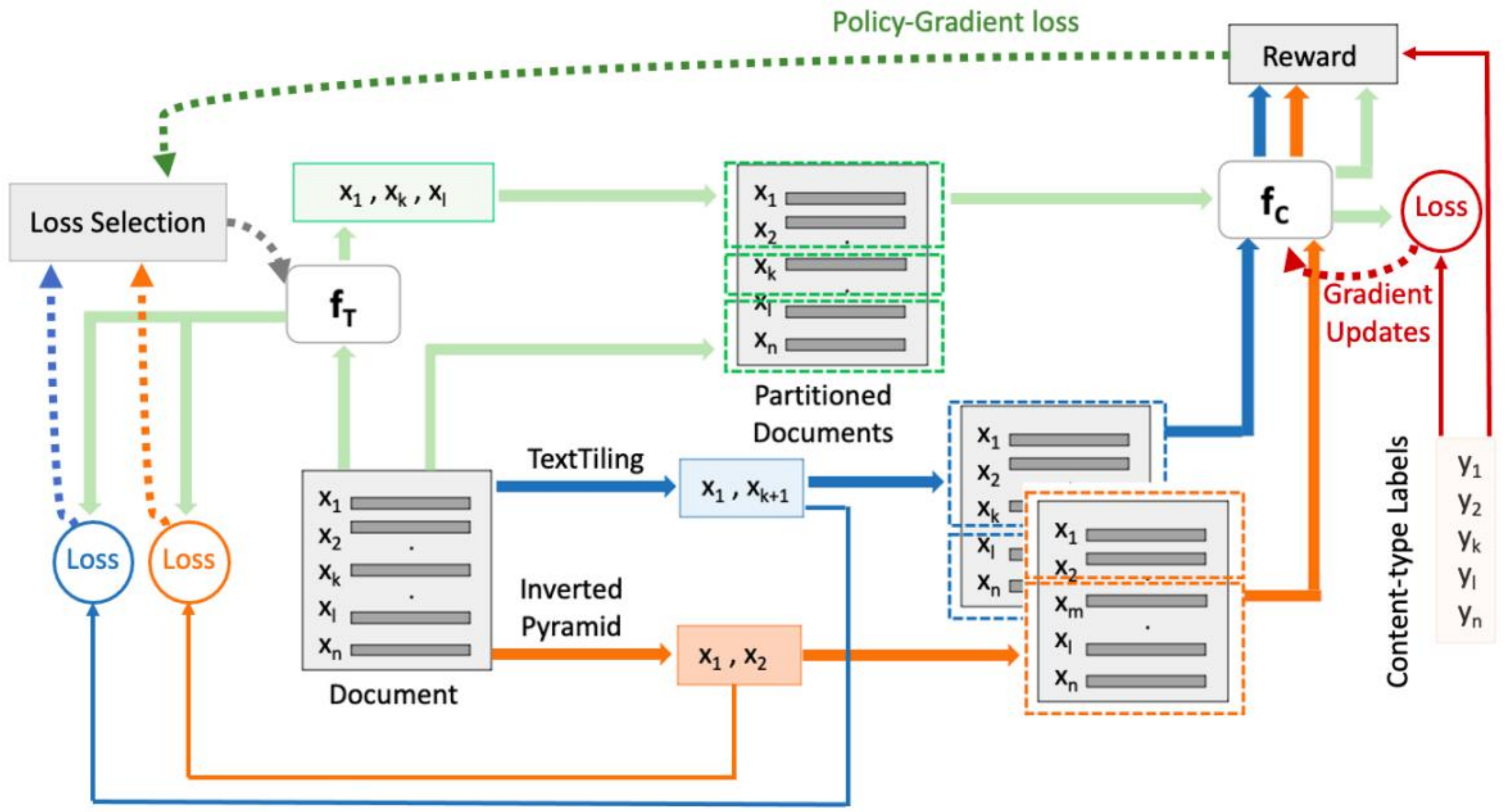

Figure 2: Neural-Network Architecture, including Gradient Flow Paths, for Incorporating Document-level Content Structures in a Discourse Profiling System

Method

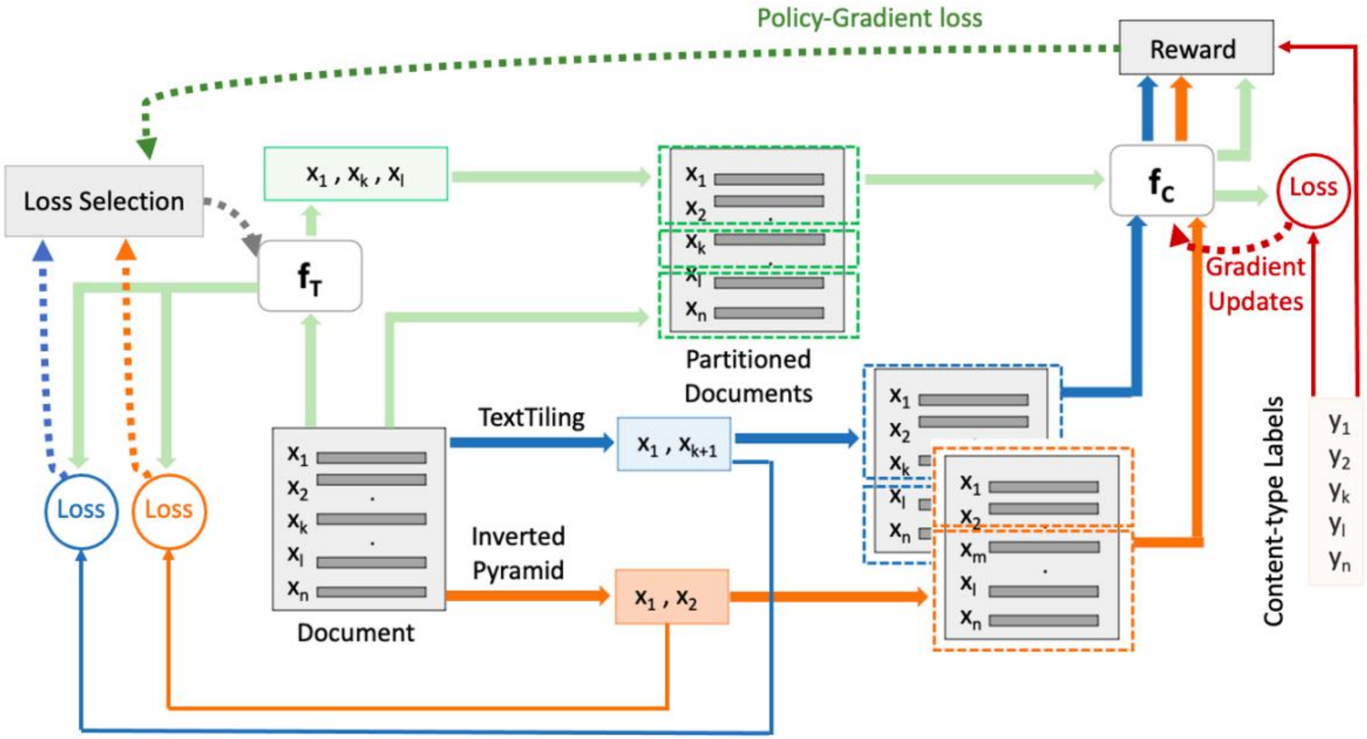

Figure 2: Neural-Network Architecture, including Gradient Flow Paths, for Incorporating Document-level Content Structures in a Discourse Profiling System

$$X \quad : \{H, x_1, x_2, .., x_n\}$$
$$x_i \ \{w_{i1}, w_{i2}, .., w_{im}\}$$
$$Y \quad : \{y_1, y_2, .., y_n\}$$

Hierarchical Sentence Encoder

$$\begin{aligned}
[E_{i1}, E_{i2}, .., E_{im}] &= ELMo([w_{i1}, w_{i2}, .., w_{im}]) \\
[H_{i1}, H_{i2}, .., H_{im}] &= biLSTM^L([E_{i1}, E_{i2}, .., E_{im}]) \\
\alpha_i[k] &= W_{\alpha 1}(tanh(W_{\alpha 2} E_{ik} + b_{\alpha 2})) + b_{\alpha 1} \in R \\
A_i &= softmax(\alpha_i) \in R^m \qquad (1) \\
S_i^L &= \sum_k A_i[k] H_{ik} \in R^{2d_{rnn}} \\
[H^C, S_1^C, .., S_n^C] &= biLSTM^C([H^L, S_1^L, .., S_n^L])
\end{aligned}$$

Method

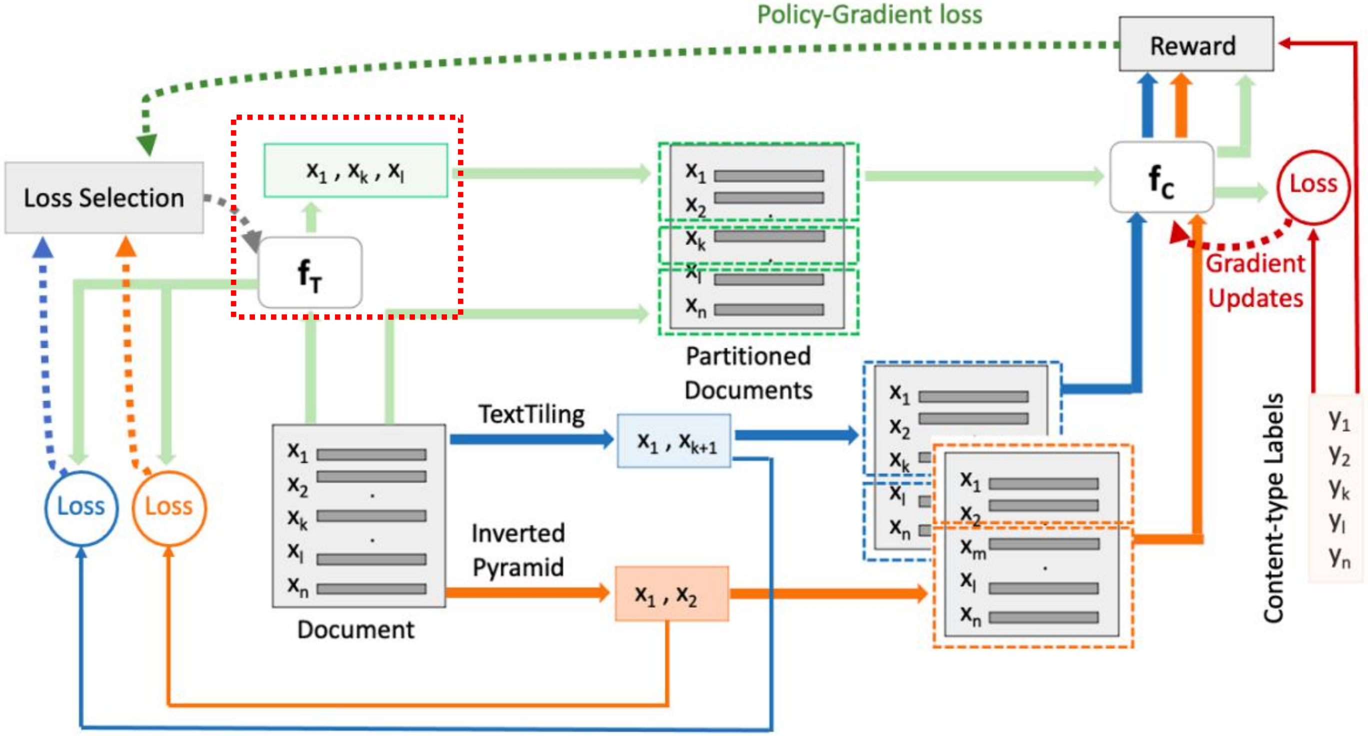

Figure 2: Neural-Network Architecture, including Gradient Flow Paths, for Incorporating Document-level Content Structures in a Discourse Profiling System

Pointer Decoder Network

$d^h_{k=1}$=D from eq. 3 S^C_1

$$
\begin{aligned}
d^h_k &= LSTMCell(d^h_{k-1}, S^C_{T_{k-1}}) \\
u^k_i &= [W^1_p(S^C_i) * W^2_p(d^h_k); W^1_p(S^C_i) - W^2_p(d^h_k)] \\
score^k_i &= \begin{cases} v^T_p tanh(u^k_i), & i > T_{k-1} \\ -\infty, & \text{otherwise} \end{cases} \qquad (2) \\
p(T_k|T_1, .., T_{k-1}; H^C, .., S^C_n) &= softmax(score^k)
\end{aligned}
$$

Method

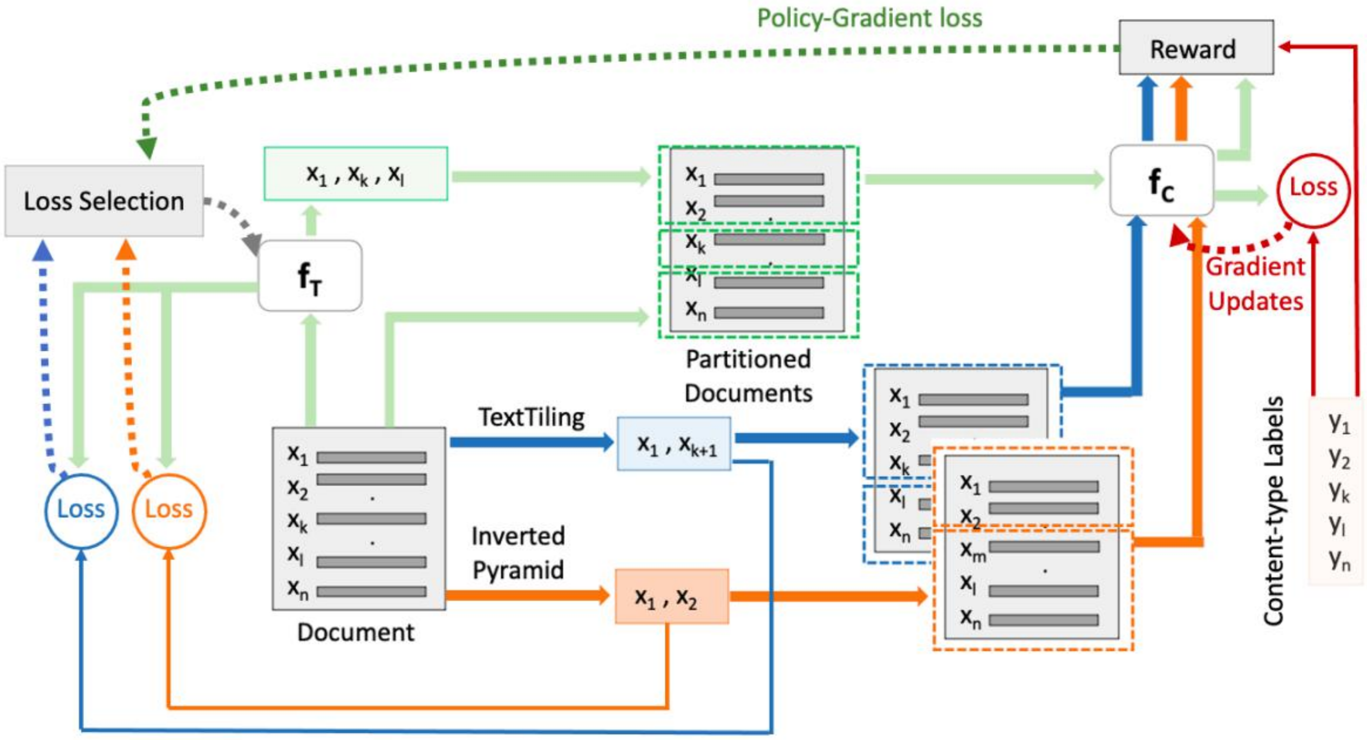

Figure 2: Neural-Network Architecture, including Gradient Flow Paths, for Incorporating Document-level Content Structures in a Discourse Profiling System

Discourse Profiling

T_L f_C

$$\alpha_s[i] = W_{s1}(tanh(W_{s2}S_i^C + b_{s2})) + b_{s1} \in R$$

$$A_T = softmax(\alpha_s[T_L[j] : T_L[j+1]]) \in R^{T_L[j]-T_L[j+1]}$$

$$T = \sum_{k=T_L[j]}^{T_L[j+1]} A_T[k].S^C[k] \in R^{2d_{rnn}}$$

$$A_s = softmax(\alpha_s) \in R^n$$

$$D = \sum_i A_s[i].H_s[i] \in R^{2d_{rnn}}$$

$$u_i = [S_i^C - T; S_i^C * T; T - D; T * D] \in R^{8d_{rnn}}$$

$$\hat{y}_i = softmax(W_{c1}(tanh(W_{c2}u_i + b_{c2})) + b_{c1}) \in R^9 \quad (3)$$

Method

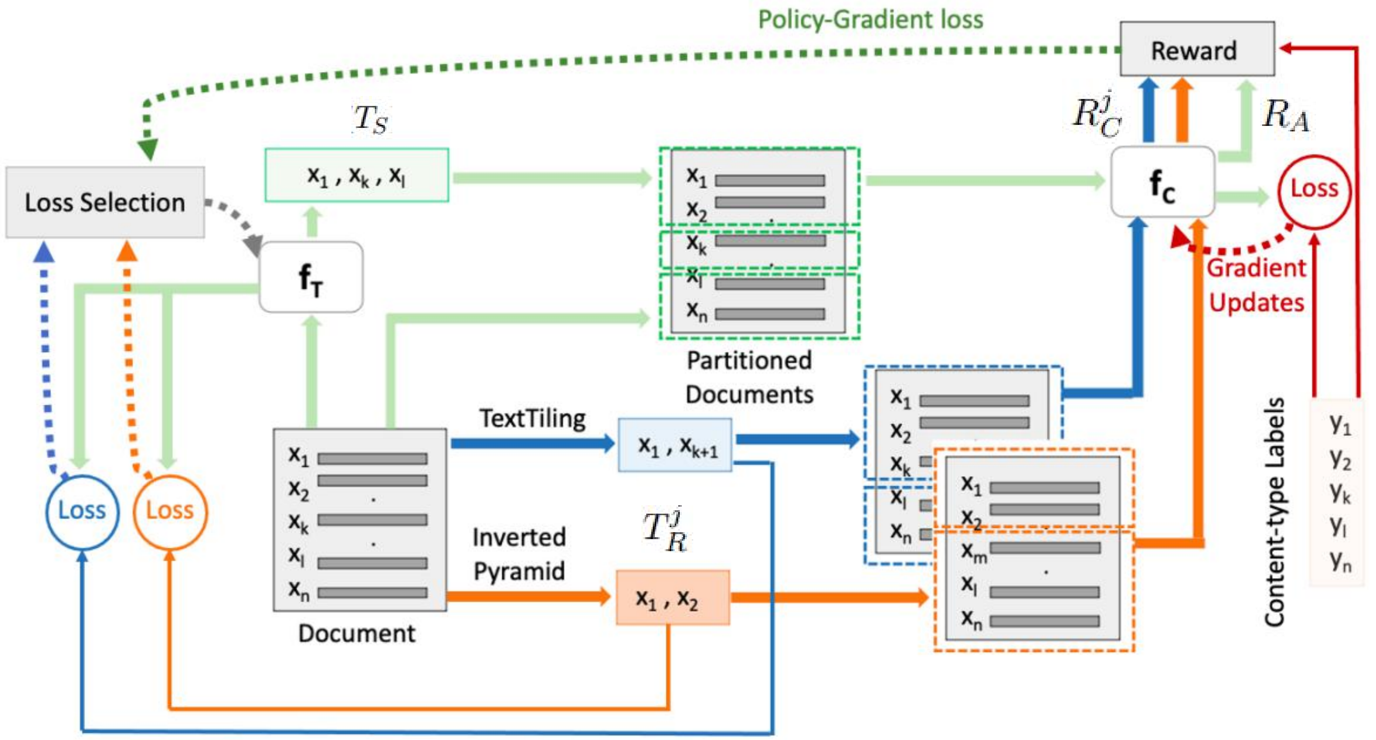

Figure 2: Neural-Network Architecture, including Gradient Flow Paths, for Incorporating Document-level Content Structures in a Discourse Profiling System

Learning f_T through Subtopic Structures-guided Critic

$$R_A > R_C^j \forall j$$

$$L_{RL} = (R_A - \bar{R_C})(\sum_i -log \frac{exp(T_S[i])}{\sum_{T_k \in T_S[i-1:]} exp(T_k)})$$

$$\bar{R_C} = \sum_{j=1}^{j=J} R_C^j / J$$

$$L_{IL} = \sum_i -log \frac{exp(T_R[i])}{\sum_{T_k \in T_R[i-1:]} exp(T_k)}$$

$$T_R = argmax_{T_R^j}(R_C^j)$$

$$L_C = \sum_i^n \sum_{c \in labels} -y_i^c log(\hat{y_i^c}) \tag{4}$$

Experiments

Models	Macro			Micro
	P	R	F1	F1
Hierarchical	55.60	51.10	51.70	58.24
Self-Critic	**58.61**	50.09	51.87	57.65
TextTiling	53.72	52.13	51.47	57.62
Joint-IP	55.74	51.34	52.45	58.65
RL-TT	57.67	52.91	53.02	58.12
RL-IP	56.04	53.76	54.15	59.07
RL-IP/TT	56.42	**55.20**	**54.42**	**59.21**

Table 1: Results for the best-performing systems on validation dataset.

Experiments

Models	M1	M2	C1	C2	D1	D2	D3	D4	Macro			Micro
	F1								P	R	F1	F1
Hierarchical	49.6	27.9	22.5	58.1	**64.1**	48.1	67.4	57.6	56.9	53.7	54.4(±0.80)	60.9(±0.70)
Self-Critic	51.5	29.4	27.2	58.2	61.4	55.3	67.5	59.7	59.0	55.1	56.1(±0.49)	61.6(±0.71)
TextTiling	50.7	**31.2**	26.1	57.6	61.1	52.3	66.5	58.7	58.5	54.0	55.5(±0.98)	60.6(±1.40)
Joint-IP	52.2	27.6	27.9	58.5	62.7	52.0	67.3	59.4	59.0	54.2	55.8(±0.56)	61.4(±0.70)
RL-TT	51.8	29.2	28.5	57.9	63.2	55.7	67.5	60.1	59.1	55.4	56.6(±0.46)	61.7(±0.61)
RL-IP	52.0	28.1	**28.9**	**58.7**	62.6	56.4	67.4	**60.6**	**59.3**	55.3	56.7(±0.37)	61.9(±0.38)
RL-IP/TT	**52.6**	28.7	26.6	58.0	63.5	**59.2**	**68.3**	**60.6**	58.7	**56.4**	**57.0**(±0.38)	**62.2**(±0.59)

Table 2: Performance of different systems on test dataset. All results correspond to average of 10 training runs with random seeds. In addition, we report standard deviation for both macro and micro F1 scores. Statistical significance tests show that both the macro and micro F1 scores for RL-IP/TT model are significantly better compared to the hierarchical, self-critic, TextTiling and joint-IP models with $p<0.05$ on paired t test (Dietterich, 1998). Similarly, the macro F1 scores for RL-TT and RL-IP models are significantly better compared to the hierarchical, TextTiling and Joint-IP models with $p<0.05$.

Experiments

	RM-IP	RM-TT	IP-TT	RM-TT-IP
Overlap	324	236	139	83

Table 3: Subtopic boundary sentences overlap between TextTiling and inverted pyramid subtopic structures and RL-IP/TT model on validation dataset. There are total 952 subtopic boundary sentences identified by RL-IP/TT model, and 589 and 540 subtopic boundary sentences identified by inverted pyramid and TextTiling structures respectively.

Experiments

	RM	IP	TT
Temporal frames	13	18	7

Table 4: Subtopic boundary sentences overlap between Temporal-frames based subtopic structure and TextTiling, inverted pyramid, and RL-IP/TT model on a subset of 10 documents from validation dataset. There are total 68 subtopic boundary sentences identified by RL-IP/TT model, and 79, 52 and 58 subtopic boundary sentences identified by inverted pyramid, TextTiling, and temporal frames-based structures respectively.

Thanks