Open-Domain Aspect-Opinion Co-Mining with Double-Layer Span Extraction

2022/11/29

Advisor : Jia-Ling, Koh

Speaker: Chi-Yu-Fan

Source: KDD 2022

Outline

- Introduction
- Method
- Experiment
- Conclusion

Introduction(Task)

- Understanding customer requirements is crucial for business
- Review analysis to enhance their services

- **Aspect**: The product or service attribute
- **Opinion**: The reviewer's opinion towards the corresponding product or aspects(service) of a product
- **Span** : Each word we call span
- **Open-domain task**: lack training data, simultaneously mines find correspondence join in the model (can help it cross domain)

Introduction(Input)

Aspect term extraction(ATE)

$$A = \{A_1, A_2, .., A_i\}$$

" wine list"

$$R = \{w_1, w_2, ..., w_k\}$$

the wine list is extensive and impressive

Opinion term extraction(OTE)

$$(O = \{O_1, O_2, ..., O_j\}$$

"extensive" and "impressive"

Aspect-Opinion pair extraction(AOPE)

$$P = \{(A_i, O_j), ..\}$$

("wine list","extensive")
("wine list","impressive")

Introduction (Method)

- Supervised extraction methods of the review require large-scale **human-annotated label data**.

- **Weak label:** generates training data without **human annotations** by **rules-based** on **universal dependency parsing**
- **Doubled layer**
- **Early stopping** to avoid the model over-fitting to the noise to tackle the noisy weak supervision
- **Self-train: Semi-supervised enrich label**

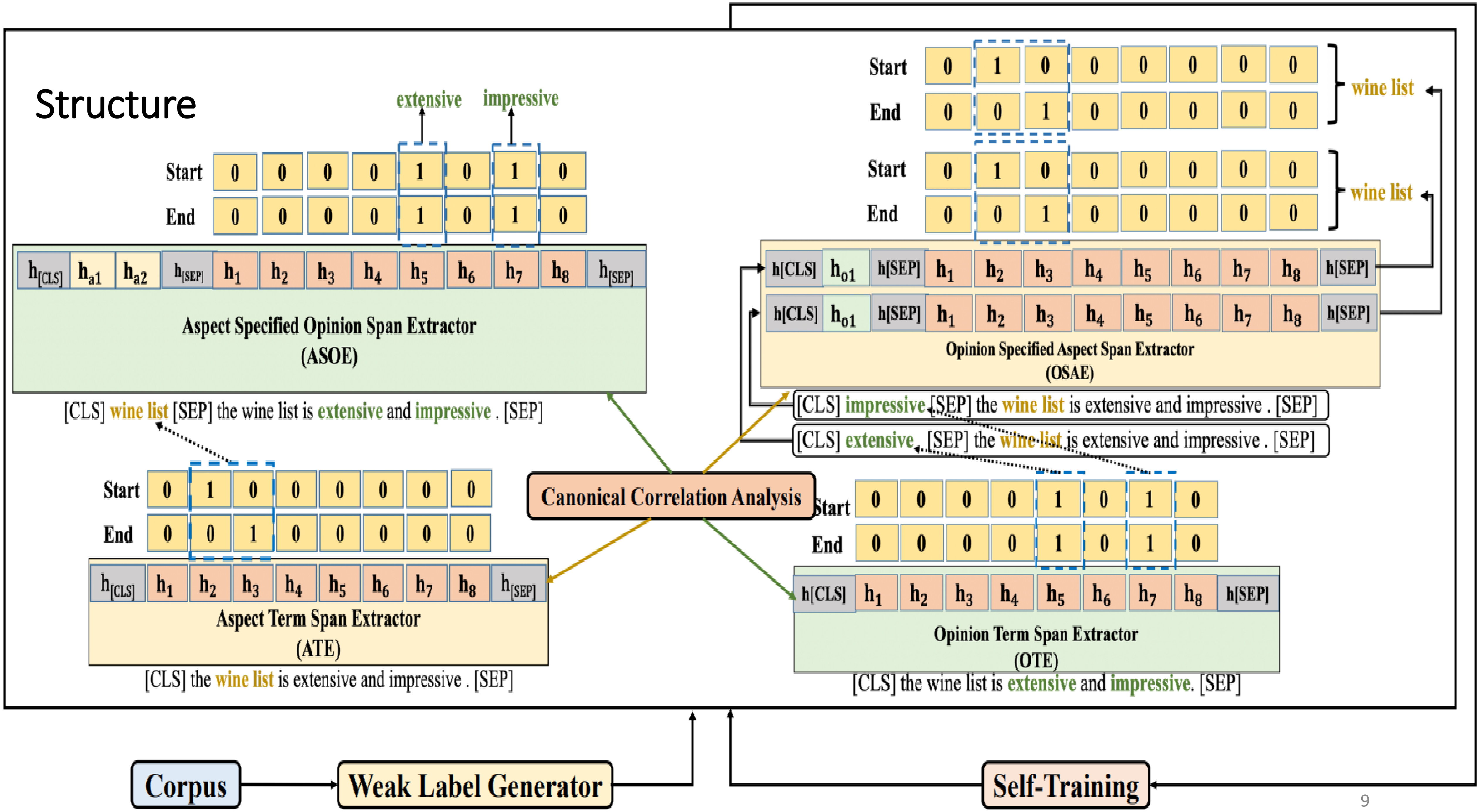
Outline

- Introduction
- Method
- Experiment
- Conclusion

Method

- Open-Domain
- ATE-OTE
- AOPE
- Loss
- Early Stop
- Self-training

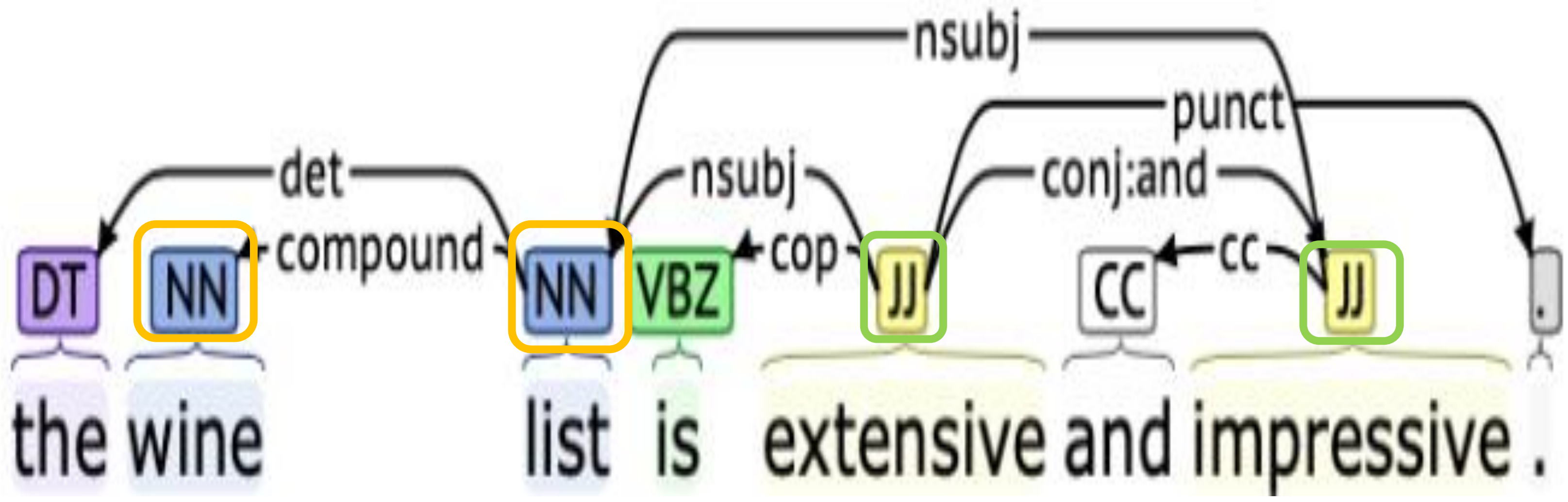
Structure



extensive
impressive
Start 0 0 0 0 1 0 1 0
End 0 0 0 0 1 0 1 0
$h_{[CLS]}$ h_{a1} h_{a2} $h_{[SEP]}$ h_1 h_2 h_3 h_4 h_5 h_6 h_7 h_8 $h_{[SEP]}$
Aspect Specified Opinion Span Extractor
(ASOE)
[CLS] wine list [SEP] the wine list is extensive and impressive . [SEP]
Start 0 1 0 0 0 0 0 0
End 0 0 1 0 0 0 0 0
$h_{[CLS]}$ h_1 h_2 h_3 h_4 h_5 h_6 h_7 h_8 $h_{[SEP]}$
Aspect Term Span Extractor
(ATE)
[CLS] the wine list is extensive and impressive . [SEP]
Canonical Correlation Analysis
Start 0 1 0 0 0 0 0 0
End 0 0 1 0 0 0 0 0
wine list
Start 0 1 0 0 0 0 0 0
End 0 0 1 0 0 0 0 0
wine list
h[CLS] h_{o1} h[SEP] h_1 h_2 h_3 h_4 h_5 h_6 h_7 h_8 h[SEP]
h[CLS] h_{o1} h[SEP] h_1 h_2 h_3 h_4 h_5 h_6 h_7 h_8 h[SEP]
Opinion Specified Aspect Span Extractor
(OSAE)
[CLS] impressive [SEP] the wine list is extensive and impressive . [SEP]
[CLS] extensive [SEP] the wine list is extensive and impressive . [SEP]
Start 0 0 0 0 1 0 1 0
End 0 0 0 0 1 0 1 0
h[CLS] h_1 h_2 h_3 h_4 h_5 h_6 h_7 h_8 h[SEP]
Opinion Term Span Extractor
(OTE)
[CLS] the wine list is extensive and impressive. [SEP]
Corpus
Weak Label Generator
Self-Training

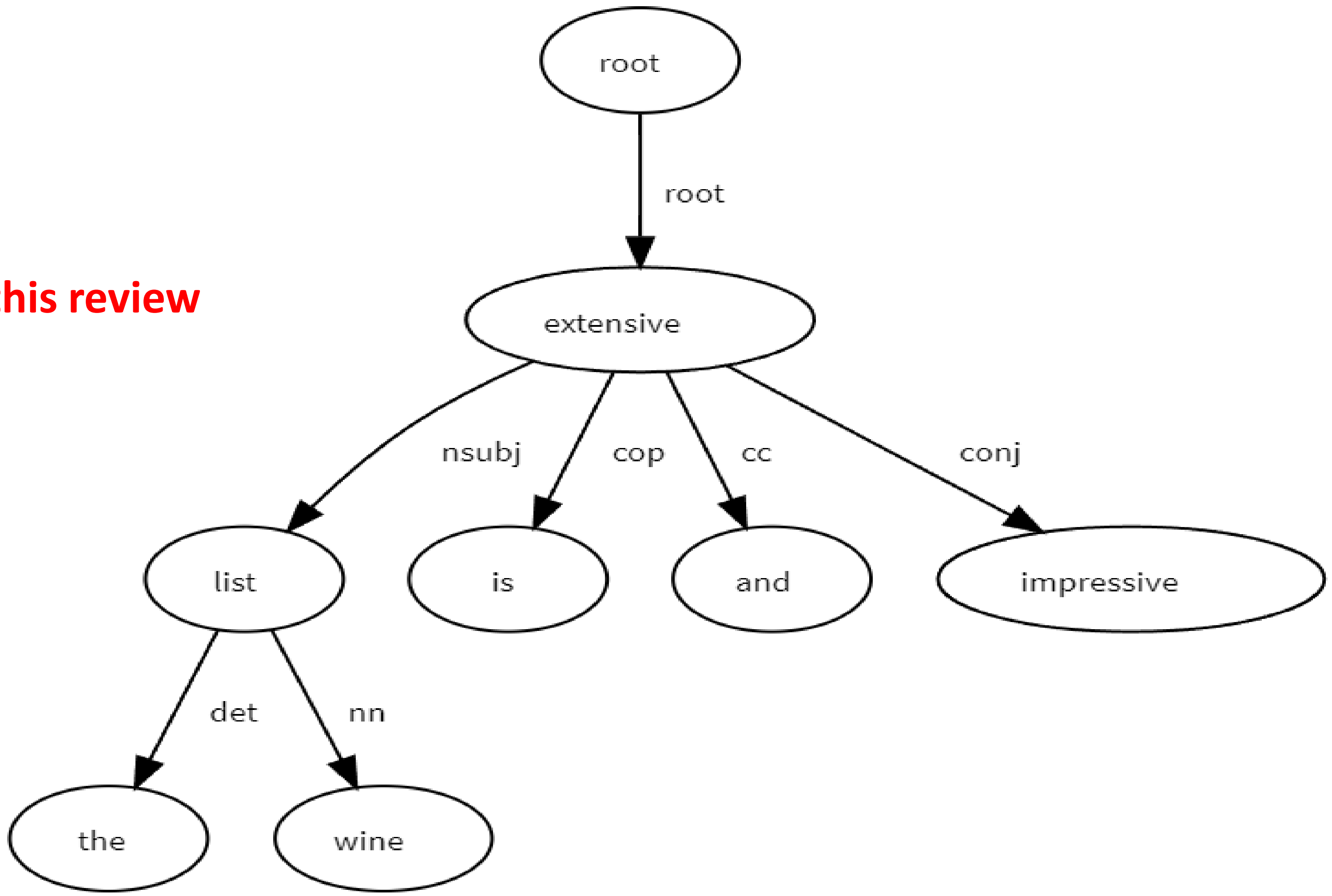
Weak-label(open domain)

- Use the **Parser tree.**
- **With out Human annotation train data**
- **This rule states that noun word (NN) is aspect term and adjective word (JJ) is an opinion term**
- **The rule can only label small portion of data(25.68% of data) but high precision of it**



parser tree

adjective word is the root of this review

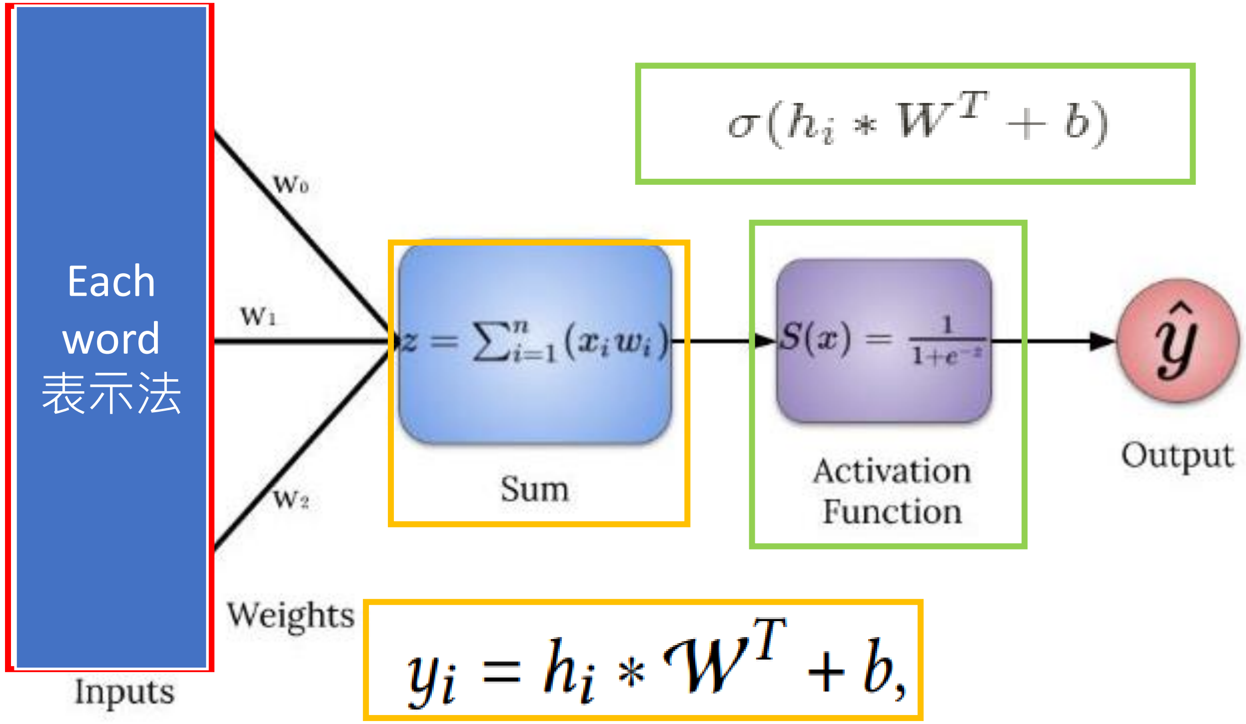


Model Description

- **Encoder**: Provide rich semantic, syntactic, and context-sensitive information for each token in the input review.(We use bert base encoder for our model)

- **Framework:** Build **four independent** encoders to tackle the tasks of $AT\ E$, $OT\ E$, $OSAE$, and ASOE

Encoder output



$$\sigma(h_i * W^T + b)$$

$$y_i = h_i * W^T + b,$$

$$H = \{h_{[CLS]}, h_1, h_2, \ldots, h_{[SEP]}\}$$

If the probability we label it 1

$$\hat{y}_i^s = \begin{cases} 1, & \text{if } h_{i_s} > 0; \\ 0, & else. \end{cases}$$

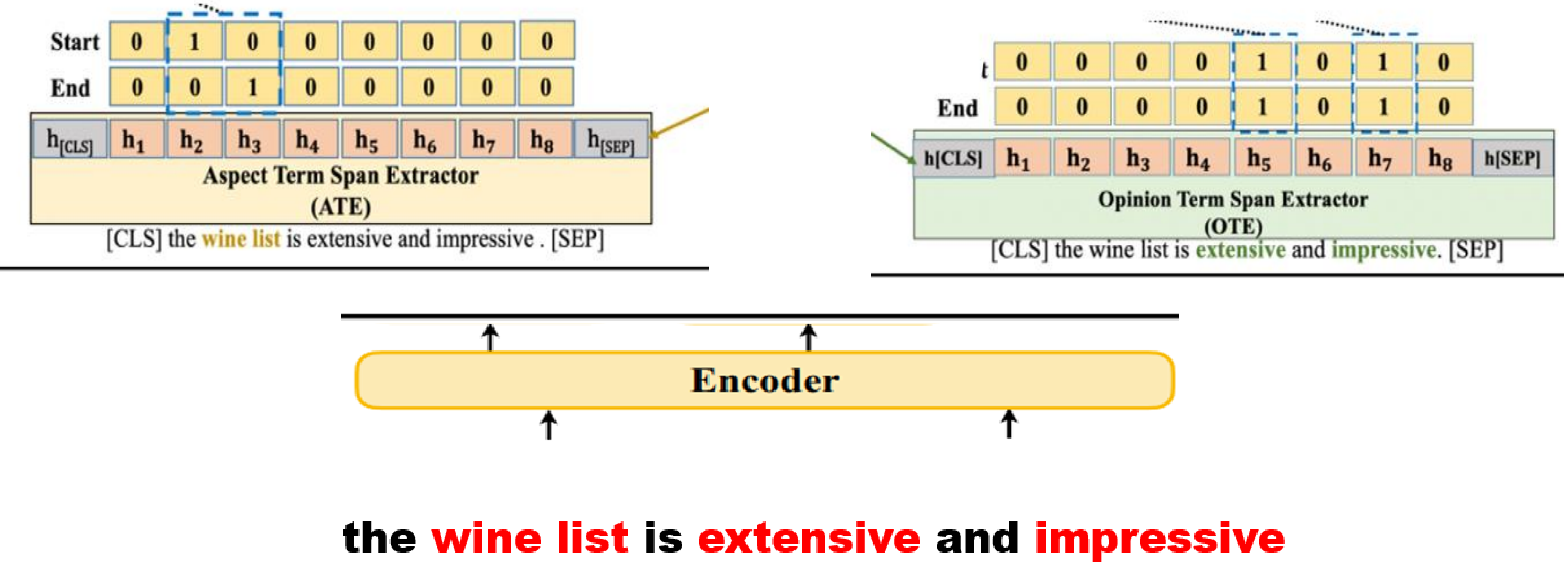
$$\hat{y}_i^e = \begin{cases} 1, & \text{if } h_{i_e} > 0; \\ 0, & else. \end{cases}$$

$$h_{i_s} = y_i[0],$$

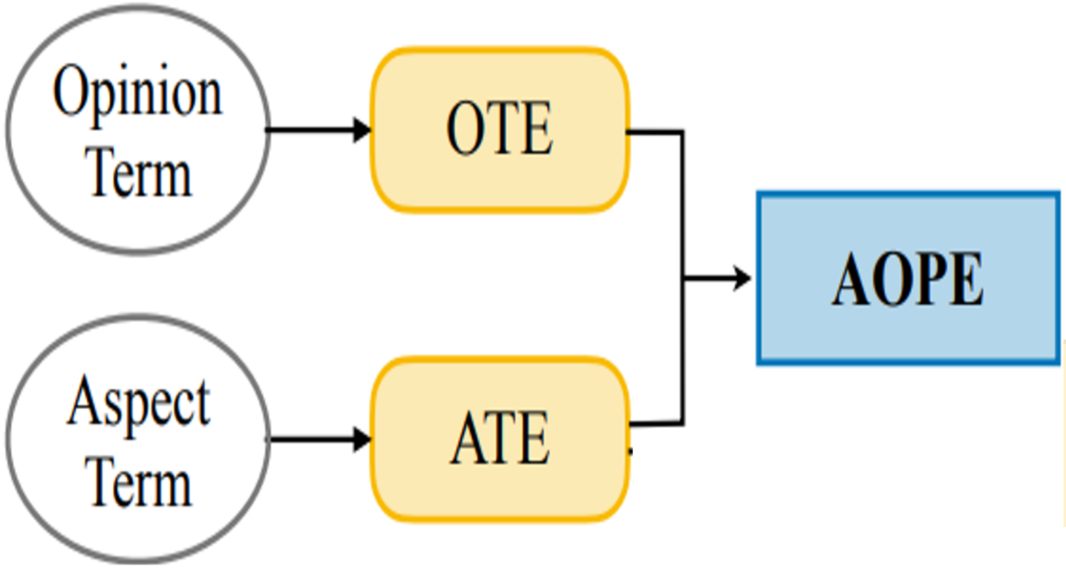
$$h_{i_e} = y_i[1],$$

The prediction from the start span and end span

ATE-OTE


Start
End
Aspect Term Span Extractor
(ATE)
[CLS] the wine list is extensive and impressive . [SEP]
End
Opinion Term Span Extractor
(OTE)
[CLS] the wine list is extensive and impressive. [SEP]
Encoder
the wine list is extensive and impressive

Aspect-Opinion Pair Extraction

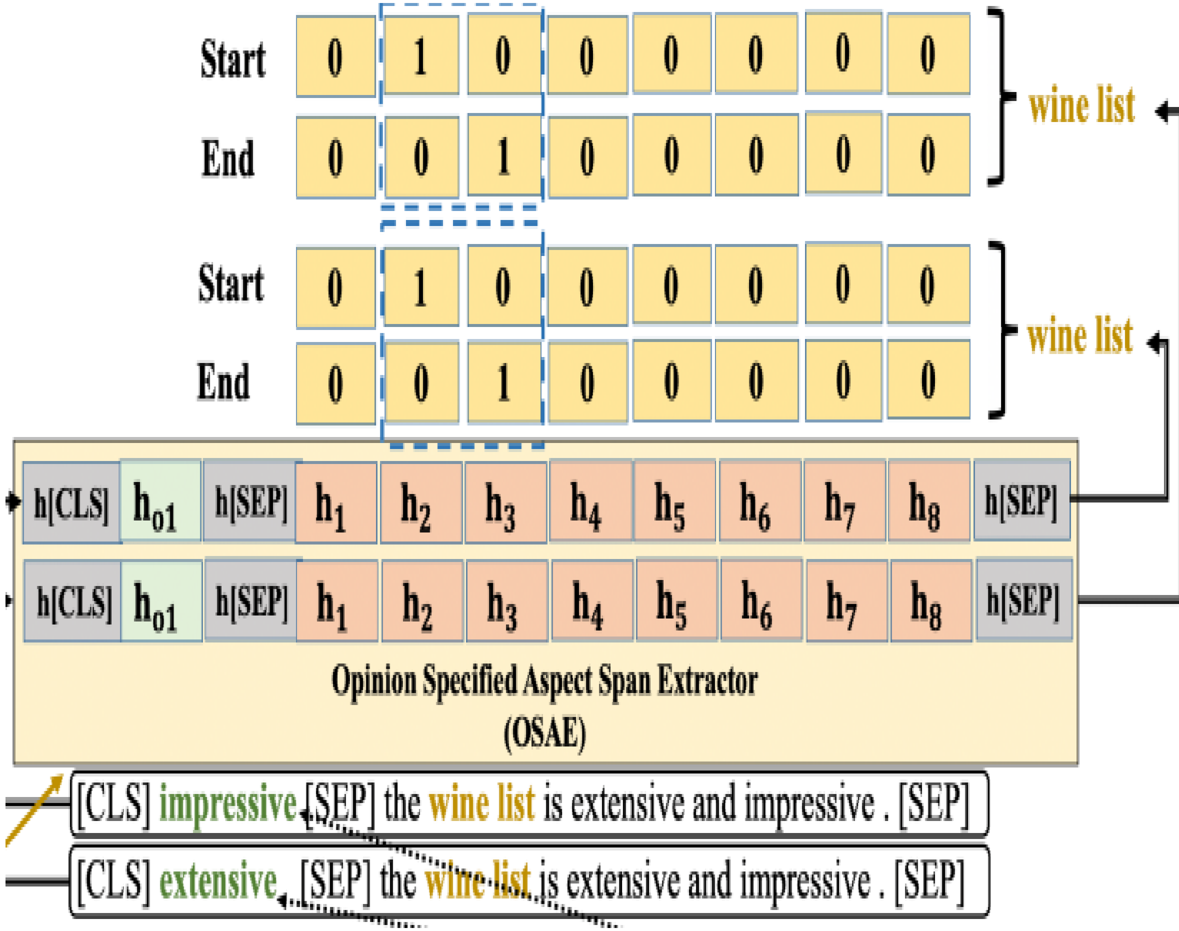


- Get the Pair Extraction from the **Opinion** &**Aspect**

Aspect Specified Opinion Span Extractor
(ASOE)

Opinion Specified Aspect Span Extractor
(OSAE)

Aspect-Opinion Pair Extraction



Loss function

$$\mathcal{L}_{ASOE} = \frac{\mathcal{L}^{s}_{ASOE} + \mathcal{L}^{e}_{ASOE}}{2} = \frac{\sum_{i=1}^{N'} \sum_{sp \in \{s,e\}} BCE\left(\boxed{\hat{y}_i^{sp}}, \boxed{y_i^{sp}}\right)}{2}. \quad (4)$$

$$\frac{1}{N} \sum_{i=1}^{N} -\left(\hat{y}_i^{sp} * \log(y_i^{sp}) + (1-\hat{y}_i^{sp}) * \log(1-y_i^{sp})\right)$$

- The loss function is also the averaged binary cross-entropy loss (BCE) between the predicted spans and labeled spans.

Loss function

$$\mathcal{L} = \mathcal{L}_{ATE} + \mathcal{L}_{OTE} + \mathcal{L}_{ASOE} + \mathcal{L}_{OSAE}. \qquad (5)$$

- Sum of loss from $\boldsymbol{ATE}$, $\boldsymbol{OTE}$, $\boldsymbol{ASOE}$, and $\boldsymbol{OSAE}$ modules

Canonical correlation analysis:

The hidden representations of the reviews to measure this **correlation** and use CCA as **early stopping** criteria during training

Representation(Aspect &Opinion) → Calculate
- **Cross Covariance Matrix**
- **Covariance Matrix**

→ Canonical correlation → Calculate Eigenvalue → Find out argmax

Formula

$$\mathrm{Cov}(x,y) = \boxed{\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}}$$

Cross Covariance Matrix

$$\frac{u^{\top}\,\Sigma_{AO}\,v}{\sqrt{(u^{\top}\;\Sigma_{AA}\;u)(v^{\top}\;\Sigma_{OO}\;v)}}.$$

Correlation $corr$

Covariance Matrix

$$\boxed{\mathrm{var}(x) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \qquad \boxed{\mathrm{var}(y) = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}}$$

Canonical correlation analysis

$$Cross\ Covariance\ Matrix$$

$$C_{AO} = \begin{pmatrix} C_{A_1O_1} & C_{A_1O_2} \\ C_{A_2O_1} & C_{A_2O_2} \end{pmatrix}$$

$$Canonical\ Correlation$$

$$C_A = C_{AA}^{-1} C_{AO} C_{OA} C_{OO}^{-1}$$

$$Canonical\ Correlation$$

$$C_O = C_{OO}^{-1} C_{OA} C_{AO} C_{AA}^{-1}$$

Calculate Eigenvalue

$$R_A = u$$

$$R_O = v$$

Find out max correlation

$$\rho_1 = \frac{u^\top \Sigma_{AO}\, v}{\sqrt{(u^\top\ \Sigma_{AA}\ u)(v^\top\ \Sigma_{OO}\ v)}}.$$

$$(u^{'}, v^{'}) = argmax\ corr(u^T H_A, v^T H_O)$$

Early Stopping:

$$(u', v') = \underset{u,v}{\operatorname{argmax}}\ corr(u^\top H_{ATE}, v^\top H_{OSAE}),$$

$$\rho_1 = corr(u^\top H_{ATE}, v^\top H_{OSAE})$$

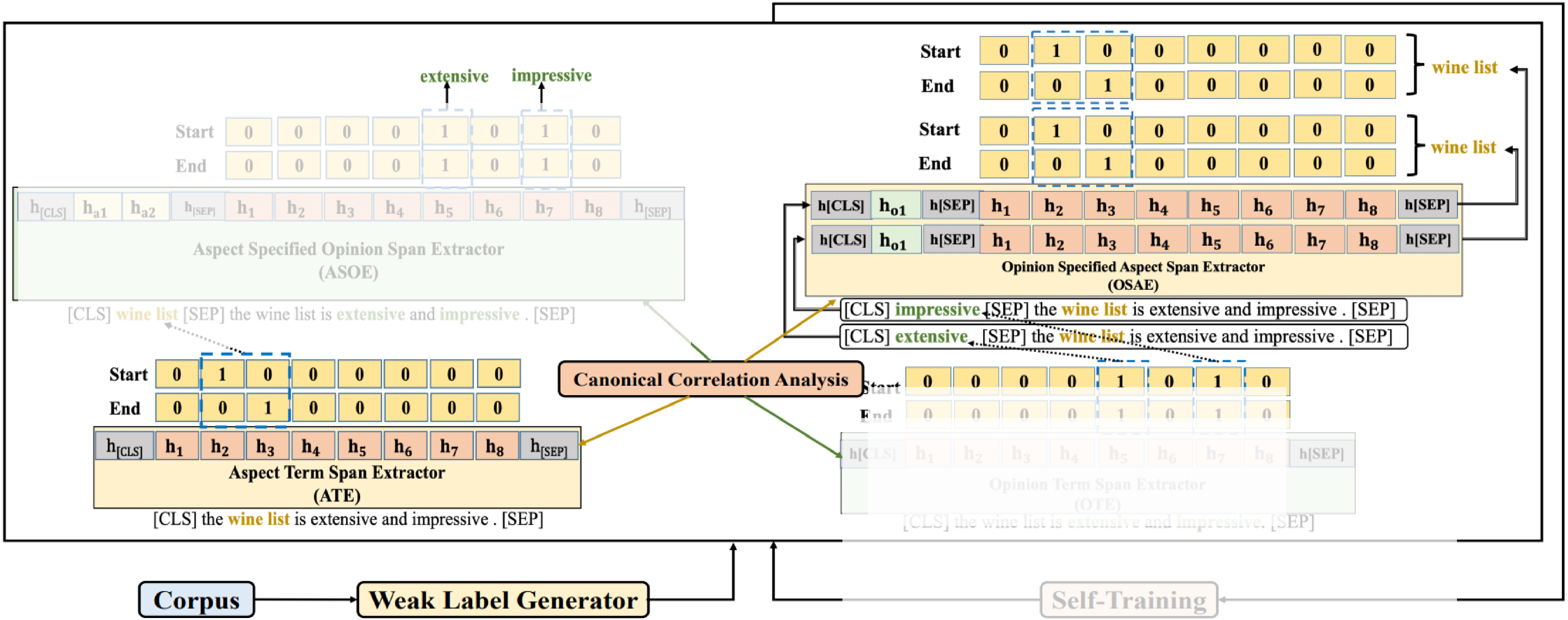
$$\rho_1 = \frac{u^\top \Sigma_{AO}\, v}{\sqrt{(u^\top\ \Sigma_{AA}\ u)(v^\top\ \Sigma_{OO}\ v)}}.$$

$$\rho = \frac{\sum_M (\rho_1 + \rho_2)}{\boxed{M}},$$

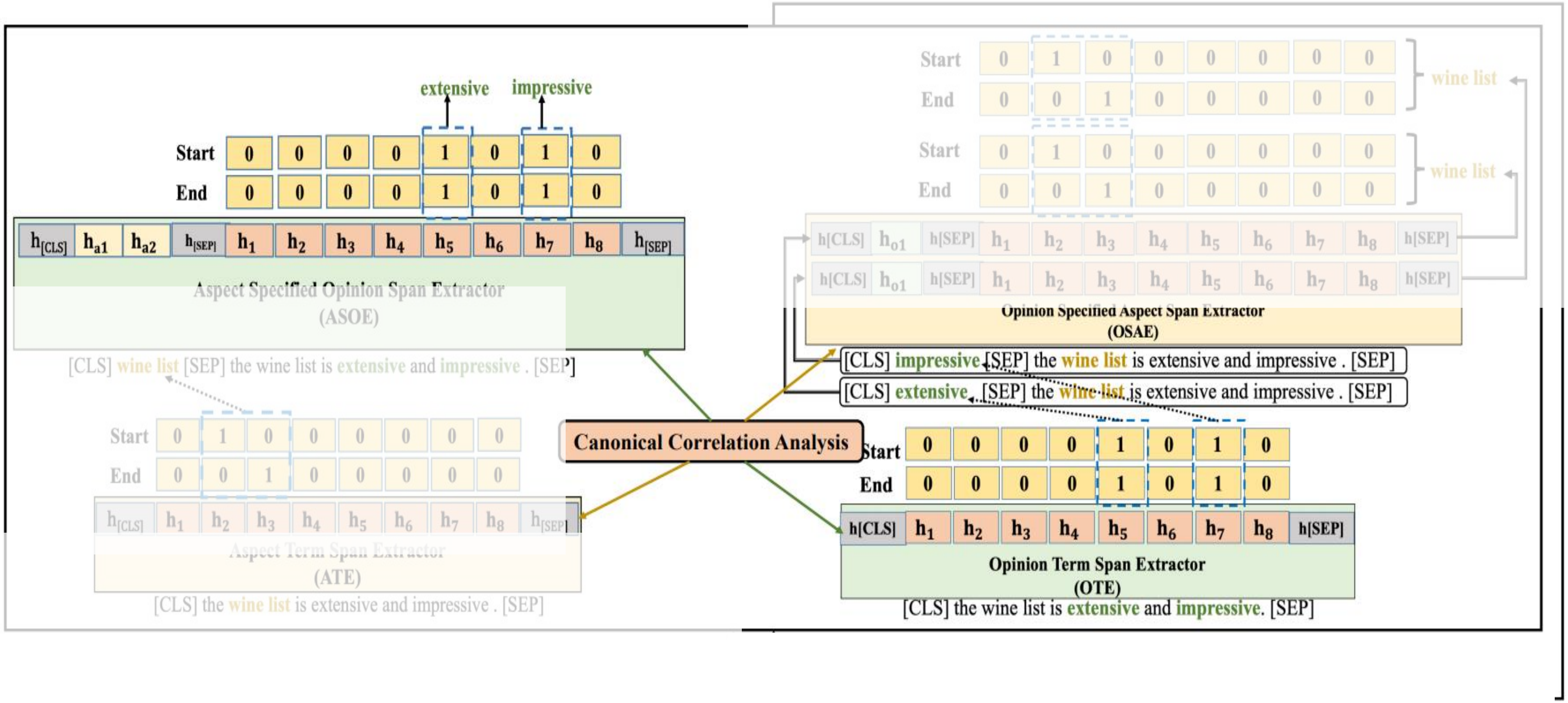
the number of reviews

Early stopping can prevent the model from over-fitting to the label noise.

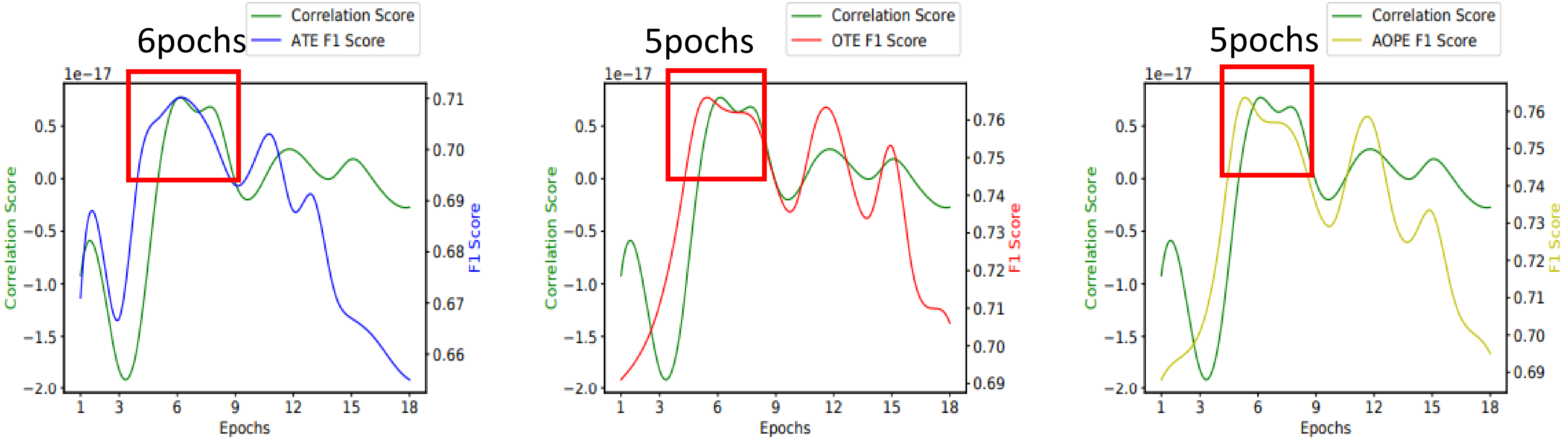
Early Stoppingρ1



Early Stoppingρ2



Early Stopping



(a) CCA score and ATE F1 score (b) CCA score and OTE F1 score (c) CCA score and AOPE F1 score

Use the following observations to solve this problem

Self-Training

$$\gamma_R = \boxed{A'_{ATE}} \Delta A'_{OSAE} + O'_{OTE} \Delta O'_{ASOE},$$

$$A \Delta B = (A - B) \bigcup (B - A)$$

- **Correctly predicted** reviews
- Is added in D$'$ $labeled$ to enrich the training data

Outline

- Domain
- Method
- **Experiment**
- Conclusion

Datasets:SemEval14~16

Datasets	$\mathbb{S}_{14}l$		$\mathbb{S}_{14}r$		$\mathbb{S}_{15}r$		$\mathbb{S}_{16}r$	
	Train	Test	Train	Test	Train	Test	Train	Test
#sentences	3045	800	3041	800	1315	685	2000	676
#aspects	2359	653	3693	1134	1205	542	1757	622
#opinions	2500	677	3512	1014	1217	516	1381	475

Datasets	$\mathbb{S}_{14}l$		$\mathbb{S}_{14}r$		$\mathbb{S}_{15}r$		$\mathbb{S}_{16}r$	
	Train	Test	Train	Test	Train	Test	Train	Test
#sentences	1158	343	1627	500	754	325	1079	329
#pairs	1634	482	2643	865	1076	436	1512	457

SemEval Datasets : The current state of the arts **semantic analysis and annotated datasets**

Datasets:SemEval14~16

The pizza is delicious.

Aspect Term	*pizza*
Aspect Category	food
Sentiment Polarity	POS

SemEval Datasets : The current state of the arts **semantic analysis and annotated datasets**

Base-Line(Method reduce)

Methods	$\mathbb{S}_{14}l$	$\mathbb{S}_{14}r$	$\mathbb{S}_{15}r$	$\mathbb{S}_{16}r$
ODAO	76.14	80.73	80.72	79.24
-Pair Extraction Modules	50.13	57.53	60.86	60.71
-Self Training	62.06	72.19	72.13	71.0

Base-Line(Human Effort)

Method	**Human Effort**	$\mathbb{S}_{14}l$	$\mathbb{S}_{14}r$	$\mathbb{S}_{15}r$	$\mathbb{S}_{16}r$
RINANTE	Gold Annotation	80.16	86.45	69.90	-
QDSL		84.27	87.85	77.72	83.34
PSTD		**86.91**	88.75	75.82	82.56
DeepWMaxSat		81.33	85.33	-	73.67
FS-ODAO		85.93	**88.77**	**83.39**	**86.15**
ABAE	None	32.9	40.2		
LCC+GBC		36.1	41.2		
GMTCMLA	Sample Annotation	56.08	76.51	61.75	-
AutoNER	Dictionary	65.44	-	-	-
DP	Rule Design	19.19	38.72	27.32	-
ODAO		**76.14**	**80.73**	**80.72**	**79.24**

FS-ODAO, the fully supervised version of ODAO, achieves state-of-the-art performance

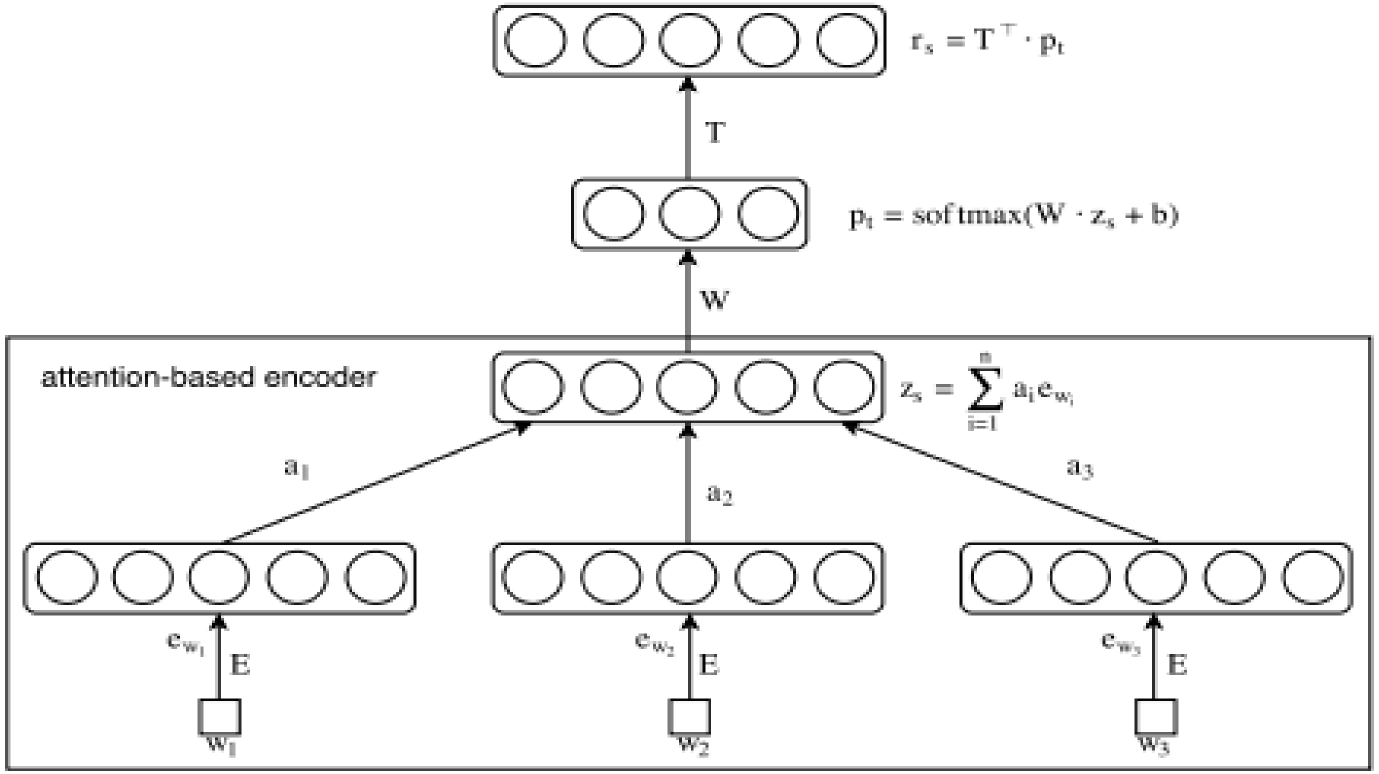
Human Effort

- Gold Annotation(Supervised) **Gold standard annotated corpora**
- None (Unsupervised)
- Sample Annotation (few-shot)
- Dictionary
- Rule Design

FS-ODAO

- Remove the **early stop** and **self-training** steps.
- FS-ODAO is trained in the same setting with other supervised baseline methods

ABAE(Unsupervised)



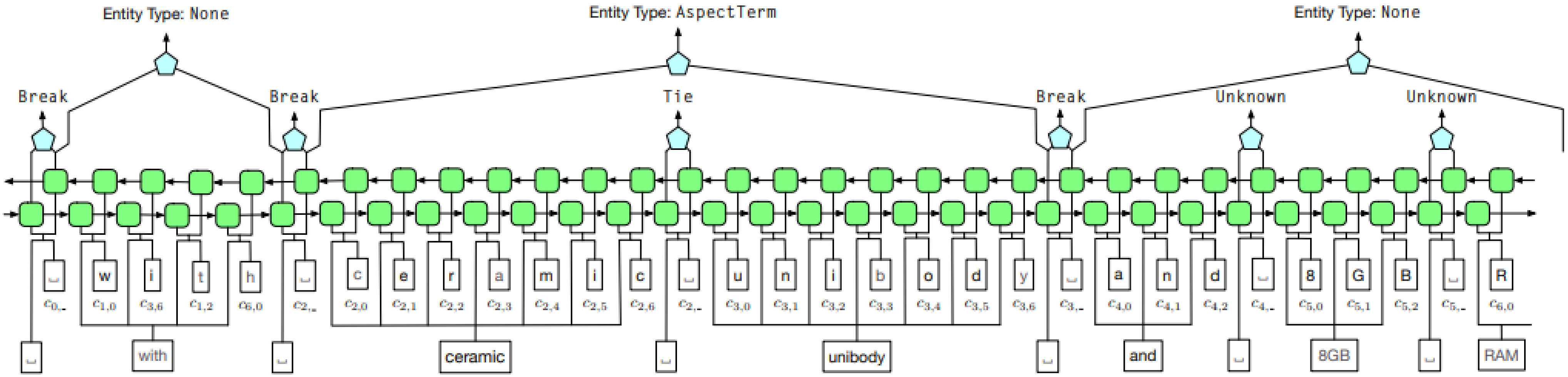
Sample Annotation

Method	**Human Effort**	$\mathbb{S}_{14}l$	$\mathbb{S}_{14}r$	$\mathbb{S}_{15}r$	$\mathbb{S}_{16}r$
RINANTE	Gold Annotation	80.16	86.45	69.90	-
QDSL		84.27	87.85	77.72	83.34
PSTD		**86.91**	88.75	75.82	82.56
DeepWMaxSat		81.33	85.33	-	73.67
FS-ODAO		85.93	**88.77**	**83.39**	**86.15**
ABAE	None	32.9	40.2		
LCC+GBC		36.1	41.2		
GMTCMLA	Sample Annotation	56.08	76.51	61.75	-
AutoNER	Dictionary	65.44	-	-	-
DP	Rule Design	19.19	38.72	27.32	-
ODAO		**76.14**	**80.73**	**80.72**	**79.24**

Dictionary

<table>
<tr><th>Method</th><th>Human Effort</th><th>$\mathbb{S}_{14}l$</th><th>$\mathbb{S}_{14}r$</th><th>$\mathbb{S}_{15}r$</th><th>$\mathbb{S}_{16}r$</th></tr>
<tr><td>RINANTE</td><td rowspan="5">Gold Annotation</td><td>80.16</td><td>86.45</td><td>69.90</td><td>-</td></tr>
<tr><td>QDSL</td><td>84.27</td><td>87.85</td><td>77.72</td><td>83.34</td></tr>
<tr><td>PSTD</td><td>86.91</td><td>88.75</td><td>75.82</td><td>82.56</td></tr>
<tr><td>DeepWMaxSat</td><td>81.33</td><td>85.33</td><td>-</td><td>73.67</td></tr>
<tr><td>FS-ODAO</td><td>85.93</td><td>88.77</td><td>83.39</td><td>86.15</td></tr>
<tr><td>ABAE</td><td rowspan="2">None</td><td>32.9</td><td colspan="3">40.2</td></tr>
<tr><td>LCC+GBC</td><td>36.1</td><td colspan="3">41.2</td></tr>
<tr><td>GMTCMLA</td><td>Sample Annotation</td><td>56.08</td><td>76.51</td><td>61.75</td><td>-</td></tr>
<tr><td>AutoNER</td><td>Dictionary</td><td>65.44</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td>DP</td><td rowspan="2">Rule Design</td><td>19.19</td><td>38.72</td><td>27.32</td><td>-</td></tr>
<tr><td>ODAO</td><td>76.14</td><td>80.73</td><td>80.72</td><td>79.24</td></tr>
</table>

AutoNer(Dictionary)



Rule Design

Method	**Human Effort**	$\mathbb{S}_{14}l$	$\mathbb{S}_{14}r$	$\mathbb{S}_{15}r$	$\mathbb{S}_{16}r$
RINANTE	Gold Annotation	80.16	86.45	69.90	-
QDSL		84.27	87.85	77.72	83.34
PSTD		**86.91**	88.75	75.82	82.56
DeepWMaxSat		81.33	85.33	-	73.67
FS-ODAO		85.93	**88.77**	**83.39**	**86.15**
ABAE	None	32.9	40.2		
LCC+GBC		36.1	41.2		
GMTCMLA	Sample Annotation	56.08	76.51	61.75	-
AutoNER	Dictionary	65.44	-	-	-
DP	Rule Design	19.19	38.72	27.32	-
ODAO		**76.14**	**80.73**	**80.72**	**79.24**

Outline

- Domain
- Method
- Experiment
- **Conclusion**

Conclusion

- Aspect-opinion co-extraction and pair extraction tasks show that ODAO can achieve competitive or even better performance.
- ODAO can handle the noise and bias of the weak supervision.
- The double-layer design's effectiveness.