SEMGraph: Incorporating Sentiment Knowledge and Eye Movement into Graph Model for Sentiment Analysis

Bingbing Wang[1,2], **Bin Liang**[1,2], **Jiachen Du**[1,2], **Min Yang**[3], **and Ruifeng Xu**[1,2,4*]
[1] Harbin Insitute of Technology, Shenzhen, China
[2] Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies
[3] SIAT, Chinese Academy of Sciences, Shenzhen, China
[4] Peng Cheng Laboratory, Shenzhen, China
{bingbing.wang,bin.liang}@stu.hit.edu.cn, xuruifeng@hit.edu.cn,
jacobvan199165@gmail.com, min.yang@siat.ac.cn

Abstract

This paper investigates the sentiment analysis task from a novel perspective by incorporating sentiment knowledge and eye movement into a graph architecture, aiming to draw the eye movement-based sentiment relationships for learning the sentiment expression of the context. To be specific, we first explore a linguistic probing eye movement paradigm to extract eye movement features based on the close relationship between linguistic features and the early and late processes of human reading behavior. Furthermore, to derive eye movement features with sentiment concepts, we devise a novel weighting strategy to integrate sentiment scores extracted from affective commonsense knowledge into eye movement features, called sentiment-eye movement weights. Then, the sentiment-eye movement weights are exploited to build the sentiment-eye movement guided graph (SEMGraph) model, so as to model the intricate sentiment relationships in the context. Experimental results on two sentiment analysis datasets with eye movement signals and three sentiment analysis datasets without eye movement signals show that the proposed SEMGraph achieves state-of-the-art performance, and can also be directly generalized to those sentiment analysis datasets without eye movement signals.

1 Introduction

As one of the hot research directions in Natural Language Processing (NLP), sentiment analysis (SA), aiming to extract knowledge and analyze the sentiment of texts, is crucial in a wide range of applications across multiple domains such as customer opinion identification (Kumar et al., 2019; Bose et al., 2019), product recommendation (Dang et al., 2021), and mental health monitoring based on patients' social media posts (Rajput, 2019), etc.

Some pioneering research efforts in cognitive science affirm the association between cognitive processing and emotional manifestations (Isen and Means, 1983; Storbeck and Clore, 2007). Intuitively, humans are quite apt in extracting subjective content and emphases from texts when reading a context. Based on the intuition that the cognitive data which covers eye movement signals or electroencephalography (EEG) signals might bring substantial improvements to sentiment analysis, recent research on sentiment analysis relying on eye movement signals has been attached importance (Mishra et al., 2017; Long et al., 2021; Chen et al., 2022; Joshi et al., 2014; Hollenstein et al., 2018).

Despite promising progress made by existing eye movement-based sentiment analysis studies, they largely rely on the data containing eye movement signals. In the real-world scenario, however, the acquisition of eye movement signals is expensive, and many commonly used sentiment analysis datasets lack eye movement signals. In addition, existing eye movement-based sentiment analysis work ignores the role of sentiment concepts in the learning of eye movement signals.

Therefore, in this paper, we propose a novel framework to incorporate sentiment knowledge and eye movement, so as to derive the sentiment-eye movement weights. Then, the sentiment-eye movement weights are exploited to build the sentiment-eye movement guided graph (SEMGraph) model. The most significant thing to note here is that the SEMGraph model can not only tackle the sentiment analysis of the data with eye movement signals, but also be directly generalized to deal with sentiment analysis of the data without eye movement signals. To be specific, we first propose a linguistic probing eye movement paradigm to extract eye movement features from linguistic features. Motivated by Roberts and Siyanova-Chanturia (2013) and Clifton et al. (2007) about the early and late processes of human reading behavior, we establish relationships from two perspectives based on a regression model by using eye movement features

* Corresponding Authors

Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 7521–7531
December 7-11, 2022 ©2022 Association for Computational Linguistics

as dependent variables and the linguistic features as independent variables. One is the relationship of two linguistic features (character and word features) and *first fixation duration* (FFD), the other is three linguistic features (character, word, and complexity feature illustrated in Section 3.2) and *total reading time* (TRT).

Furthermore, leveraging sentiment concepts is a key to improving the learning of sentiment analysis (Pang et al., 2008; Liu, 2012). Therefore, we extract the sentiment concepts from an affective commonsense knowledge (Cambria et al., 2010), and then devise a novel weighting strategy to integrate the sentiment concepts into eye movement features, so as to derive sentiment-eye movement weights including FFD weight and TRT weight. Based on it, to model the intricate sentiment and eye movement relationships between contextual words, we incorporate the sentiment-eye movement weights into a Gated Graph Neural Network (GGNN) architecture to build a sentiment-eye movement guided graph (SEMGraph) model. More concretely, considering the close relationships between FFD and word features, we adopt GGNN to study the embeddings of words, and add FFD weight for better information interaction. Meanwhile, according to the connection of TRT and more complex linguistic features, after obtaining the word embedding by GGNN, soft attention weight and TRT weight are combined to enhance graph-level information aggregation for learning the sentiment cues.

To sum up, the contributions of this work can be summarized as follows:

- The SA task is approached from a novel perspective by introducing sentiment-eye movement weights to improve the learning of sentiment information.

- An effective paradigm termed linguistic probing eye movement paradigm is proposed to extract eye movement features based on the close relationship between linguistic features and the early and late processes of human reading.

- A novel sentiment-eye movement guided graph (SEMGraph) model is explored to draw the eye movement-based sentiment relationships in SA.

- Experimental results on two sentiment analysis datasets with eye movement signals and three sentiment analysis datasets without eye movement signals demonstrate that the proposed SEMGraph achieves state-of-the-art performance in SA and simultaneously can be generalized to the sentiment analysis datasets without eye movement signals.

2 Related work

2.1 Sentiment Analysis

Sentiment analysis can be regarded as a text classification problem (Schuller et al., 2015; Choi and Lee, 2017; Ju and Li, 2013; Liang et al., 2022b), and the process ends up with classifying whether a given text expresses a positive or negative sentiment. It has been commonly investigated at document, sentence level or the aspect level (Liang et al., 2022a; Behdenna et al., 2016; Do et al., 2019; Liang et al., 2021). Many existing research efforts focus on SA with deep learning methods to capture the significant feature of sentiment detection (Kim, 2014; Li et al., 2020; Zhou et al., 2014). Furthermore, researchers have attempted to analyze sentiment depending on integrating with some other algorithm components instead of single deep learning method (Zharmagambetov and Pak, 2015; Li et al., 2014).

Another trend leverages graph-based approaches, which not only can capture the overall structural characteristics of the network, but also possess excellent representation and reasoning ability. Li and Li (2022) proposed a sentiment analysis model of Weibo comments based on GNN. This model establishes an interpretable feature space and expresses the significance of GNN semantic parse and Long Short-Time Memory (LSTM) update. AlBadani et al. (2022) built a Sentiment Transformer Graph Convolutional Network (ST-GCN) to study a new graph structure on a heterogeneous graph.

2.2 Eye movement

The effect of eye movement in human language processing is verified by numerous research in psycholinguistics and psychologists in the 20th century and beyond (Hollenstein et al., 2019). It's well proven that lexical processing goes hand in hand with eye movements, not excepting the higher-level syntactic features of text (Staub and Rayner, 2007). Given that humans pay attention to different linguistic features during early and late two different processes, two eye movement measures including early measures and late measures are defined by Clifton et al. (2007). Early measures including

FFD and gaze duration (GD) are susceptible to the early processes of the comprehension of a text in the characters and words characteristic. While late measures such as TRT are considered to be sensitive to the late processes in more complex linguistic features (Roberts and Siyanova-Chanturia, 2013).

Furthermore, recent research on SA based on eye movement has gradually stepped into the limelight. Mishra et al. (2017) introduced a framework to automatically extract cognitive eye movement features of human reading the text and applied them as features alone with textual features to the sentiment analysis and sarcasm detection. Long et al. (2021) proposed an LSTM model enhanced with a cognition-based attention model based on CNN which learns features from both eye movement data and text and applies them to classify the input text. Chen et al. (2022) recorded the eye movement data with an eye tracker device, and built a human behavior-inspired sentiment prediction model.

3 Method

In this section, we describe our proposed sentiment-eye movement guided graph neural network (SEMGraph) in detail. As illustrated in Figure 1, our SEMGraph comprises five components: 1) Embedding module, which utilizes BERT to create embeddings of the vertices. 2) Eye movement feature extraction, which derives eye movement features from the text by the linguistic probing eye movement paradigm. 3) Sentiment-eye movement feature incorporation, which derives sentiment-eye movement weights by incorporating sentiment knowledge and eye movement features. 4) Graph-based information interaction, which adopts GGNN and injects FFD weight to implement word interaction. 5) Attention-based readout function, which designs an attention layer consisting of soft attention and TRT weight to obtain a graph-level representation for sentiment prediction.

3.1 Embedding Module

In this paper, we form a linguistic feature vector for eye movement feature extraction, and create word representations as input of graph-based information interaction.

Give a word $w_i \in S_j$ in sentence S_j, and its linguistic feature vector $V_{w_i} = [f_1^{w_i}, f_2^{w_i}, ..., f_m^{w_i}]$, where the linguistic feature f^{w_i} consists of character feature c^{w_i}, word feature r^{w_i}, and complexity feature o^{w_i} smoothing by min-max normalization (Yu et al., 2009). Besides, we used BERT (Devlin et al., 2019) to generate words representations which capture syntactic and semantic information in vector space, defined as $H \in \mathbb{R}^{|V|\times d}$ where d represents the embedding dimension.

3.2 Eye Movement Feature Extraction

Considering two measures including early and late measures of eye movement, we propose a linguistic probing eye movement paradigm and take FFD from early measures and TRT from late measures into account. Furthermore, we select related linguistic features from the level of character, word, and complexity based on the study from Demberg and Keller (2008) and Tomanek et al. (2010). The selective linguistic features of a word in three levels are shown below:

- Character feature: number of characters, words in phrase starting with a capital letter.
- Word feature: test whether an entity-critical word in annotation phrase, number of senses in wordnet of the word.
- Complexity feature: number of dominated nodes, complexity score referred to the number of words in the sentence, maximum dependency distance.

We use eye movement features as dependent variables and linguistic features as independent variables based on a regression model. A word w_i in sentence S_j contains its linguistic feature vector $V_{w_i} = [f_1^{w_i}, f_2^{w_i}, ..., f_m^{w_i}]$, where the linguistic feature f^{w_i} contains character feature c^{w_i}, word feature r^{w_i}, and complexity feature o^{w_i} which are smoothed by min-max normalization (Yu et al., 2009). We use ridge regression (RR) to implement our paradigm, and a mapping function g between FFD t_{FFD} and TRT t_{TRT} is shown as below.

$$\begin{aligned} t_{FFD,w_i\in S_j} =& g(\alpha_1 c_1^{w_i}, ..., \alpha_k c_k^{w_i}, \alpha_{k+1} r_1^{w_i}, ..., \\ & \alpha_{k+p} r_p^{w_i} + \varepsilon) \\ t_{TRT,w_i\in S_j} =& g(\beta_1 c_1^{w_i}, ..., \beta_k c_k^{w_i}, \beta_{k+1} r_1^{w_i}, ..., \\ & \beta_{k+p} r_p^{w_i}, \beta_{k+p+1} o_1^{w_i}, ..., \\ & \beta_{k+p+q} o_q^{w_i} + \delta) \end{aligned} \tag{1}$$

where t_{FFD} and t_{TRT} are the prediction FFD and TRT for a word $w_i \in S_j$, ε is a constant. Note that k, p, q are the space size for character feature c^{w_i}, word feature r^{w_i}, and complexity feature o^{w_i}.

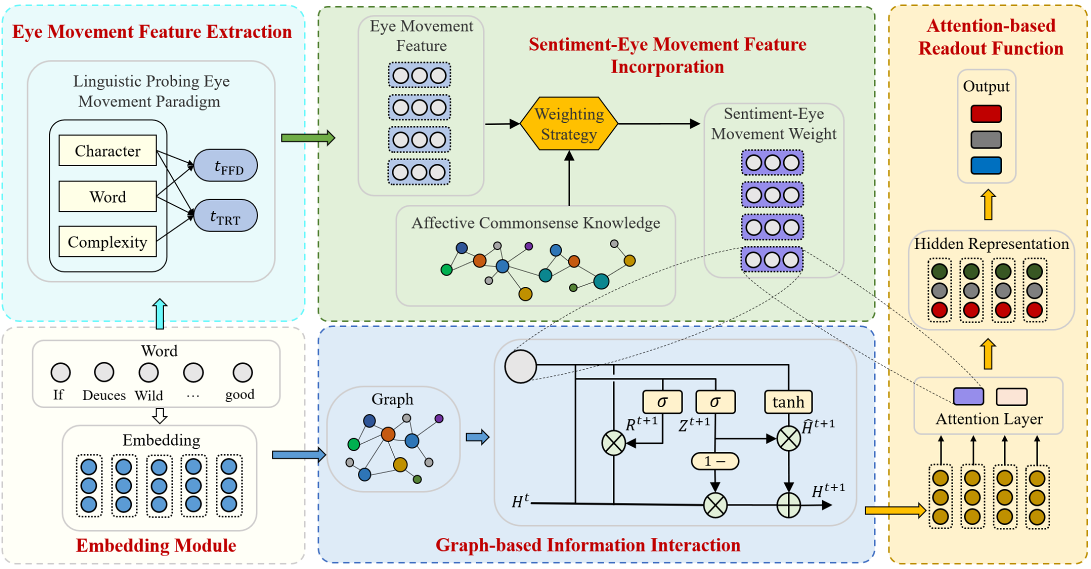

Figure 1: The architecture of our proposed SEMGraph

And the set of $\alpha_i(i = 1, 2, ..., k + p)$ and $\beta_i(i = 1, 2, ..., k + p + q)$ form the weight vector $\vec{\alpha}_{w_i}$ and $\vec{\beta}_{w_i}$ for t_{FFD} and t_{TRT} respectively. The objective function for t_{FFD} and t_{TRT} can be set as follows:

$$\begin{aligned} \boldsymbol{J}_{FFD} &= \arg \underset{\alpha}{min}\{\|t_{FFD} - y_{FFD}\|_2^2 + \eta\|\alpha\|_2^2\} \\ \boldsymbol{J}_{TRT} &= \arg \underset{\beta}{min}\{\|t_{TRT} - y_{TRT}\|_2^2 + \mu\|\beta\|_2^2\} \end{aligned} \tag{2}$$

where y_{FFD} and y_{TRT} are the true eye movement value of FFD and TRT, η and μ are the regularization weight.

3.3 Sentiment-Eye Movement Feature Incorporation

After attaining FFD and TRT eye movement features of each word in sentences, a weighting strategy is introduced to incorporate sentiment knowledge and eye movement features. Specifically, we obtain the sentiment score from SenticNet (Cambria et al., 2010) of each word in the range from 0 (bad) to +2 (good).

For a sentence $S_j = w_1, w_2, ..., w_i, ..., w_{n_j}$ with length n_j, each word w_i in sentence S_j enjoys a corresponding FFD t_{FFD,w_i} and TRT t_{TRT,w_i}, and let W_{FFD} and W_{TRT} denote the weight of FFD and TRT for each word as bellow.

$$\begin{aligned} W_{FFD} &= \frac{t_{FFD,w_i} + x_{w_i}}{\sum_{i=1,w_i \in S_j}^{n_j} (t_{FFD,w_i} + x_{w_i})} \\ W_{TRT} &= \frac{t_{TRT,w_i} + x_{w_i}}{\sum_{i=1,w_i \in S_j}^{n_j} (t_{TRT,w_i} + x_{w_i})} \end{aligned} \tag{3}$$

3.4 Graph-based Information Interaction

We represent a sentence as a statistical word co-occurrence network to construct the graph, denoted as $G = (v, e)$. Each word v in the sentence is represented by a vertice in G. The undirected edge in the graph connects two words that occur within a fixed-size sliding window of default size 3 over-spanning word. The sentence is preprocessed in the same way as Zhang et al. (2020) with stopword removal and tokenization. And we use BERT (Devlin et al., 2019) to initialize the embeddings of the vertices with word feature, defined as $H \in \mathbb{R}^{|V| \times d}$ where d represents the embedding dimension.

Since individual graphs are constructed for each sentence, the word feature information is propagated and merged contextually during the interaction phase. Given the close relation between FFD and word feature, the sentiment-eye movement weight of FFD is added on the basis of word feature information. Furthermore, we adopt the GGNN (Li et al., 2015) to study the word embeddings. A word node can receive the information I comprising eye movement and word feature information from adjacent neighbors, and merge with its representation to update. Referring to Yu et al. (2009) for obtaining the high-order feature interaction, we stack the graph layer operating on the first-order neighbors t times. The detailed information interaction formulas are described as below:

$$
\begin{aligned}
I^{t+1} &= AH^tW_a + AH^tW_{FFD} \\
R^{t+1} &= \sigma(W_rI_{t+1} + U_rH^t + b_r) \\
Z^{t+1} &= \sigma(W_zI_{t+1} + U_zH^t + b_z) \\
\hat{H}^{t+1} &= \tanh(W_hI_{t+1} + U_h(R^{t+1} \odot H^t) + b_h) \\
H^{t+1} &= \hat{H}^{t+1} \odot Z^{t+1} + H^t \odot (1 - Z^{t+1})
\end{aligned} \tag{4}
$$

where $A \in \mathbb{R}^{|v|\times|v|}$ represents the adjacency matrix, and $\sigma(x) = 1/(1 + \exp^{-x})$ is the sigmoid function. W_{FFD} is the FFD weights, and all W except W_{FFD}, U and b are trainable weights and biases. GRU cell deploys an update gate R and a reset gate Z to determine what degree the neighbor information conduces to the node embedding.

3.5 Attention-based Readout Function

After word nodes are sufficiently updated, we obtain a matrix $H^T \in \mathbb{R}^{|v|\times|v|}$ containing a word-level representation for the sentence. The readout function obtains feature representation of the whole graph by aggregating the node features, considering the connection of TRT and complex linguistic features, we define the readout function as:

$$s_v^T = \tanh(W_v^T H_v^T + b_v^T) \tag{5}$$

$$a_v^T = \frac{1}{(1 + \exp(-s_v^T))} \tag{6}$$

$$H_v = (a_v^T + W_{TRT}) \odot \tanh(W_v H_v^T + b_v) \tag{7}$$

$$H_G = \frac{1}{|V|} \sum_{v \in V} H \tag{8}$$

where a_v^T and W_{TRT} are the soft attention weight and TRT weight. we average the weighted word features to form the final graph representation H_G because each word has an impact on the sentence. Last but not least, the prediction of sentiment result is obtained by applying the graph-level vector H_G into a softmax layer, and the loss is minimized through a cross-entropy function:

$$\hat{y}_G = \mathrm{softmax}(WH_G + b) \tag{9}$$

$$L = -\sum_i y_{G_i} \log(\hat{y}_{G_i}) \tag{10}$$

where W and b are weights and bias, and y_{G_i} is one-hot label of the i^{th} sentence.

Dataset	Sentence	Token	Participant
Provo	134	56212	84
GECO	5424	68606	17
Mishra	994	68543	7
Zuco	1100	21629	12

Table 1: Statistics of four eye movement corpora.

Dataset	Class	Avg.vocab	Max.vocab	Train	Test
MR	2	18.46	46	9596	1066
SST-2	2	9.5	53	67349	1821
STS-Gold	2	14.84	31	1831	203

Table 2: Statistics of three sentiment datasets. The vocab means the number of unique words in a sentence.

4 Experiment

4.1 Dataset

Eye Movement Dataset: In this work, we use four high-quality eye movement datasets established from online reviews involving Provo (Luke and Christianson, 2018), the Ghent Eye-Tracking Corpus (GECO) (Cop et al., 2017), Mishra's dataset abbreviated as Mishra (Mishra et al., 2016), and ZuCo (Hollenstein et al., 2018). Their lengths of sentences, tokens, and the number of participants are listed in Table 1. It's remarkable that Mishra and ZuCo both have eye movement signals and sentiment labels, and there are 1,100 English sentences including 399 sentiment sentences in Zuco.

Sentiment Dataset: To explore the influence of eye movement data for the sentiment analysis task, three sentiment datasets were used. The first dataset MR (Pang and Lee, 2005) is a binary-category dataset including positive and negative movie review sentences collected from the Rotten Tomatoes website. SST-2 (Socher et al., 2013) is a public dataset of movie reviews with binary classes from the Stanford Sentiment Treebank. The Stanford Twitter Sentiment Gold (STS-Gold) (Saif et al., 2013) dataset consists of 2034 tweets with 632 positive tweets and 1402 negative tweets. The details of these three datasets are shown in Table 2.

4.2 Experimental Set-up

The experiment was mainly conducted on three sentiment datasets without eye movement signals. We first extract the eye movement features of FFD and TRT based on our proposed linguistic probing eye movement paradigm from the four eye movement datasets. Especially, the eye movement datasets record the eye movement signals from several par-

Model	MR	SST2	STS-Gold
CNN-static-fastText (Ouyang et al., 2015)	85.20	84.50	-
LSTM (Xiang et al., 2021)	76.70	80.20	-
BiLSTM-AT (Xiang et al., 2021)	76.90	79.50	-
Single-layered BiLSTM (Hameed and Garcia-Zapirain, 2020)	80.50	85.78	-
BiGRU+CNN (Zhang et al., 2018)	78.30	85.40	-
CNN-GRU-multilevel and multitype fusion (Usama et al., 2019)	80.20	85.70	-
SentiStrength (Krouska et al., 2017)	-	-	82.10
DeepCNN (Jianqiang et al., 2018)	-	-	86.00
One-layer CNN (Wang, 2021)	-	-	86.50
MC-CNN (Islam et al., 2019)	-	-	90.70
EMTCNN (Wang, 2021)	-	-	90.90
BERT+SG-OPT (Kim et al., 2021)	82.47	86.20	-
KESA (Zhao et al., 2022b)	86.29	91.56	-
BESA (Yang et al., 2021)	84.30	92.40	-
BT-TAPT (Lee et al., 2021)	86.40	92.40	-
BERT (Kodiyala and Mercer, 2021)	-	-	93.60
GNN (Li et al., 2015; Zhang et al., 2020)	83.91	92.13	91.63
SEMGraph-G	85.96	93.08	92.89
SEMGraph-Z	86.12	93.01	92.58
SEMGraph-M	86.54	93.34	93.42
SEMGraph-P	**87.42**	**94.23**	**94.78**

Table 3: Accuracy (%) of different methods on the three sentiment datasets.

ticipants in the same sentence. We thus use the box plot method (Kwak and Kim, 2017) to identify the outliers which lie outside the upper or lower fence lines. After removing the outliers, we average eye movement data of each word of each participant. Then, we incorporate sentiment knowledge and eye movement features to obtain the sentiment-eye movement weight and build a sentiment-eye movement guided graph. We randomly split the training set into actual training and validation in the ratio of 9:1. The hyperparameters are tuned relying on the validation set. The learning rate is set to 0.0005 with Adam (Kingma and Ba, 2014) optimizer, the dropout is set as 0.2. Concerning the word embeddings, we adopt the BERT (Devlin et al., 2019) with 768 as the input features[1].

4.3 Comparison Models

A few mainstream models for SA without eye movement signals are applied for comparison including deep learning models without BERT pretrained model, BERT-based models, and our model without eye movement data and its variations.

Deep learning models: CNN-static-fastText (Ouyang et al., 2015), LSTM (Xiang et al., 2021), BiLSTM-AT (Xiang et al., 2021), Single-layered BiLSTM (Hameed and Garcia-Zapirain, 2020), BiGRU+CNN (Zhang et al., 2018), CNN-GRU-multilevel and multitype fusion (Usama et al., 2019) of MR and SST-2 dataset, SentiStrength (Krouska et al., 2017), DeepCNN (Jianqiang et al., 2018), One-layer CNN (Wang, 2021), MC-CNN (Islam et al., 2019), and EMTCNN (Wang, 2021) of STS-Gold dataset.

BERT-based models: BERT+SG-OPT (Kim et al., 2021), KESA (Zhao et al., 2022b), BESA (Yang et al., 2021) and BT-TAPT (Lee et al., 2021) of MR and SST-2 dataset. BERT (Kodiyala and Mercer, 2021) of STS-Gold dataset.

GNN: Our proposed model without eye movement data (Li et al., 2015; Zhang et al., 2020).

Our models: Three variations involving SEMGraph-G, SEMGraph-Z, SEMGraph-M, and SEMGraph-P (G, Z, M, P represent the eye movement features derived from GECO, Zuco, Mishra and Provo datasets)

4.4 Main Results

The results of the baselines and our models are presented in Table 3, from which several observations

[1]The source code of this work is released at: `https://github.com/HITSZ-HLT/SEMGraph`

can be noted. First, the BERT-based models generally outperform the other types of baseline models, because BERT performs better in sentences with high complexity and sentences with many ambiguous words as the three datasets in our work. Further, our proposed models have outstanding performance, suggesting that the eye movement data benefits the sentiment analysis. Particularly, the results of SEMGraph-P are remarkably higher on three datasets since the better implementation of paradigm in the Provo dataset, which would be described in section 4.9.

4.5 Statistical Significance Tests

In order to check whether the differences in performance among the baseline models without eye movement data and our model are statistically significant, we perform the pairwise two-tailed t-test on GNN and three variations of our model in three datasets. The difference between the GNN and SEMGraph-M is statistically significant at an alpha level of 0.05 with a p-value of 0.045, and the p-value is 0.049 between GNN and SEMGraph-Z. Moreover, the difference between the GNN and SEMGraph-P is significant at an alpha level of 0.05 with a p-value of 0.02. However, it can be found that the difference between the GNN and SEMGraph-G is statistically not very significant, where the p-value is 0.052.

4.6 Ablation Study

To understand the influence of the sentiment-eye movement weight, we further conduct an ablation study. We present the results on the SEMGraph-P model without eye movement features and sentiment knowledge for comparison. From Table 4, we can observe that SEMGraph-P is improved by using eye movement and sentiment knowledge on three sentiment datasets without eye movement signals. Compared with GNN, the performance of SEMGraph-P without either FFD or TRT achieve higher accuracy, which validates that adding the eye movement signals including FFD and TRT can effectively enhance the performance of SA. Furthermore, sentiment knowledge of each word allows our model to be robust, suggesting that the sentiment-eye movement features are of importance for the sentiment of a sentence.

Model	MR	SST-2	STS-Gold
GNN	83.91	92.13	91.63
SEMGraph-P	87.42	94.23	94.78
- w/o FFD	84.23	93.21	93.10
- w/o TRT	85.31	93.43	92.09
- w/o sentiment score	86.34	93.56	93.61

Table 4: Ablation study on the importance of sentiment-eye movement features. w/o means without.

4.7 Visualization

As shown in Figure 2, we visualize sentiment-eye movement features including FFD and TRT, and some words are fixated upon more time (dark blue) and others lesser (light blue). The highlighted words are proportional to the attention weights and they present a close correlation to the sentiment label, which interprets how sentiment-eye movement features work in sentiment analysis. Besides, FFD is not as obvious as that of TRT, which suggests that a person would read the words related to sentiment repeatedly.

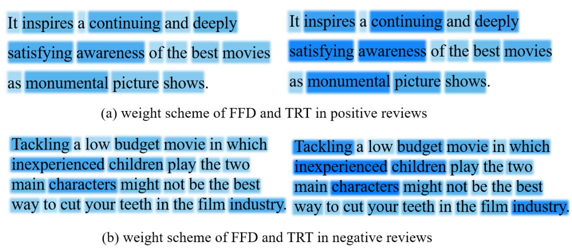

Figure 2: Visualizations of positive and negative samples in MR dataset.

4.8 Performance in Eye Movement Sentiment Dataset

To analyze our proposed method is also well-behaved in the existing sentiment analysis datasets with eye movement signals, tenfold cross-validation was performed on Mishra and ZuCo two eye movement sentiment datasets with available sentiment labels. Rather than extracting eye movement features, we utilize the FFD and TRT eye movement features provided by the datasets themselves. Table 5 compares our approach with the previous results which employ the eye movement feature as attention, and reports the precision (P), recall (R), and F1 score (F1). As is shown, the common application of eye movement features and sentiment knowledge is beneficial to sentiment analysis, and our proposed method achieves strong performance both in sentiment datasets with eye movement signals and without eye movement signals.

Model	Mishra			ZuCo		
	P	R	F1	P	R	F1
SVM (Mishra et al., 2016)	73.30	73.60	73.50	-	-	-
Multi-Task learning (Majumder et al., 2019)	83.67	83.10	83.03	-	-	-
LSTM layer (Hollenstein et al., 2019)	-	-	-	59.80	60.00	59.80
BERT (Zhao et al., 2022a)	-	-	-	73.46	74.10	72.54
GNN	81.42	80.81	81.03	73.02	71.79	71.32
SEMGraph	**85.00**	**84.85**	**84.91**	**78.57**	73.33	**75.86**

Table 5: Results (%) of different methods on sentiment analysis datasets with eye movement signals.

Eye movement feature	Datasets	LSTM	SVM	RNN	GRU	RR
FFD	Provo	55.7678	42.3335	55.7321	59.3299	30.7693
	GECO	79.4685	75.3986	54.9739	79.4847	44.1300
	Mishra	54.9739	57.2110	54.9660	54.9659	53.9477
	Zuco	56.7743	54.9552	56.7261	54.1131	52.2803
TRT	Provo	103.3025	89.7172	103.3468	105.4021	57.0098
	GECO	102.8497	106.5619	101.9863	101.8437	100.3379
	Mishra	167.7762	161.5456	101.5186	167.7809	101.2218
	Zuco	103.2423	98.8732	104.2341	103.2539	93.0365

Table 6: RMSE of different methods for the implementation of Paradigm.

4.9 Effectiveness of Eye Movement Feature Extraction

To compare the effectiveness of different methods for eye movement feature extraction based on our proposed paradigm, we experiment on the four eye movement datasets, and report the root mean square error (RMSE) as the indicator, which means a statistic value bringing about large with either too large or too small deviation from average.

We take 90 percent of sentences as training data and the rest 10 percent as test data. We compare the RR model with more complex machine learning methods including LSTM, SVM, RNN, and GRU in FFD and TRT data of four eye movement datasets. The best performance results for each model are shown in Table 6. It's obvious from Table 6 that the RR model outperforms other machine learning methods benefiting from the regularization in RR to alleviate the overfitting problem. The linguistic features, their types and the corresponding coefficients in RR based on three eye movement datasets are illustrated in Table 7.

Feature	Linguistic feature	Type	Coefficient
FFD	Constant	Num	48.0787
	Number of characters	Num	14.4777
	Start with capital letter	Bool	-5.3184
	Is entity critical word	Bool	29.8971
	Number of senses in wordnet	Num	0.3074
TRT	Constant	Num	38.4023
	Number of characters	Num	27.4164
	Start with capital letter	Bool	-9.2777
	Is entity critical word	Bool	34.5984
	Number of senses in wordnet	Num	0.2100
	Number of dominated nodes	Num	-0.5454
	Complexity score	Num	0.0792
	Max dependency distance	Num	0.6010

Table 7: Eye movement and linguistic features used for RR method (Num stands for numerical type and Bool stands for Boolean type).

5 Conclusion

In the current work, we investigate the SA task from a novel perspective by incorporating sentiment knowledge and eye movement into a graph architecture. Moreover, a linguistic probing eye movement paradigm is proposed to extract eye movement features based on the close connection between linguistic features and the processes of human reading. We then devise a novel weighting strategy to obtain sentiment-eye movement weights based on sentiment scores extracted from affective commonsense knowledge and eye movement features, which are exploited to build the sentiment-eye movement guided graph. Experimental results on two eye movement sentiment analysis datasets with eye movement signals and three sentiment analysis datasets without eye movement signals demonstrated the effectiveness and excellent generalization of our proposed method.

Limitations

This work proposes a paradigm based on the existing eye movement data instead of capturing human eye movement data online with an eye tracker. Therefore, one direction for future work is to explore online sentiment analysis with eye tracker. Secondly, we mainly focus on FFD and TRT among the most eye movement data. It would be better if more useful eye movement data can be introduced to study the relationship between eye movement and sentiment in more depth.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China 62006062 and 62176076, Shenzhen Foundational Research Funding JCYJ20200109113441941, JCYJ20210324115614039, The Major Key Project of PCL2021A06, Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies 2022B1212010005.

References

Barakat AlBadani, Ronghua Shi, Jian Dong, Raeed Al-Sabri, and Oloulade Babatounde Moctard. 2022. Transformer-based graph convolutional network for sentiment analysis. *Applied Sciences*, 12(3).

Salima Behdenna, Fatiha Barigou, and Ghalem Belalem. 2016. Sentiment analysis at document level. In *Smart Trends in Information Technology and Computer Communications*, pages 159–168, Singapore. Springer Singapore.

Rajesh Bose, Raktim Dey, Sandip Roy, and Debabrata Sarddar. 2019. *Sentiment Analysis on Online Product Reviews*, pages 559–569. Springer, Singapore.

Erik Cambria, Robyn Speer, Catherine Havasi, and Amir Hussain. 2010. Senticnet: A publicly available semantic resource for opinion mining. In *2010 AAAI fall symposium series*.

Xuesong Chen, Jiaxin Mao, Yiqun Liu, Min Zhang, and Shaoping Ma. 2022. Investigating human reading behavior during sentiment judgment. *International Journal of Machine Learning and Cybernetics*, pages 1–14.

Youngseok Choi and Habin Lee. 2017. Data properties and the performance of sentiment classification for electronic commerce applications. *Information Systems Frontiers*, 19(5):993–1012.

Charles Clifton, Adrian Staub, and Keith Rayner. 2007. Eye movements in reading words and sentences. In Roger P.G. Van Gompel, Martin H. Fischer, Wayne S. Murray, and Robin L. Hill, editors, *Eye Movements*, pages 341–371. Elsevier, Oxford.

Uschi Cop, Nicolas Dirix, Denis Drieghe, and Wouter Duyck. 2017. Presenting geco: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior research methods*, 49(2):602–615.

Cach N. Dang, María N. Moreno-García, and Fernando De la Prieta. 2021. An approach to integrating sentiment analysis into recommender systems. *Sensors*, 21(16).

Vera Demberg and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Hai Ha Do, PWC Prasad, Angelika Maag, and Abeer Alsadoon. 2019. Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Systems with Applications*, 118:272–299.

Zabit Hameed and Begonya Garcia-Zapirain. 2020. Sentiment classification using a single-layered bilstm model. *Ieee Access*, 8:73992–74001.

Nora Hollenstein, Maria Barrett, Marius Troendle, Francesco Bigiolli, Nicolas Langer, and Ce Zhang. 2019. Advancing nlp with cognitive language processing signals. *arXiv preprint arXiv:1904.02682*.

Nora Hollenstein, Jonathan Rotsztejn, Marius Troendle, Andreas Pedroni, Ce Zhang, and Nicolas Langer. 2018. Zuco, a simultaneous eeg and eye-tracking resource for natural sentence reading. *Scientific data*, 5(1):1–13.

Alice M. Isen and Barbara Means. 1983. The influence of positive affect on decision-making strategy. *Social Cognition*, 2(1):18–31.

Jumayel Islam, Robert E. Mercer, and Lu Xiao. 2019. Multi-channel convolutional neural network for Twitter emotion and sentiment recognition. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1355–1365, Minneapolis, Minnesota. Association for Computational Linguistics.

Zhao Jianqiang, Gui Xiaolin, and Zhang Xuejun. 2018. Deep convolution neural networks for twitter sentiment analysis. *IEEE Access*, 6:23253–23260.

Aditya Joshi, Abhijit Mishra, Nivvedan Senthamilselvan, and Pushpak Bhattacharyya. 2014. Measuring sentiment annotation complexity of text. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 36–41.

Shengfeng Ju and Shoushan Li. 2013. Active learning on sentiment classification by selecting both words and documents. In *Chinese Lexical Semantics*, pages 49–57, Berlin, Heidelberg. Springer Berlin Heidelberg.

Taeuk Kim, Kang Min Yoo, and Sang-goo Lee. 2021. Self-guided contrastive learning for BERT sentence representations. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2528–2540, Online. Association for Computational Linguistics.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. *CoRR*, abs/1408.5882.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Vishwa Sai Kodiyala and Robert E. Mercer. 2021. Emotion recognition and sentiment classification using bert with data augmentation and emotion lexicon enrichment. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 191–198.

Akrivi Krouska, Christos Troussas, and Maria Virvou. 2017. Comparative evaluation of algorithms for sentiment analysis over social networking services. *J. Univers. Comput. Sci.*, 23(8):755–768.

Sudhanshu Kumar, Mahendra Yadava, and Partha Pratim Roy. 2019. Fusion of eeg response and sentiment analysis of products review to predict customer satisfaction. *Information Fusion*, 52:41–52.

Sang Kyu Kwak and Jong Hae Kim. 2017. Statistical data preparation: management of missing values and outliers. *Korean journal of anesthesiology*, 70(4):407.

Junghoon Lee, Jounghee Kim, and Pilsung Kang. 2021. Back-translated task adaptive pretraining: Improving accuracy and robustness on text classification. *arXiv preprint arXiv:2107.10474*.

Changliang Li, Bo Xu, Gaowei Wu, Saike He, Guanhua Tian, and Hongwei Hao. 2014. Recursive deep learning for sentiment analysis over social data. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, volume 2, pages 180–185.

Weijiang Li, Fang Qi, Ming Tang, and Zhengtao Yu. 2020. Bidirectional lstm with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing*, 387:63–77.

Yan Li and Nianfeng Li. 2022. Sentiment analysis of weibo comments based on graph neural network. *IEEE Access*, 10:23497–23510.

Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. 2015. Gated graph sequence neural networks. *arXiv preprint arXiv:1511.05493*.

Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. 2015. Gated Graph Sequence Neural Networks. *arXiv e-prints*, page arXiv:1511.05493.

Bin Liang, Xiang Li, Lin Gui, Yonghao Fu, Yulan He, Min Yang, and Ruifeng Xu. 2022a. Few-shot aspect category sentiment analysis via meta-learning. *ACM Transactions on Information Systems (TOIS)*.

Bin Liang, Hang Su, Lin Gui, Erik Cambria, and Ruifeng Xu. 2022b. Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowledge-Based Systems*, 235:107643.

Bin Liang, Rongdi Yin, Jiachen Du, Lin Gui, Yulan He, Min Yang, and Ruifeng Xu. 2021. Embedding refinement framework for targeted aspect-based sentiment analysis. *IEEE Transactions on Affective Computing*.

Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, Cambridge, UK.

Yunfei Long, Rong Xiang, Qin Lu, Chu-Ren Huang, and Minglei Li. 2021. Improving attention model based on cognition grounded data for sentiment analysis. *IEEE Transactions on Affective Computing*, 12(4):900–912.

Steven G Luke and Kiel Christianson. 2018. The provo corpus: A large eye-tracking corpus with predictability norms. *Behavior research methods*, 50(2):826–833.

Navonil Majumder, Soujanya Poria, Haiyun Peng, Niyati Chhaya, Erik Cambria, and Alexander Gelbukh. 2019. Sentiment and sarcasm classification with multitask learning. *IEEE Intelligent Systems*, 34(3):38–43.

Abhijit Mishra, Kuntal Dey, and Pushpak Bhattacharyya. 2017. Learning cognitive features from gaze data for sentiment and sarcasm classification using convolutional neural network. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 377–387, Vancouver, Canada. Association for Computational Linguistics.

Abhijit Mishra, Diptesh Kanojia, Seema Nagar, Kuntal Dey, and Pushpak Bhattacharyya. 2016. Leveraging cognitive features for sentiment analysis. *CoNLL 2016*, page 156.

Xi Ouyang, Pan Zhou, Cheng Hua Li, and Lijun Liu. 2015. Sentiment analysis using convolutional neural network. In *2015 IEEE International Conference on*

Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, pages 2359–2364.

Bo Pang and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 115–124.

Bo Pang, Lillian Lee, et al. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135.

Adil Rajput. 2019. Natural Language Processing, Sentiment Analysis and Clinical Analytics. *arXiv e-prints*, page arXiv:1902.00679.

Leah Roberts and Anna Siyanova-Chanturia. 2013. Using eye-tracking to investigate topics in l2 acquisition and l2 processing. *Studies in Second Language Acquisition*, 35(2):213–235.

Hassan Saif, Miriam Fernández, Yulan He, and Harith Alani. 2013. Evaluation datasets for twitter sentiment analysis: a survey and a new dataset, the sts-gold. In *1st Interantional Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and Perspectives from AI (ESSEM 2013)*.

Björn Schuller, Amr El-Desoky Mousa, and Vasileios Vryniotis. 2015. Sentiment analysis and opinion mining: on optimal parameters and performances. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(5):255–263.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642.

Adrian Staub and Keith Rayner. 2007. Eye movements and on-line comprehension processes. *The Oxford handbook of psycholinguistics*, 327:342.

Justin Storbeck and Gerald L. Clore. 2007. On the interdependence of cognition and emotion. *Cognition and Emotion*, 21(6):1212–1237. PMID: 18458789.

Katrin Tomanek, Udo Hahn, Steffen Lohmann, and Jürgen Ziegler. 2010. A cognitive cost model of annotations based on eye-tracking data. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1158–1167.

Mohd Usama, Wenjing Xiao, Belal Ahmad, Jiafu Wan, Mohammad Mehedi Hassan, and Abdulhameed Alelaiwi. 2019. Deep learning based weighted feature fusion approach for sentiment analysis. *IEEE Access*, 7:140252–140260.

Quansen Wang. 2021. Learning from other labels: Leveraging enhanced mixup and transfer learning for twitter sentiment analysis. In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 336–343.

Rong Xiang, Emmanuele Chersoni, Qin Lu, Chu-Ren Huang, Wenjie Li, and Yunfei Long. 2021. Lexical data augmentation for sentiment analysis. *Journal of the Association for Information Science and Technology*, 72(11):1432–1447.

Xinghao Yang, Weifeng Liu, Dacheng Tao, and Wei Liu. 2021. Besa: Bert-based simulated annealing for adversarial text attacks. In *International Joint Conference on Artificial Intelligence*.

Liping Yu, Yuntao Pan, and Yishan Wu. 2009. Research on data normalization methods in multi-attribute evaluation. In *2009 International Conference on Computational Intelligence and Software Engineering*, pages 1–5.

Dejun Zhang, Long Tian, Mingbo Hong, Fei Han, Yafeng Ren, and Yilin Chen. 2018. Combining convolution neural network and bidirectional gated recurrent unit for sentence semantic classification. *IEEE Access*, 6:73750–73759.

Yufeng Zhang, Xueli Yu, Zeyu Cui, Shu Wu, Zhongzhen Wen, and Liang Wang. 2020. Every document owns its structure: Inductive text classification via graph neural networks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 334–339, Online. Association for Computational Linguistics.

Lei Zhao, Yingyi Zhang, and Chengzhi Zhang. 2022a. Does attention mechanism possess the feature of human reading? a perspective of sentiment classification task. *Aslib Journal of Information Management*, (ahead-of-print).

Qinghua Zhao, Shuai Ma, and Shuo Ren. 2022b. Kesa: A knowledge enhanced approach for sentiment analysis. *arXiv preprint arXiv:2202.12093*.

Arman S. Zharmagambetov and Alexandr A. Pak. 2015. Sentiment analysis of a document using deep learning approach and decision trees. In *2015 Twelve International Conference on Electronics Computer and Computation (ICECCO)*, pages 1–4.

Shusen Zhou, Qingcai Chen, and Xiaolong Wang. 2014. Fuzzy deep belief networks for semi-supervised sentiment classification. *Neurocomputing*, 131:312–322.