

情感分析-评论挖掘

董天智 2017100937

2018/6/12

- 绪论
- 相关工作
- 模型
- 实验结果

绪论

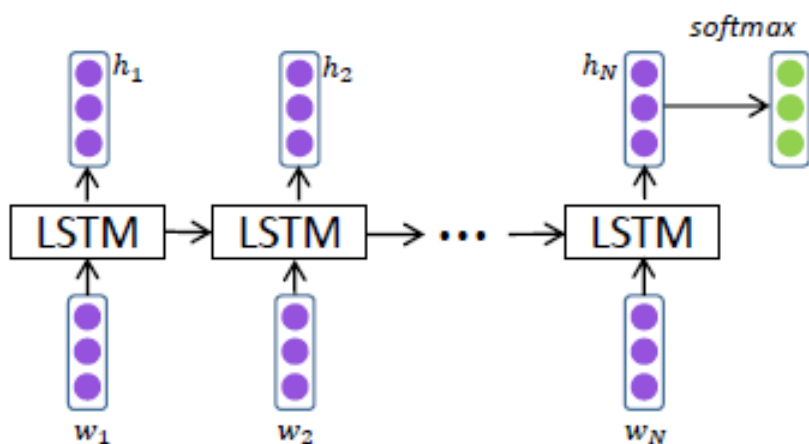
- 情感极性同时取决于评论内容和视角
--“服务员态度不好，但是口味还是不错的。”
- 带视角的情感更有价值

相关工作

- 传统的视角级别情感分析方法：
基于词典；
基于句法分析；
- 利用深度学习的视角级别情感分析方法：
Target-Dependent LSTM

模型

- 1、Long Short-term Memory (LSTM)



$$X = \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \quad (1)$$

$$f_t = \sigma(W_f \cdot X + b_f) \quad (2)$$

$$i_t = \sigma(W_i \cdot X + b_i) \quad (3)$$

$$o_t = \sigma(W_o \cdot X + b_o) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c \cdot X + b_c) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

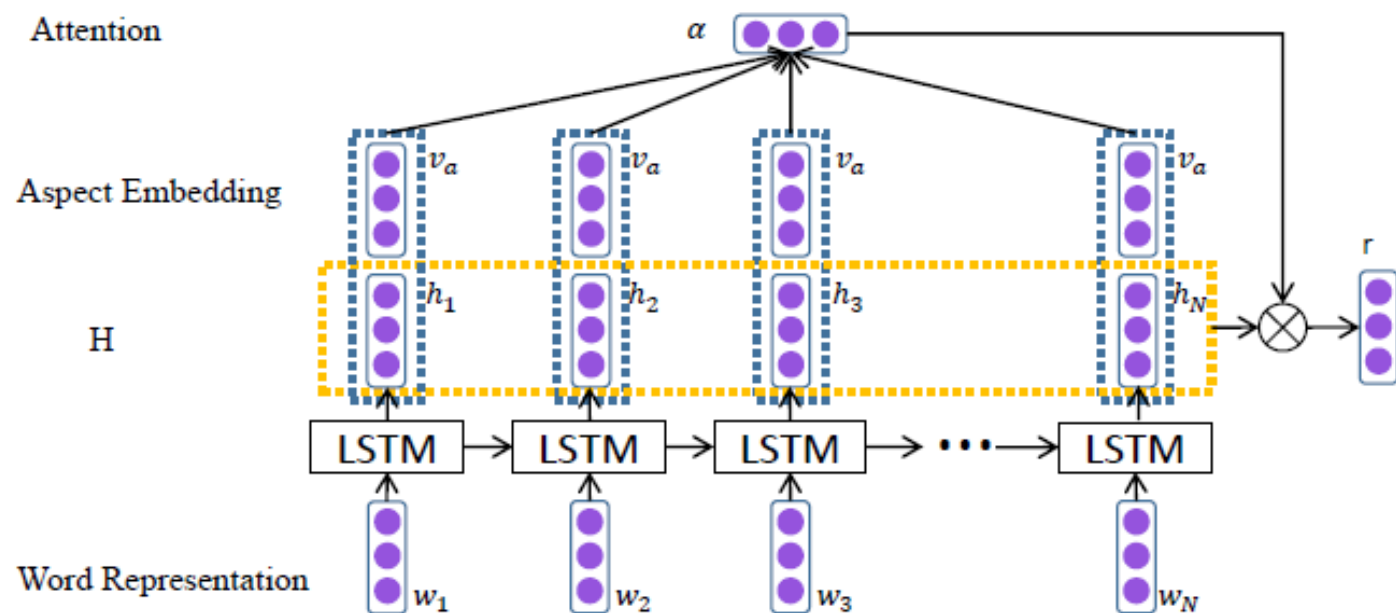
```
h, c = T.zeros_like(self.bf, dtype=theano.config.floatX), T.zeros_like(self.bc, dtype=theano.config.floatX)
```

```
def encode(x_t, h_fore, c_fore, tar_vec):  
    v = T.concatenate([h_fore, x_t, tar_vec])  
    f_t = T.nnet.sigmoid(T.dot(self.Wf, v) + self.bf)  
    i_t = T.nnet.sigmoid(T.dot(self.Wi, v) + self.bi)  
    o_t = T.nnet.sigmoid(T.dot(self.Wo, v) + self.bo)  
    c_next = f_t * c_fore + i_t * T.tanh(T.dot(self.Wc, v) + self.bc)  
    h_next = o_t * T.tanh(c_next)  
    return h_next, c_next
```

```
scan_result, _ = theano.scan(fn=encode, sequences=[self.seq_matrix], outputs_info=[h, c], non_sequences=[self.tar_vector])  
embedding = scan_result[0]
```

模型

- 2、Attention-based LSTM (AT-LSTM)



Attention:

$$M = \tanh\left(\begin{bmatrix} W_h H \\ W_v v_a \otimes e_N \end{bmatrix}\right) \quad (7) \text{ projected hidden matrix}$$

$$\alpha = \text{softmax}(w^T M) \quad (8) \text{ attention weights vector}$$

$$r = H\alpha^T \quad (9) \text{ weighted hidden representation}$$

Sentence representation:

$$h^* = \tanh(W_p r + W_x h_N) \quad (10) \text{ feature representation of a sentence given an input aspect}$$

Classification:

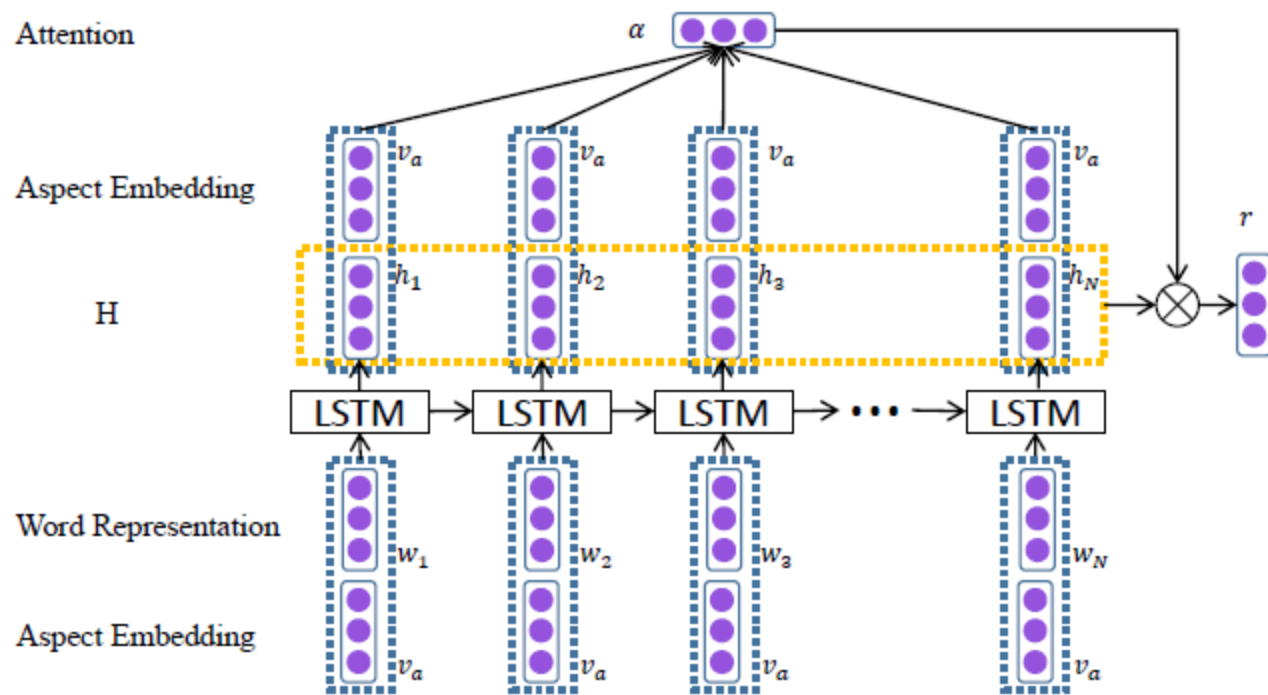
$$y = \text{softmax}(W_s h^* + b_s) \quad (11) \text{ probability distribution}$$

```
# attention
matrix_aspect = T.zeros_like(embedding, dtype=theano.config.floatX)[:,:self.dim_aspect] + self.tar_vector
hhhh = T.concatenate([T.dot(embedding, self.Wh), T.dot(matrix_aspect, self.Wv)], axis=1)
M_tmp = T.tanh(hhhh)
alpha_tmp = T.nnet.softmax(T.dot(M_tmp, self.w))
r = T.dot(alpha_tmp, embedding)
h_star = T.tanh(T.dot(r, self.Wp) + T.dot(embedding[-1], self.Wx))
```

```
self.pred_for_train = T.nnet.softmax(T.dot(embedding_for_train, self.Ws) + self.bs)
self.pred_for_test = T.nnet.softmax(T.dot(embedding_for_test, self.Ws) + self.bs)
```

模型

- Attention-based LSTM with Aspect Embedding (ATAE-LSTM)



实验结果

- 我们标注了约18000条商品评论

Asp.	Positive		Negative		Neural	
	Train	Test	Train	Test	Train	Test
快递运费	211	21	1370	165	3578	115
快递员服务	1389	150	763	75	3525	112
快递速度	1744	186	885	100	3260	95
发货速度	955	123	574	70	3522	118
客服体验	1047	115	1591	166	3254	100
安装体验	1503	189	1144	134	3265	100
退换货体验	302	32	686	73	3790	113
包装体验	968	114	752	93	3637	106
Total	8119	930	7765	876	27831	859

- Accuracy: 87.3%

1: 发货速度: 发货快性价比高: 发货快性价比高, 值得信赖的牌子。

0: 快递速度: None: 理都不理我, 也是上火。

-1: 快递员服务: 分钟就能到, 可我等了一个多, 上楼: 快递师傅提前打电话说20分钟就能到, 可我等了一个多小时才到, 而且还不给送上楼!

1: 安装体验: 安装很简单也便宜: 京东快递就是快啊风扇安装很简单也便宜风力很大

1: 发货速度: 速度超快, 的: 价格实惠, 发货速度超快, 很满意的一次网购

0: 包装体验: None: 后来又说星期一, 再后来直接打网点电话不接, 一天以后打通了直接跟我说让我退货选别家吧, 按不了, 要安装得等一个礼拜再看情况!

-1: 包装体验: 超级烂, 二次包装, 会坑死: 包装超级烂, 二次包装, 买了就会坑死你, 坑的不要不要的。

0: 发货速度: None: 规格不达标, 联系未回复

1: 安装体验: 棒…, 态度好: 安装师傅超级棒……办事认真, 态度好!

1: 快递员服务: 及时也很贴心, 四楼: 送货师傅很及时也很贴心, 直接抗到了四楼。

- 谢谢！