Interactive Attention Transfer Network for Cross-domain Sentiment Classification

Kai Zhang[†], Hefu Zhang[†], Qi Liu[†§*], Hongke Zhao[†], Hengshu Zhu[‡], Enhong Chen[†§]
[†]Anhui Province Key Laboratory of Big Data Analysis and Application, University of Science and Technology of China
[§]School of Data Science, University of Science and Technology of China
[‡]Baidu Talent Intelligence Center, Baidu Inc
{sa517494, zhf2011, zhhk}@mail.ustc.edu.cn, {qiliuql, cheneh}@ustc.edu.cn, zhuhengshu@gmail.com

Kai Zhang 2018/1/28

Outline

1 Background
2 Related Work
3 IATN Framework
4 Experiment
5 Conclusion

01 Background

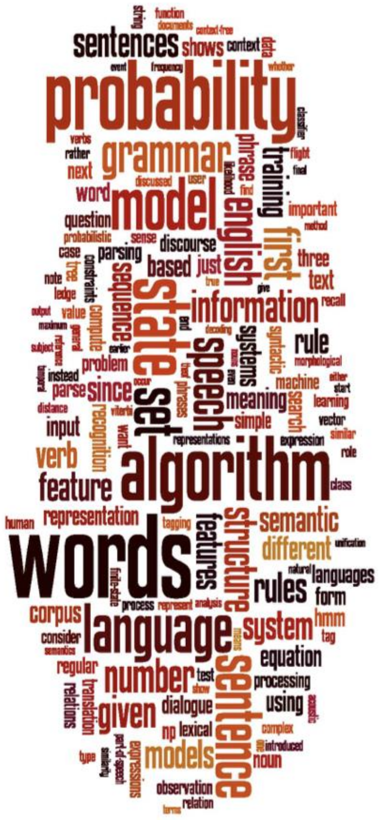
Background

- **What is Sentiment classification ?**
  - Opinion mining
  - Opinion extraction
  - **Sentiment analysis**
  - Sentiment mining

- **Words**
- **Pictures**
- **Behaviors**

?

Attitude



Emotion Tendency

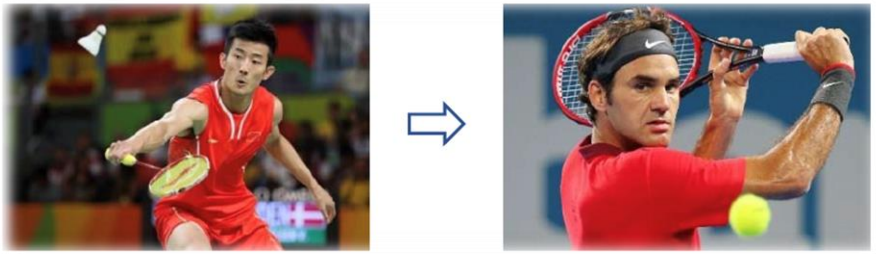
Use of natural language processing (NLP) and computational techniques to automate the extraction or classification of sentiment from typically unstructured text.

Background

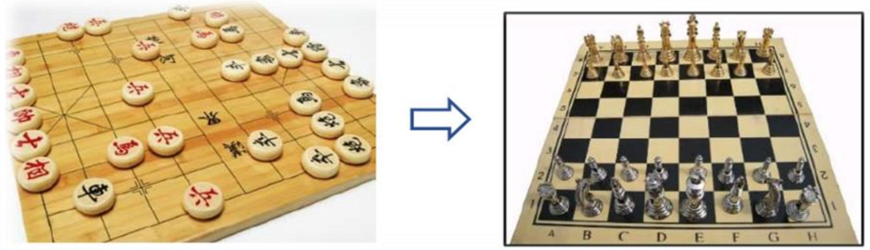
- **What is Transfer Learning?**
  - Refers to the use of similarities between tasks to apply the knowledge learned in the old field to a new field.

Advantages :

- Massive unlabeled data
- Cold start
- Model versatility
- weak computing …



Source → Target



Domain:

- Source: with labeled data
- Target: less or no labeled data

不同领域、不同背景下的
文本翻译、舆情分析

语料匮乏条件下不同语言
的相互翻译学习

不同视角、不同背景、不同光照
的图像识别

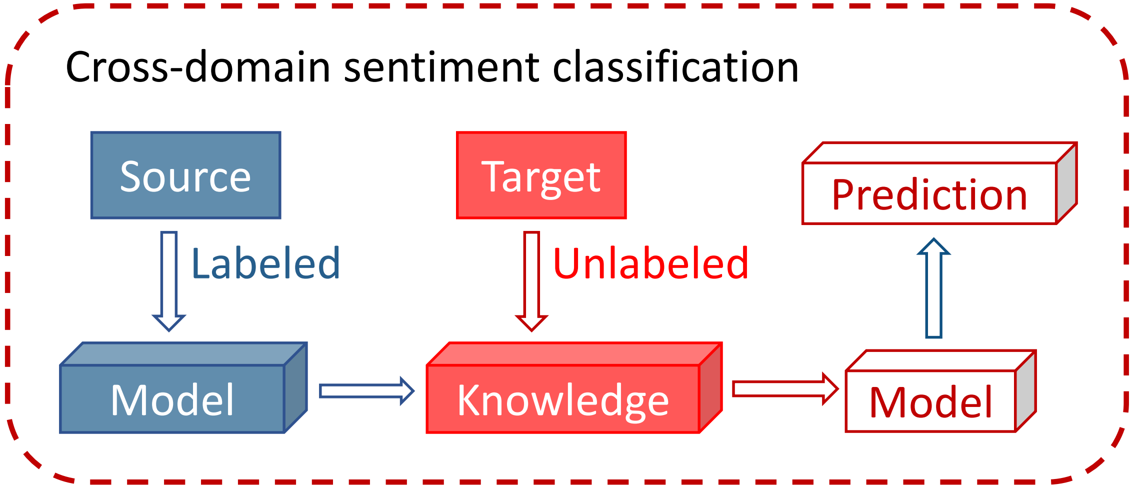
不同用户、不同设备、不同位置
的行为识别

Background

- **Cross-domain Sentiment Classification**
  - Refers to the use of similarities between text classification tasks to apply the knowledge learned in the source domain to the target domain.

Problem : Unsupervised Sentiment Classification

- From: labeled source domain data
- To: sentiment label (target domain)
- Focus on short-text mining and transfer learning methods.



02 Related Work

Related Work

- Past researches can be categorized into two perspectives.
  - Traditional (Theoretical)[Blitzer 2006; Pan et al. 2010; Chen et al. 2012)]
    - Analyzing the major features across domains.
  - Deep-learning Networks[Ganin et al. 2016, Li et al. 2017; 2018b]
    - Modeling the common sentiment relations between different domains.

> Domain shared-feature(pivots)
> Memory Network
> Hierarchical Attention Network

However, most of the previous efforts ignore the characteristics which do not express the sentiment directly, such as the individual modeling of the aspect.

Related Work

- Passed researches can be categorized as two perspectives.
  - Traditional (Theoretical)[Blitzer 2006; Pan et al. 2010; Chen et al. 2012)]
    - Analyzing the major features across domains.
  - Deep-learning Networks[Ganin et al. 2016, Li et al. 2017; 2018b]
    - Modeling the common relations between different domains.

- Domain shared-feature(pivots)
- Memory Network
- Hierarchical Attention Network

However, most of the previous efforts ignore the characteristics which do not express the sentiment directly,

For Example: "the voice quality of iPhone is great, but its battery sucks"

Voice quality – positive

Battery - negative

03 IATN Framework

IATN Framework

- **Challenge One**
  - **What is the key information in the sentence that we deserve?**

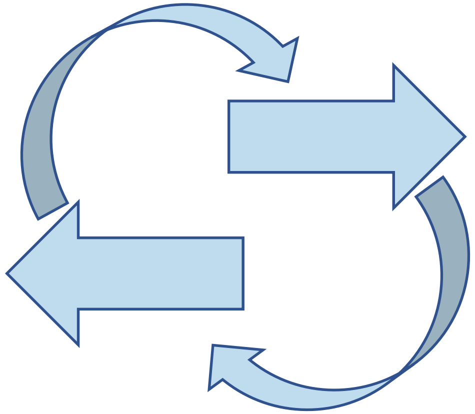
"Aspects"

- **Challenge Two**
  - **How do we make the most of this critical information?**

RAM
CPU
Price
…

The appearance of the PC looks good, but the battery life is too short.

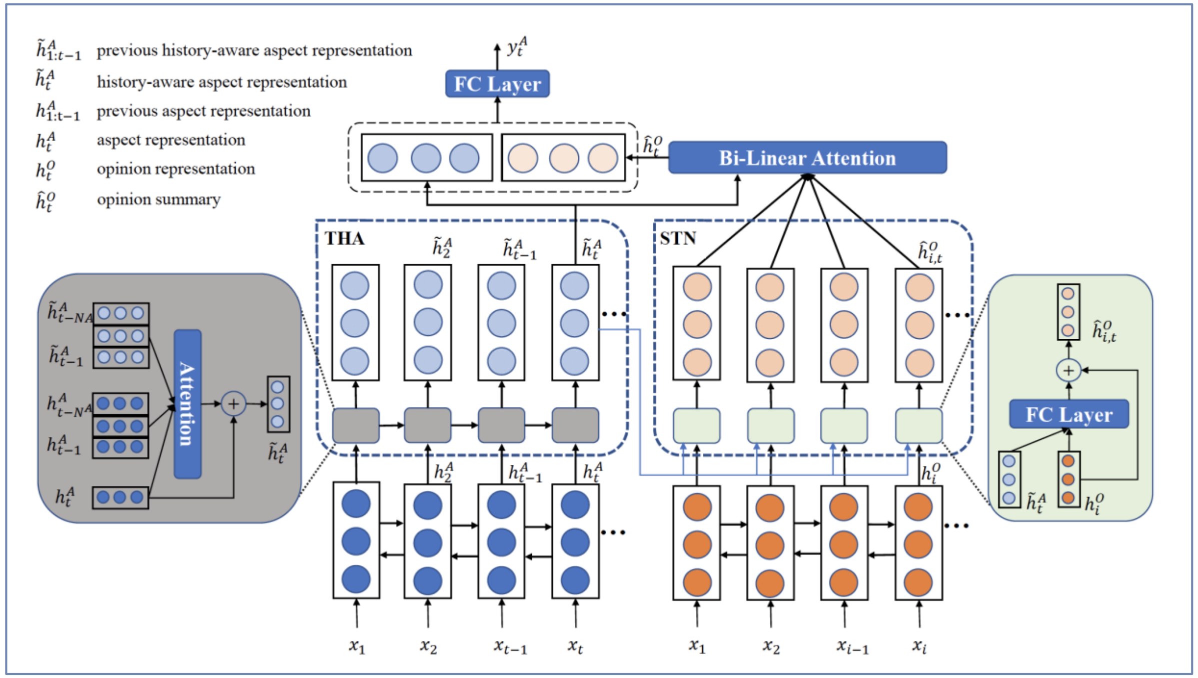
The appearance of this dress looks nice, and the fabric is not bad.



Interactive Network

IATN Framework

- **Aspect Extraction** [Li, Xin, et al 2018]
- **Method to extract aspect information?**
  - Truncated History Attention
    - Integrate the information of aspect detection history
  - Selective Transformation Network
    - Use opinion summary



Input sentences
1. *the device speaks about it self*
2. *Great survice !*
3. *Apple is unmatched in product quality, aesthetics, craftmanship, and custormer service*
4. *I am pleased with the fast log on, speedy WiFi connection and the long battery life*
5. *Also, I personally wasn't a fan of the portobello and asparagus mole*

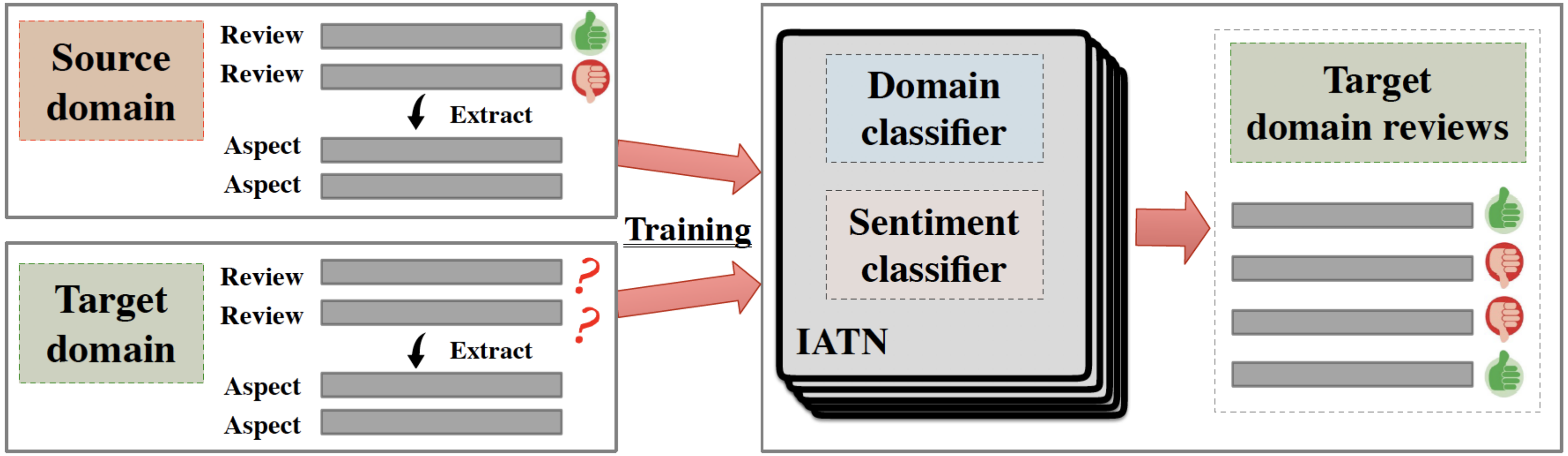
Output of OURS
NONE
survice
product quality, aesthetics, craftmanship, custormer service
log on, WiFi connection, battery life
portobello and asparagus mole

IATN Framework

- **Problem Definition**

Given two domains, D_s is the source domain and D_t is the target domain. Denote that each item (e.g., review) at both domains consists of n words marked as s= $\{w_s^1, w_s^2, \ldots, w_s^n\}$ and their aspect sequence contains m words marked as a= $\{w_a^1, w_a^2, \ldots, w_a^m\}$.

The goal is to train a robust model based on labeled data in source domain and adapt it to predict the unlabeled data in target domain.

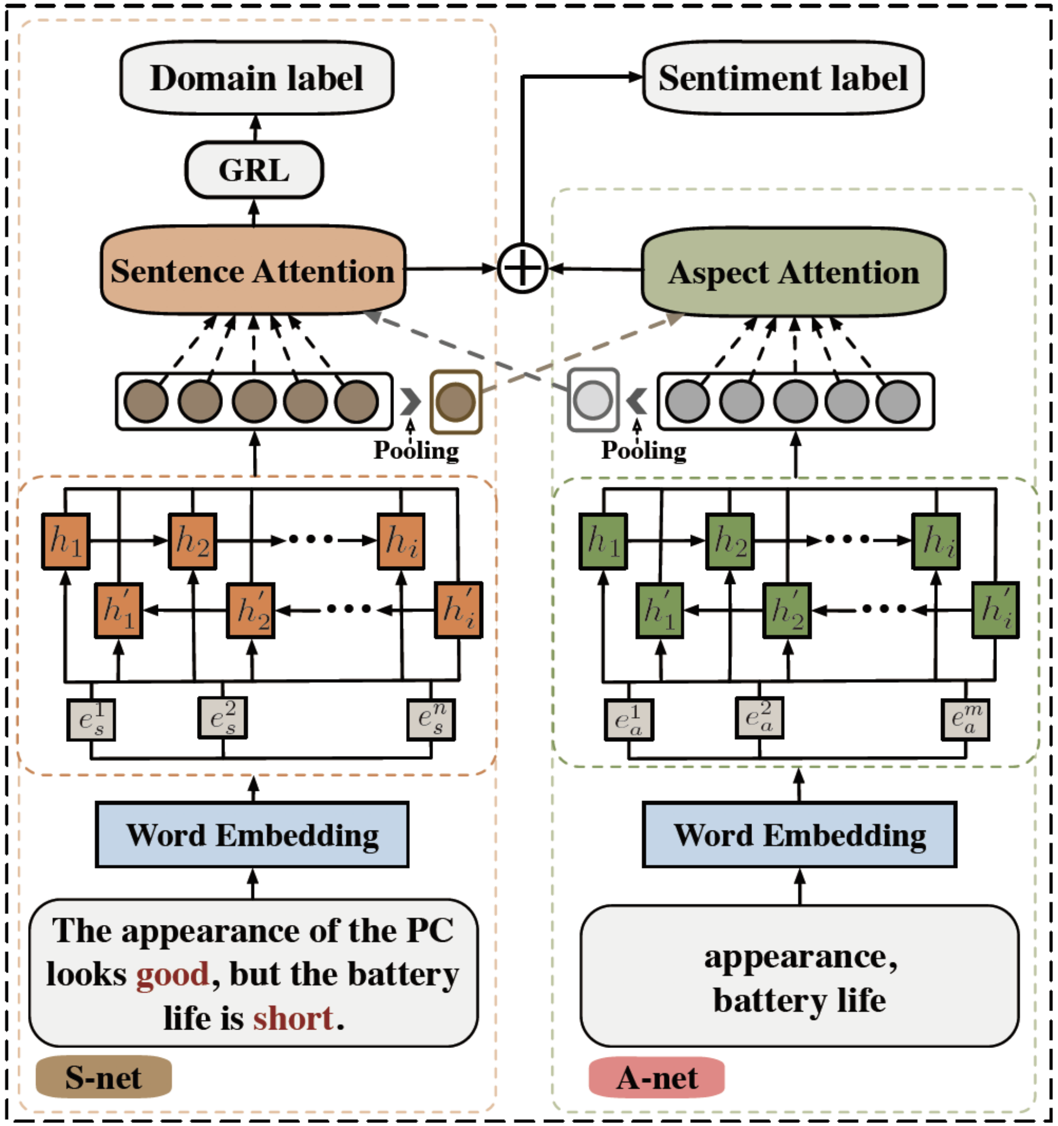


IATN Framework

- **Interactive Attention Transfer Network**
  - Construct a better sentiment transfer network
  - Make full use of the information of the aspect

- Consist of two similar parts
  - S-net
    - Learn shared information between domains
    - Pooling feature vectors to A-net
  - A-net
    - Modeling aspect information
    - Use aspect information as the auxiliary task of domain classification.

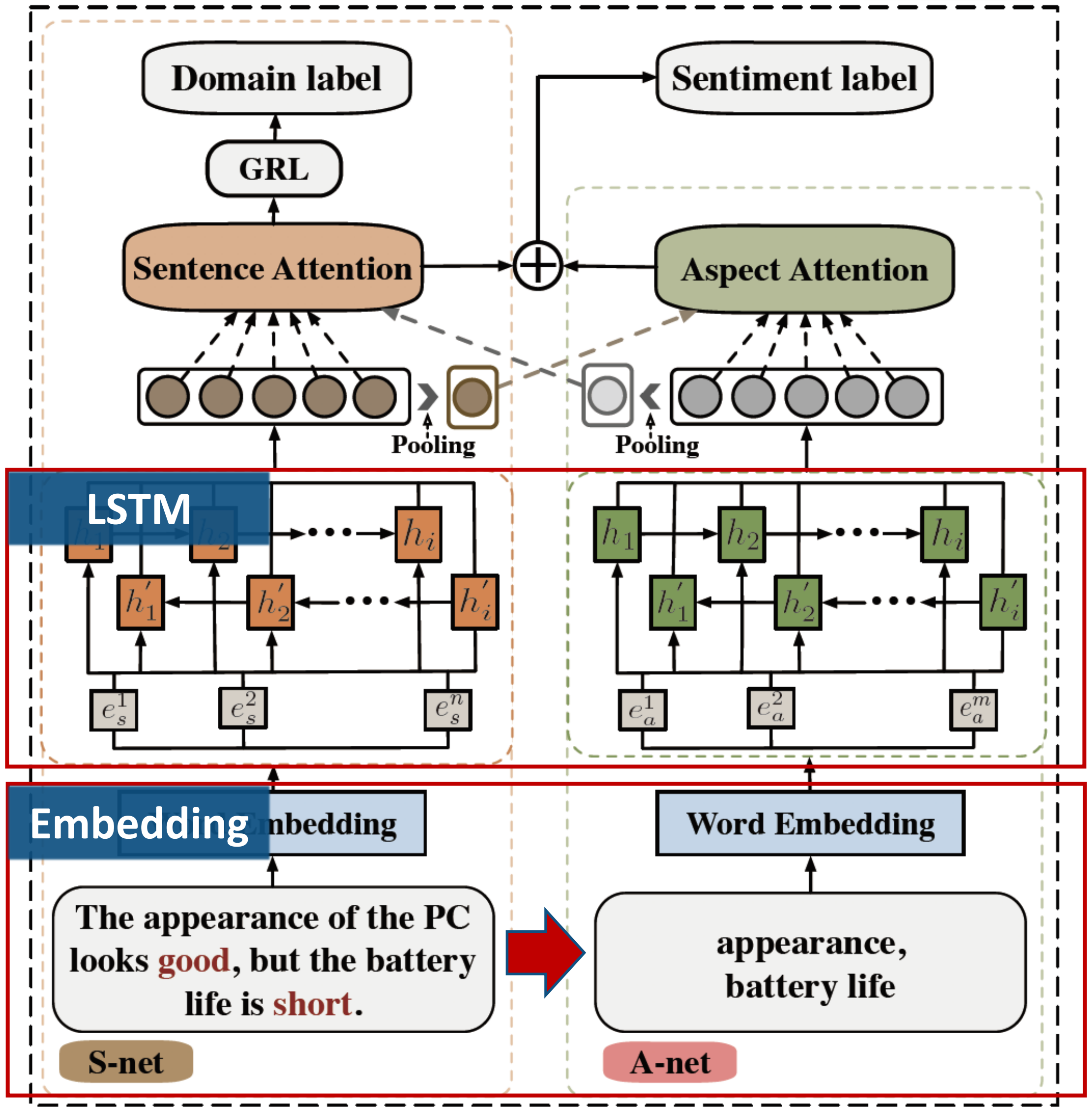
“IATN”



IATN Framework

- **IATN – Embedding Layer**
  - After aspect extraction, we map each word into a low-dimensional real-value vector.
  - We choose the pre-training method and take each word to a embedding vector.

- **IATN – LSTM Layer**
  - We adopt LSTM to learn hidden states.
  - Performs well in learning long-term dependencies
  - Solve gradient vanishing and expansion problems

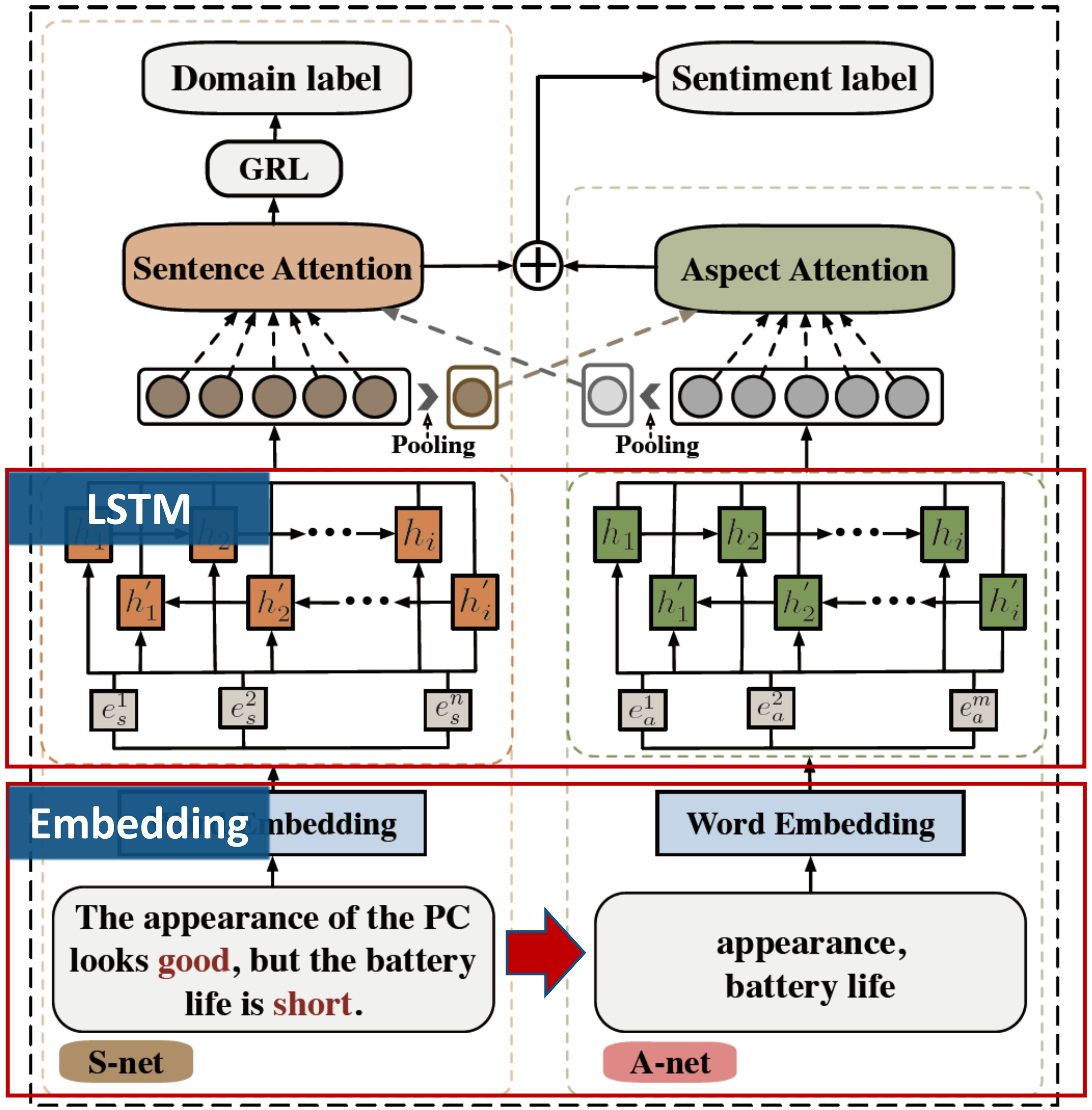


IATN Framework

- **IATN – Embedding Layer**
  - After aspect extraction, we map each word into a low-dimensional real-value vector.
  - We choose the pre-training method and take each word to a embedding vector.

- **IATN – LSTM Layer**

$$
\begin{aligned}
i_t &= \delta(W_{ei}e_s^t + W_{hi}h_{t-1} + \widehat{b}_i)\ , \\
f_t &= \delta(W_{ef}e_s^t + W_{hf}h_{t-1} + \widehat{b}_f)\ , \\
c_t &= f_t \cdot c_{t-1} + i_t \cdot \tau(W_{ec}e_s^t + W_{hc}h_{t-1} + \widehat{b}_c)\ , \\
o_t &= \delta(W_{eo}e_s^t + W_{ho}h_{t-1} + \widehat{b}_o)\ , \\
h_t &= o_t \cdot tanh(c_t)\ ,
\end{aligned}
$$



IATN Framework

- **IATN – Pooling Layer**

After getting the hidden state representations of sentences and aspects, we need to make them interactive.

$$h_s^p = \sum_{i=1}^{n}(h_s^i/\mathrm{n}) \qquad h_a^p = \sum_{i=1}^{m}(h_a^i/\mathrm{m})$$

- **IATN – Attention Layer**

$$\alpha_i = \frac{\exp\left(\gamma_s\left(h_s^i, h_a^p\right)\right)}{\sum_{j=1}^{n+1}\exp\left(\left(\gamma_s\left(h_s^i, h_a^p\right)\right)\right)} ,$$

From A-net

$$\gamma_s\left(h_s^i, h_a^p\right) = \tanh\left(\left[h_s^i, h_a^p\right]\cdot W_s + b_s\right) ,$$

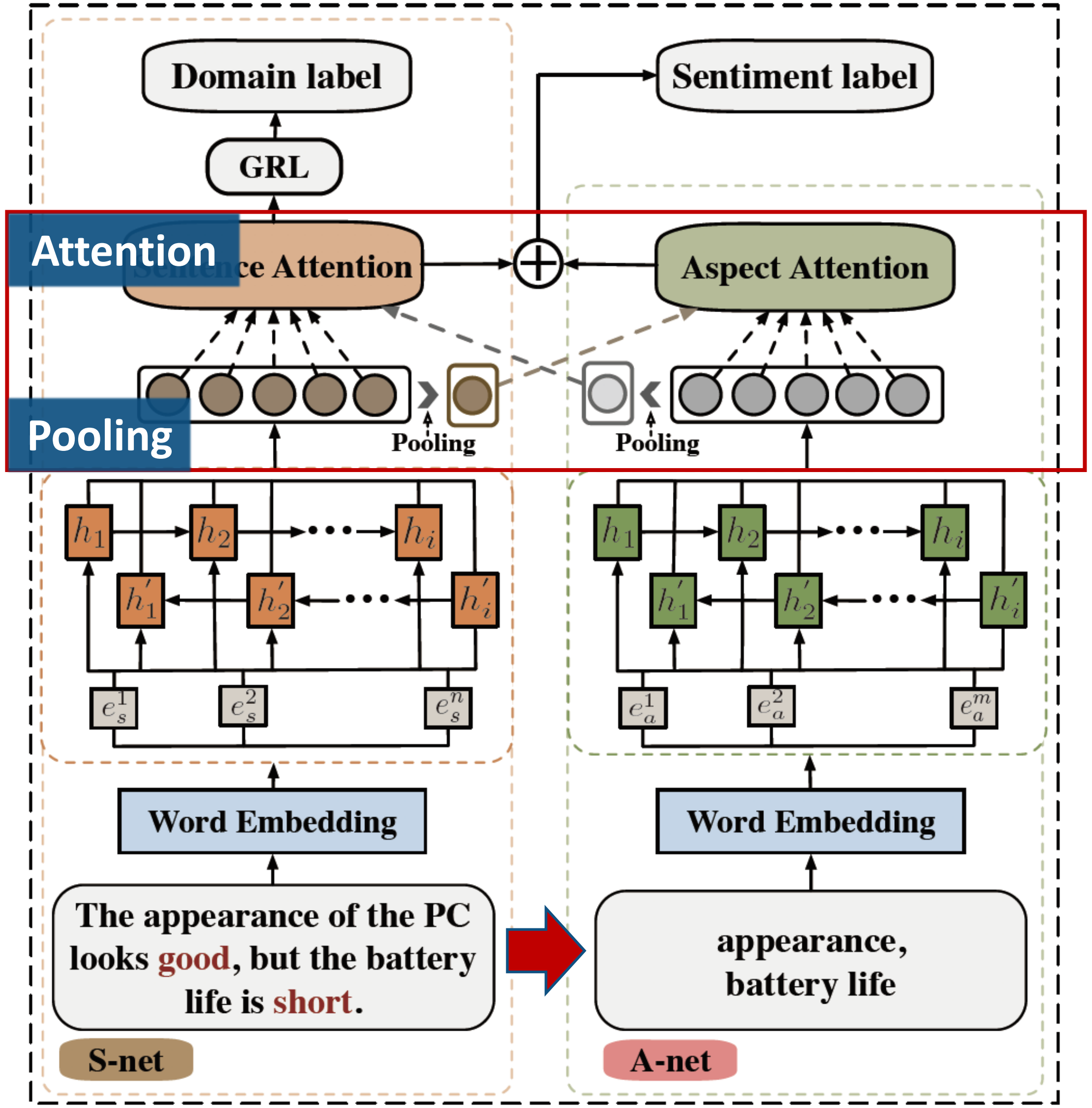
$$S_r = \sum_{i=1}^{n}\alpha_i\, h_s^i .$$

Sentence Representation

The same way

$$A_r = \sum_{i=1}^{n}\beta_i\, h_a^i .$$

Aspects Representation



IATN Framework

- **IATN – Two tasks**
  - **Domain Classification**

    GRL-Layer: $G_i = G(S_r)$.

    $y'_d = softmax(G_i)$

    $Loss_{dom} = -\frac{1}{N}\sum_{i=1}^{N} y'_d\, ln y_d + (1 - y'_d)\ln(1 - y_d)$.

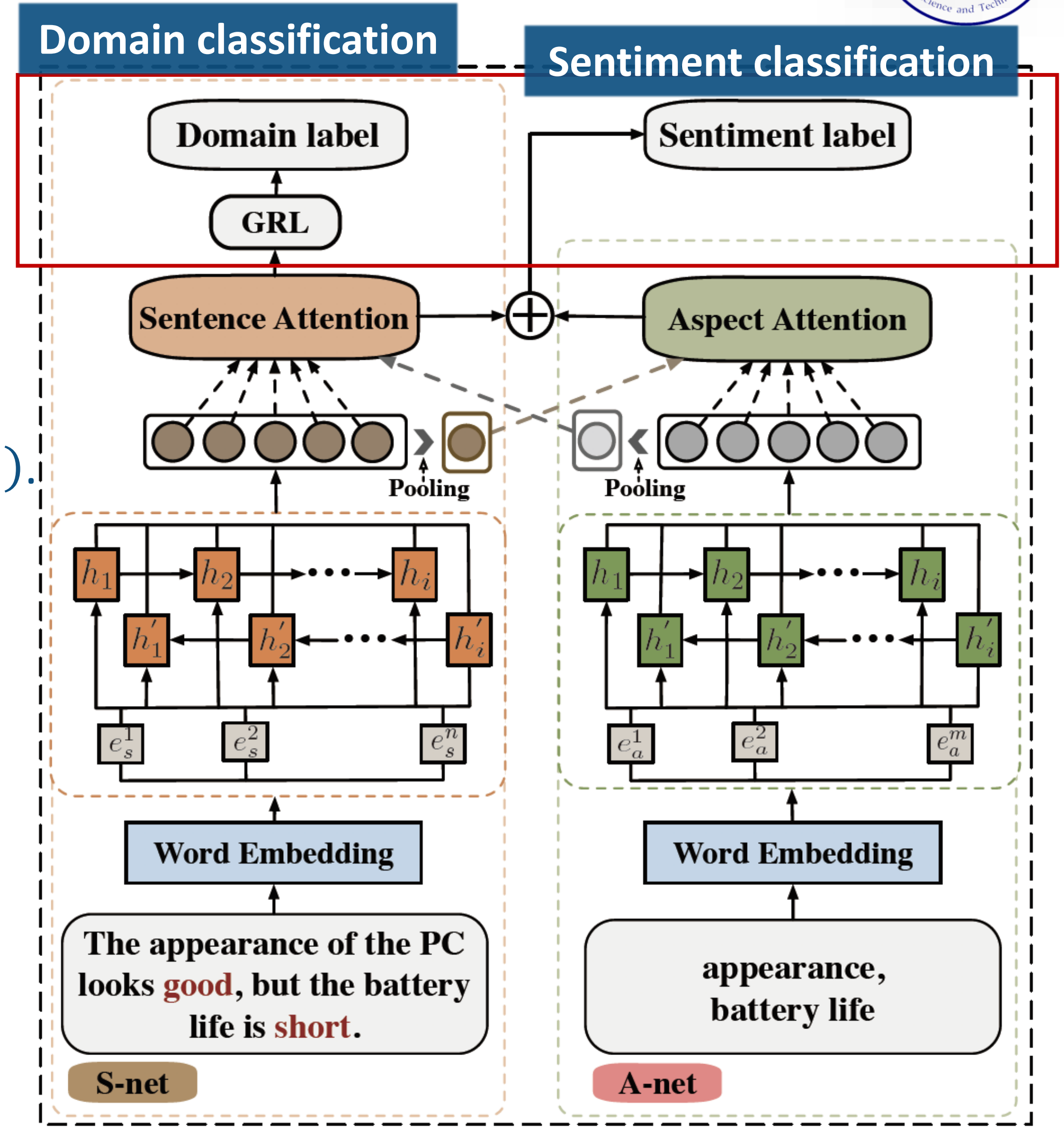
  - **Sentiment Classification**

    $V_i = S_r \bigoplus A_r$

    $y'_s = softmax(V_i)$

    $Loss_{sen} = -\frac{1}{N}\sum_{i=1}^{N} y'_s\, ln y_s + (1 - y'_s)\ln(1 - y_s)$.

Minimize: $L = L_{sen} + L_{dom} + \rho L_{reg}$



04 Experiment

Experiment

- **Data Description**
  - Amazon dataset (Main experiment)
  - Indiegogo dataset (Application verification)

（**Download**） http://www.indiegogo.com/download/

HAL 9000 Product Line from 2001: A Space Odyssey

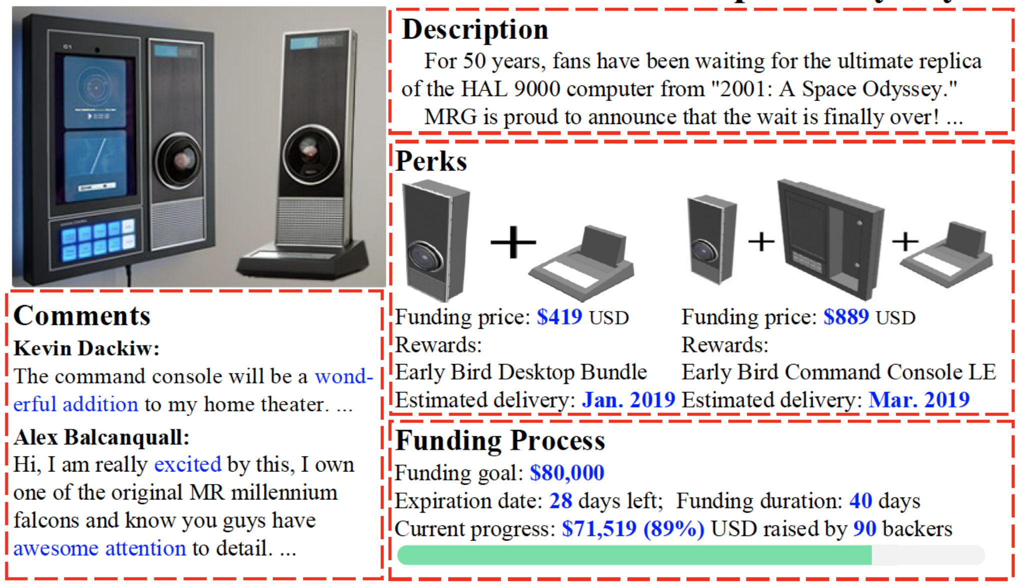


Table 1: Statistics of datasets after pre-processing.

Domains	Testing set percentage		
	# Train	# Test	# Unlabel
Books	5,000	1,000	8,000
DVD	5,000	1,000	8,000
Electronics	5,000	1,000	8,000
Kitchen	5,000	1,000	8,000

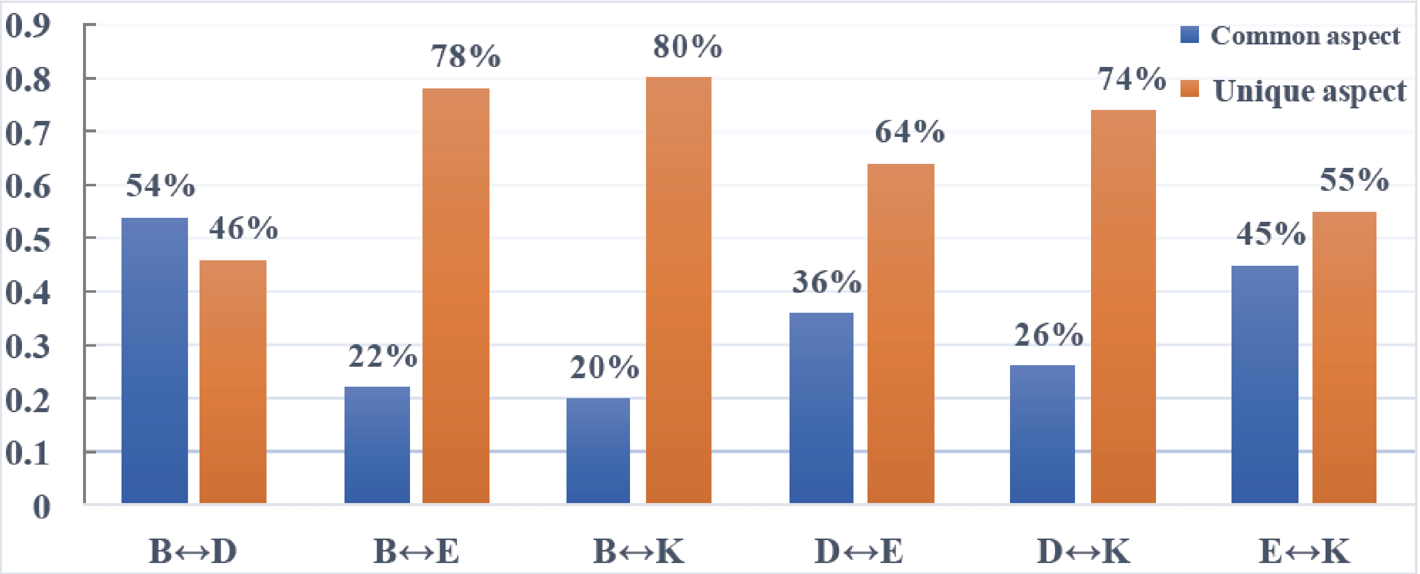
- **Amazon data preparation :**

  we choose the Amazon reviews data from four domains: Book, DVD, Electronics and Kitchen appliances. Each of the domains contains 6,000 labeled data.

Experiment

- **Data Analysis**
  - We observe that similar domains have more same aspects than different domains. (Which can prove something latter!)
  - B and D domain have shared most amount of aspects(plots, character, story, …)
  - B and K domain shared the least amount of aspects(those two domain are too far apart)

Think about those…



Top-100 aspects analysis between domains:

- We selected the top 100 Aspects with the highest frequency between every two domains.

Experiment

- **Benchmarks**
  - Naive is a non-domain-adaptive method which is trained in the source domain and predicts in target domain directly. It is designed based on LSTM.
  - SCL is a linear method, which aims to solve feature mismatch problem by aligning domain common and unique features.
  - SFA is a method which aims to build a bridge between the source and the target domains by aligning common and unique features.
  - mSDA is proposed to automatically learn a unified feature representation for sentences from a large amount of data in all the domains.
  - DANN is based on the adversarial training. DANN performs domain adaptation with the representation encoded in a 5000-dimension feature vector.
  - CNN-aux is based on Convolutional Neural Network and makes use of two auxiliary tasks to help inducing sentence embedding.
  - AMN is a method which learns domain-shared representations based on memory networks and adversarial training.
  - $HATN$ & $HATN^h$ are hierarchical attention networks to focus on both the word level and the sentence level sentiment. The former one does not contain the hierarchical positional encoding and the latter one does.

Experiment

- **Results**
  - As the table shows, IATN model has achieved the best performances on most tasks of this dataset.

	(a) **Book** →			(b) **DVD** →		
Benchmarks	Tasks			Tasks		
	B→D	B→E	B→K	D→B	D→E	D→K
Naive	0.786	0.752	0.737	0.756	0.734	0.767
SCL	0.807	0.763	0.771	0.782	0.754	0.779
SFA	0.813	0.776	0.785	0.788	0.758	0.786
mSDA	0.819	0.783	0.789	0.783	0.770	0.793
DANN	0.832	0.764	0.790	0.805	0.796	0.814
CNN-a	0.843	0.810	0.813	0.829	0.803	0.819
AMN	0.855	0.824	0.811	0.846	0.812	0.827
HATN	0.858	0.853	0.849	0.858	0.849	0.853
HATN^h	0.861	0.857	0.852	0.863	0.856	**0.862**
IATN^n	0.854	0.849	0.838	0.848	0.855	0.839
IATN	**0.868**	**0.865**	**0.859**	**0.870**	**0.869**	0.858

(c) **Electronics** →			(d) **Kitchen** →			
Tasks			Tasks			**Avg**
E→B	E→D	E→K	K→B	K→D	K→E	
0.696	0.722	0.787	0.686	0.723	0.807	**0.746**
0.716	0.745	0.817	0.713	0.752	0.818	**0.768**
0.724	0.754	0.825	0.724	0.758	0.825	**0.776**
0.738	0.761	0.837	0.730	0.755	0.831	**0.782**
0.735	0.786	0.841	0.752	0.776	0.843	**0.794**
0.749	0.793	0.843	0.779	0.803	0.855	**0.811**
0.766	0.827	0.857	0.805	0.812	0.867	**0.825**
0.808	0.838	0.868	0.824	0.841	0.868	**0.847**
0.810	0.840	0.879	0.833	**0.845**	0.870	**0.851**
0.768	0.825	0.859	0.828	0.835	0.864	**0.837**
0.818	**0.841**	**0.887**	**0.847**	0.844	**0.876**	**0.859**

From the experimental results, we can also observe that the classification accuracy between similar domains will be higher than different domain. For example, <u>"B&D" task is more accurate than "B&K" task because they have more similar aspects</u>.(As we mentioned before!)

Experiment

- **Case Study**
  - In order to validate that our model is able to identify the impact of aspect on sentiment representation, we visualize the aspect attention layer.

We can see that although "acting" and "pace" both appear with positive emotions in this review, but the more effective aspects "character" and "plot" have received negative emotions. Affected by these more powerful aspects, this review is finally labeled as negative.(Aspect influence different across domains!)

Book-pos Example:

This story captivated me right from the outset, as Vivian vague of memory . Her scenes are drawn in sensitively described and insightful detail, and she is a very realistically portrayed .

Book-neg Example:

I am not sure whatever possessed me to buy this book. Honestly , it was a complete waste of free time . To quote a friend , it was not the best use of my entertainment dollar . If you are a fan of pedestrian writing , lack-luster plots and hackneyed character development , this is your book .

Dvd-pos Example:

An amazing film ! When seeing this movie's poster , i was not too excited , but when watching it realized how awesome it really is . Not only it's story is well laid out , but the amount of special effects , great scences and good actors .

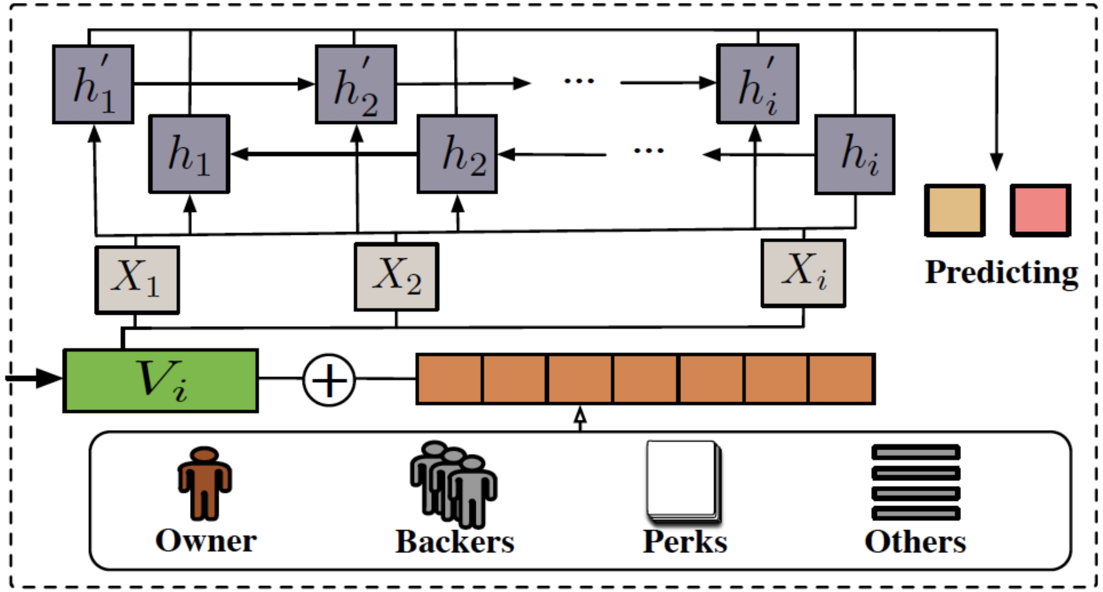
Dvd-neg Example:

The acting was good . The pace was adequate . However , the plot was predictable . The movie content just reeks of the intelligent thriller syndrome . Clive's character kept calling this the perfect robbery . I'm still don't understand what Jody Foster's character brings to the plot .

Aspect colored in deeper green means that it gains the heavier weight.

Experiment

- **Application Verification**



Benchmarks	Metrics			
	Accu.	Prec.	Rec.	F1-score.
SVM	0.7671	0.4567	0.6971	0.5536
LSTM^{one}	0.7862	0.4813	0.6732	0.5689
LSTM^{hatn}	0.7940	0.4843	0.6805	0.5674
LSTM^{iatn}	**0.8182**	**0.4977**	**0.6733**	**0.5743**

Prediction Model

- The further mission of cross-domain sentiment classification aims to solve the problem of unlabeled domain.
- Thus, to further verify the effectiveness of IATN, we design an application on an unlabeled dataset. (Crowdfunding review dataset!)

Prediction Result

- As Table shows, the SVM method without review information performs the worst and the accuracy result is 5.11% lower than $LSTM^{iatn}$.
- $LSTM^{iatn}$ improves the accuracy by 3.20% and 2.42% than $LSTM^{one}$ and $LSTM^{hatn}$ respectively.

05 Conclusion

Conclusion

- What we do ?
  - We propose to make cross-domain sentiment classification by integrating the contextual representation and aspect representation of sentences.
  - We propose a novel IATN method which utilizes the interactive attention mechanism to get important information from both the sentence and aspect.
  - We conduct extensive experiments on two real-world datasets. The experimental results validate that our method outperforms other state-of-the-art methods.
- Future
  - Add aspect information to the hierarchical attention network.
  - Do some sentiment transfer learning at the aspect-level.

Thank you!   Q&A

Kai Zhang

2018/1/28