Learning Domain-Sensitive and Sentiment-Aware Word Embeddings

Bei Shi[1], Zihao Fu[1], Lidong Bing[2] and Wai Lam[1]

[1]Department of SEEM, The Chinese University of Hong Kong
[2]Tencent AI Lab

Motivation

Recent research works look into the problem of learning task-specific word embeddings for sentiment classification.

- **Sentiment-Aware**. Some words, especially sentiment words, have similar syntactic context but opposite sentiment polarity, such as the words "good" and "bad".
- **Domain-Sensitive**. The polarity of some sentiment words varies according to their domain.
  - "lightweight(electronics)": A lightweight device is easier to carry.
  - "lightweight(movie)": The movie do not invoke deep thoughts among the audience.
- We aim at learning word embeddings that are both domain-sensitive and sentiment-aware.

Our Contribution

- Jointly model the sentiment semantics and domain specificity of words, expecting the learned embeddings to achieve superior performance for the task of sentiment classification.
- Our model exploits the information of sentiment labels and context words to distinguish domain-common and domain-specific words.
  - **Domain-common word embeddings**. The words "good" and "interesting" convey consistent semantic meanings and positive sentiments in different domains, which should have similar embeddings across domains.
  - **Domain-specific word embeddings**. The sentiments or meanings of word embeddings across domains are different.
- The learning of domain-common embeddings can allow the advantage of data augmentation of common semantics of multiple domains, and meanwhile, domain-specific embeddings allow us to capture the varied semantics of specific words in different domains.

The Objective Function

$$\mathcal{L} = \mathcal{L}^p + \mathcal{L}^q \tag{1}$$

$$\mathcal{L}^p = \sum_{r \in \mathcal{D}^p} \sum_{w \in r} \sum_{w_t \in c_w} \log p(w_t|w) + \sum_{r \in \mathcal{D}^p} \sum_{w \in r} \log p(y_w|w) \tag{2}$$

Our DSE model

- Each word w is associated with a domain-common vector U_w^c and two domain-specific vectors, namely U_w^p specific to the domain p and U_w^q specific to the domain q.
- For each word w, we use a latent variable z_w to depict its domain commonality. When $z_w = 1$, it means that w is common in both domains. Otherwise, w is specific to the domain p or the domain q.
- The probability of predicting the context words is affected by not only the relatedness with the target words but also the domain-commonality of the target word.

$$p(w_t|w, z_w = 1) = \frac{\exp(U_w^c \cdot V_{w_t})}{\sum_{w' \in \Lambda} \exp(U_w^c \cdot V_{w'})} \tag{3}$$

$$p(w_t|w, z_w = 0) = \begin{cases} \frac{\exp(U_w^p \cdot V_{w_t})}{\sum_{w' \in \Lambda} \exp(U_w^p \cdot V_{w'})}, & \text{if } w \in \mathcal{D}^p \\ \frac{\exp(U_w^q \cdot V_{w_t})}{\sum_{w' \in \Lambda} \exp(U_w^q \cdot V_{w'})}, & \text{if } w \in \mathcal{D}^q \end{cases} \tag{4}$$

- Similarly, the prediction of review sentiment depends on not only the text information but also the domain-commonality.

$$p(y_w = 1|w, z_w = 1) = \sigma(U_w^c \cdot \mathbf{s}) \tag{5}$$

$$p(y_w = 1|w, z_w = 0) = \begin{cases} \sigma(U_w^p \cdot \mathbf{s}) & \text{if } w \in \mathcal{D}^p \\ \sigma(U_w^q \cdot \mathbf{s}) & \text{if } w \in \mathcal{D}^q \end{cases} \tag{6}$$

Experiment

	B & D		B & E		B & K		D & E		D & K		E & K	
	Acc.	F1	*Acc.*	F1	*Acc.*	F1	*Acc.*	F1	*Acc.*	F1	*Acc.*	F1
BOW	0.680	0.653	0.738	0.720	0.734	0.725	0.705	0.685	0.706	0.689	0.739	0.715
EmbeddingP	0.753	0.740	0.752	0.745	0.742	0.741	0.740	0.746	0.707	0.702	0.761	0.760
EmbeddingQ	0.736	0.732	0.697	0.697	0.706	0.701	0.762	0.759	0.758	0.759	0.783	0.780
EmbeddingCat	0.769	0.731	0.768	0.763	0.763	0.763	0.787	0.773	0.770	0.770	0.807	0.803
EmbeddingAll	0.769	0.759	0.765	0.740	0.775	0.767	0.783	0.779	0.779	0.776	0.819	0.815
Yang	0.767	0.752	0.775	0.766	0.760	0.755	0.791	0.785	0.762	0.760	0.805	0.804
SSWE	0.783	0.772	0.791	0.780	**0.801**	0.792	0.825	0.815	0.795	0.790	0.835	0.824
DSE_c	0.773	0.750	0.783	0.781	0.775	0.773	0.797	0.792	0.784	0.776	0.806	0.800
DSE_w	**0.794**†‡	**0.793**†‡	**0.806**†‡	**0.802**†‡	0.797†	**0.793**†	**0.843**†‡	**0.832**†‡	**0.829**†‡	**0.827**†‡	**0.856**†‡	**0.853**†‡

Table: Results of review sentiment classification.

	B & D		B & E		B & K		D & E		D & K		E & K	
	HL	MPQA	HL	MPQA	HL	MPQA	HL	MPQA	HL	MPQA	HL	MPQA
EmbeddingP	0.740	0.733	0.742	0.734	0.747	0.735	0.744	0.701	0.745	0.709	0.628	0.574
EmbeddingQ	0.743	0.701	0.627	0.573	0.464	0.453	0.621	0.577	0.462	0.450	0.465	0.453
EmbeddingCat	0.780	0.772	0.773	0.756	0.772	0.751	0.744	0.728	0.755	0.702	0.683	0.639
EmbeddingAll	0.777	0.769	0.773	0.730	0.762	0.760	0.712	0.707	0.749	0.724	0.670	0.658
Yang	0.780	0.775	0.789	0.762	0.781	0.770	0.762	0.736	0.756	0.713	0.634	0.614
SSWE	**0.816**	**0.801**	0.831	0.817	0.822	**0.808**	**0.826**	0.785	0.784	0.772	0.707	0.659
DSE	0.802	0.788	**0.833**	**0.828**	**0.832**	0.799	0.804	**0.797**	**0.796**	**0.786**	**0.725**	**0.683**

Table: Results of lexicon term sentiment classification.

Casy Study

Term	Domain	$p(z=1)$	Sample Reviews
"lightweight"	B & D	0.999	- I find Seth Godin's books incredibly **lightweight**. There is really nothing of any substance here.(B) - I love the fact that it's small and **lightweight** and fits into a tiny pocket on my camera case so I never lose track of it.(E) - These are not "**lightweight**" actors. (D) - This vacuum does a pretty good job. It is **lightweight** and easy to use.(K)
	B & E	0.404	
	B & K	0.241	
	D & E	0.380	
	D & K	0.013	
	E & K	0.696	
"die"	B & E	0.435	- I'm glad Brando lived long enough to get old and fat, and that he didn't **die** tragically young.(B) - Like many others here, my CD-changer **died** after a couple of weeks and it wouldn't read any CD.(E) - I had this toaster for under 3 years when I came home one day and it smoked and **died**. (K)
	B & K	0.492	
	E & K	0.712	
"mysterious"	B & E	0.297	- This novel really does cover the gamut: stunning twists, genuine love, beautiful settings, desire for riches, **mysterious** murders, detective investigations, false accusations, and self vindication.(B) - Caller ID functionality for Vonage **mysteriously** stopped working even though this phone's REN is rated at 0.1b. (E)
"great"	B & D	0.760	- This is a **great** book for anyone learning how to handle dogs.(B) - This is a **great** product, and you can get it, along with any other products on Amazon up to $500 Free!(E) - I grew up with drag racing in the 50s, 60s & 70s and this film gives a **great** view of what it was like.(D) - This is a **great** mixer its a little loud but worth it for the power you get.(K)
	B & E	0.603	
	B & K	0.628	
	D & E	0.804	
	D & K	0.582	
	E & K	0.805	

Table: Learned domain-commonality for some words. $p(z=1)$ denotes the probability that the word is domain-common.