Xi-bin JIA, Ya JIN, Ning LI, Xing SU, Barry CARDIFF, Bir BHANU, 2018. Words alignment based on association rules for cross-domain sentiment classification. *Frontiers of Information Technology & Electronic Engineering*, 19(2):260-272. https://doi.org/10.1631/FITEE.1601679

Words alignment based on association rules for cross-domain sentiment classification

Front Inform Technol Electron Eng

Key words: Sentiment classification; Cross-domain; Association rules

Corresponding author: Xing SU
E-mail: xingsu@bjut.edu.cn
ORCID: http://orcid.org/0000-0002-4214-3363

Motivation

1. The field of product reviews and comments in the Internet involves a wide range of domains. It is time- and labor-consuming to obtain enough labeled data for every domain's machine learning tasks.

2. Performance of the sentiment classifier will decline sharply when different distributions exist between training data and testing data, even if there are abundant labeled data to train a sentiment classifier.

Front Inform Technol Electron Eng

Main idea

1. Inspired by the observations that even if there may be some different sentiment lexicons in different domains, they can still provide knowledge of general sentiment words or domain-shared words.

2. If the sentiment classifier can learn the relationships between domain-specific words and domain-shared words, it will be useful in cross-domain sentiment classifier training.

Front Inform Technol Electron Eng

Method

1. Establish a direct mapping relationship:

(1) Extract the domain-specific words and domain-shared words;

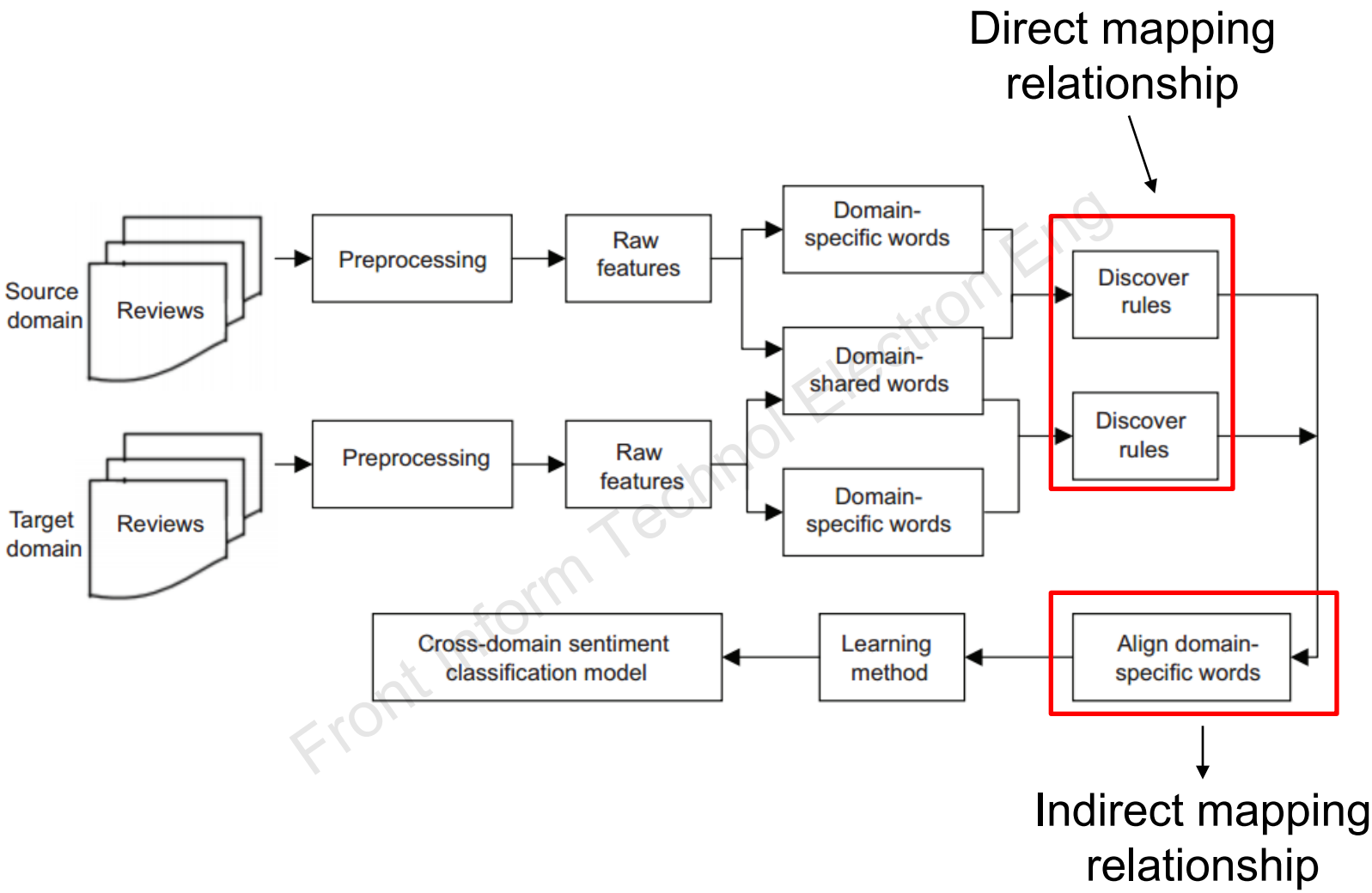
(2) Use the association rule learning method to learn the direct mapping relationship between domain-specific words and domain-shared words in the same domain.

2. Establish an indirect mapping relationship:

(1) Look for the domain-share words between source and target domains;

(2) Establish a weighted graph-based evaluation model for correlation to establish an indirect mapping relationship between domain-specific words in different domains via domain-shared words.

Front Inform Technol Electron Eng



Front Inform Technol Electron Eng

Major results

Table 3 shows a comparison of our approach with four reference algorithms on the accuracy of cross-domain sentiment classification.

Table 3 Accuracy of algorithms on 12 cross-domain sentiment classification tasks using the the Amazon® datasets

Task	Accuracy (%)			
	NoTransf	SCL	SFA	WAAR
B→D	76.80	78.50	80.54	**81.25**
E→D	71.25	75.25	75.50	**75.80**
K→D	73.05	76.65	76.70	**76.90**
D→B	73.35	78.26	77.54	**79.60**
E→B	72.42	75.02	**75.40**	73.50
K→B	71.80	72.78	74.20	**74.30**
B→E	71.28	**75.22**	72.10	73.50
D→E	72.69	74.20	76.04	**77.60**
K→E	83.02	85.04	85.02	**85.05**
B→K	74.45	77.08	**78.02**	76.90
D→K	75.02	78.94	**79.50**	77.85
E→K	85.02	85.06	**85.95**	85.03

Best results are shown in bold

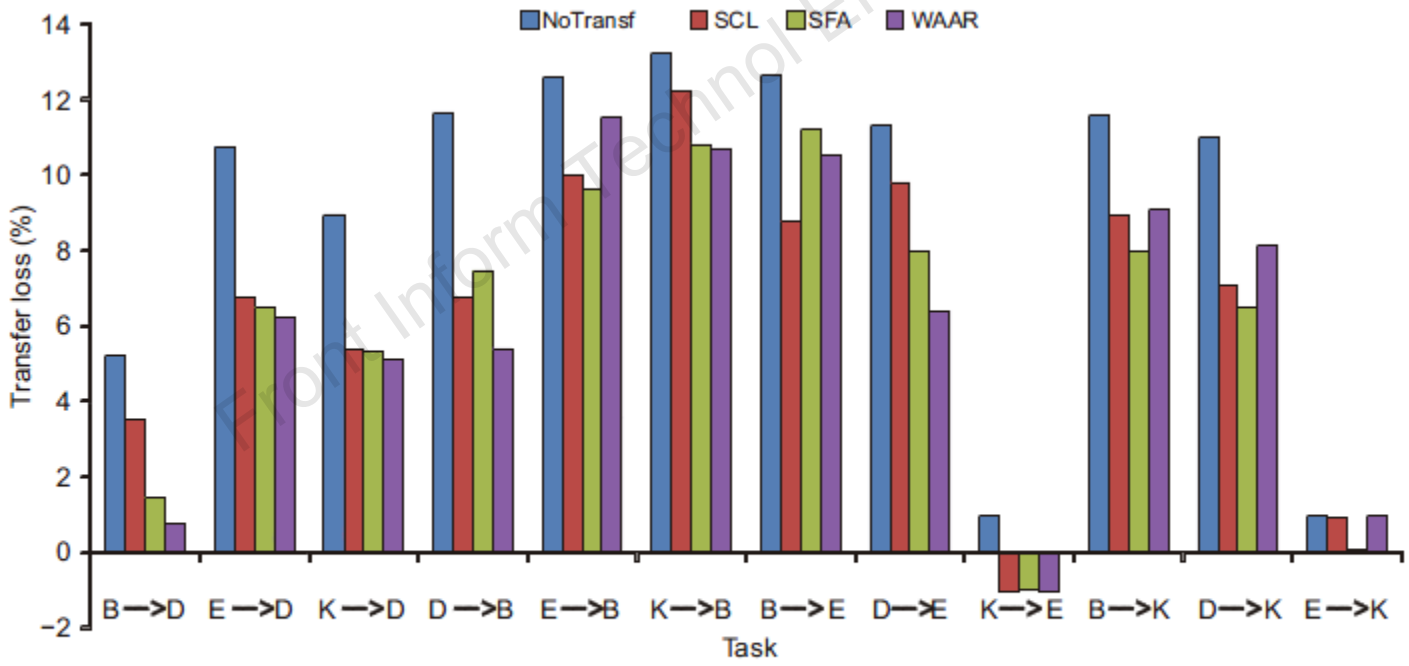
Front Inform Technol Electron Eng

The novel approach (WAAR) approach is more accurate than the other algorithms in most transfer learning tasks. This implies that knowledge can be easily transferred between domains to improve the accuracy of cross-domain sentiment classification by the WAAR approach.

Major results

Fig.3 shows the transfer loss of all cross-domain sentiment classification tasks in four methods.

The best transfer is achieved by the WAAR approach in 8 of 12 transfer learning tasks.



Front Inform Technol Electron Eng

Fig. 3 Transfer loss of algorithms on Amazon® datasets for 12 cross-domain sentiment classification tasks

References to color refer to the online version of this figure. B: Books; D: DvDs; E: Electronics; K: Kitchen

Conclusions

1. A novel approach (WAAR) was proposed to align domain-specific words in different domains for cross-domain sentiment classification.

2. Based on the association rule algorithm such as the Apriori algorithm and domain-shared words, an indirect mapping relationship between domain-specific words in different domains can be established to eliminate the divergence between different domains.

3. Experimental results show the effectiveness of the WAAR approach in improving the performance of cross-domain sentiment classification.

Front Inform Technol Electron Eng