Transfer learning on sentiment analysis

Nasrin Baratalipour

Sentiment Classification

Sentiment Analysis is denoted as the automatic detection and classification of opinions expressed in the text spans.

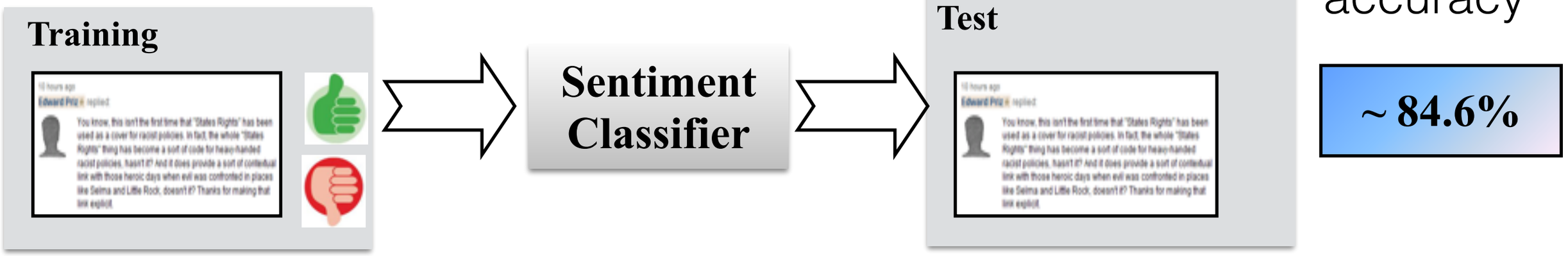
Sentiment Polarities:

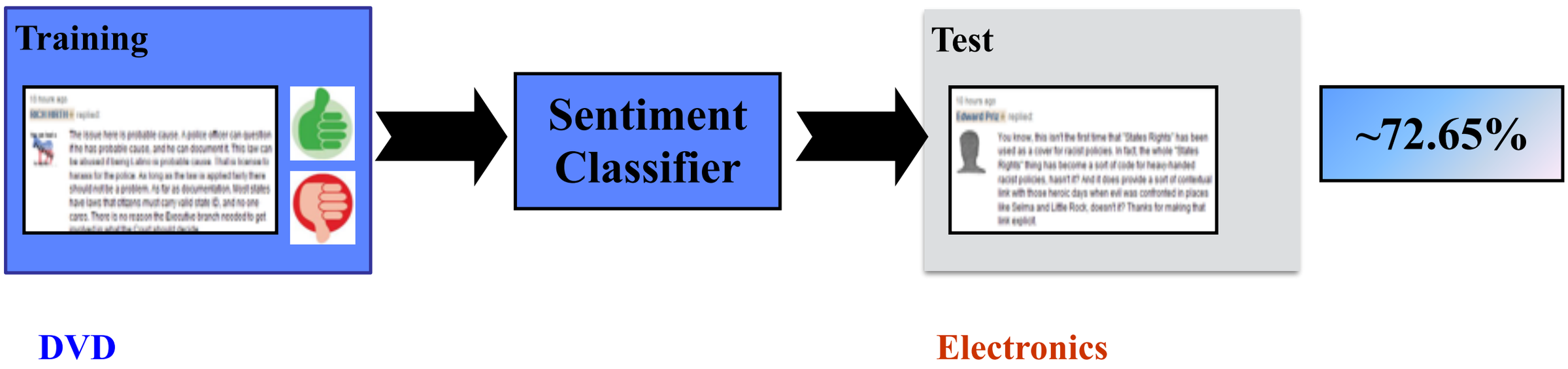
- I like the new Nokia model.
- I find the functionality of the new phone *less* practical.
- I don’t have any opinion about the new Nokia model.

Motivation

- Problem:
  annotating corpora for every possible domain of interest is impractical
- Why?
  - In some domain, labeled data are in short supply
  - Annotating new data is expensive in the terms of its cost and time
- Solution:
  - Reusing existing labeled sentiment dataset in other domain.

Traditional learning algorithms do not work!





Differences in domains

- Sentiment is expressed differently in different domain
  - Same phrase can indicate different sentiment in different domains
    - “go read the book”
- Difference in vocabularies across different domains
  - “must-read”
  - “excellent-product”

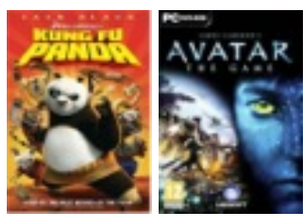
Transfer Learning

Target Domain

Training

A few labeled training data

DVD



Classification

Target Domain

Test

DVD

Source Domains

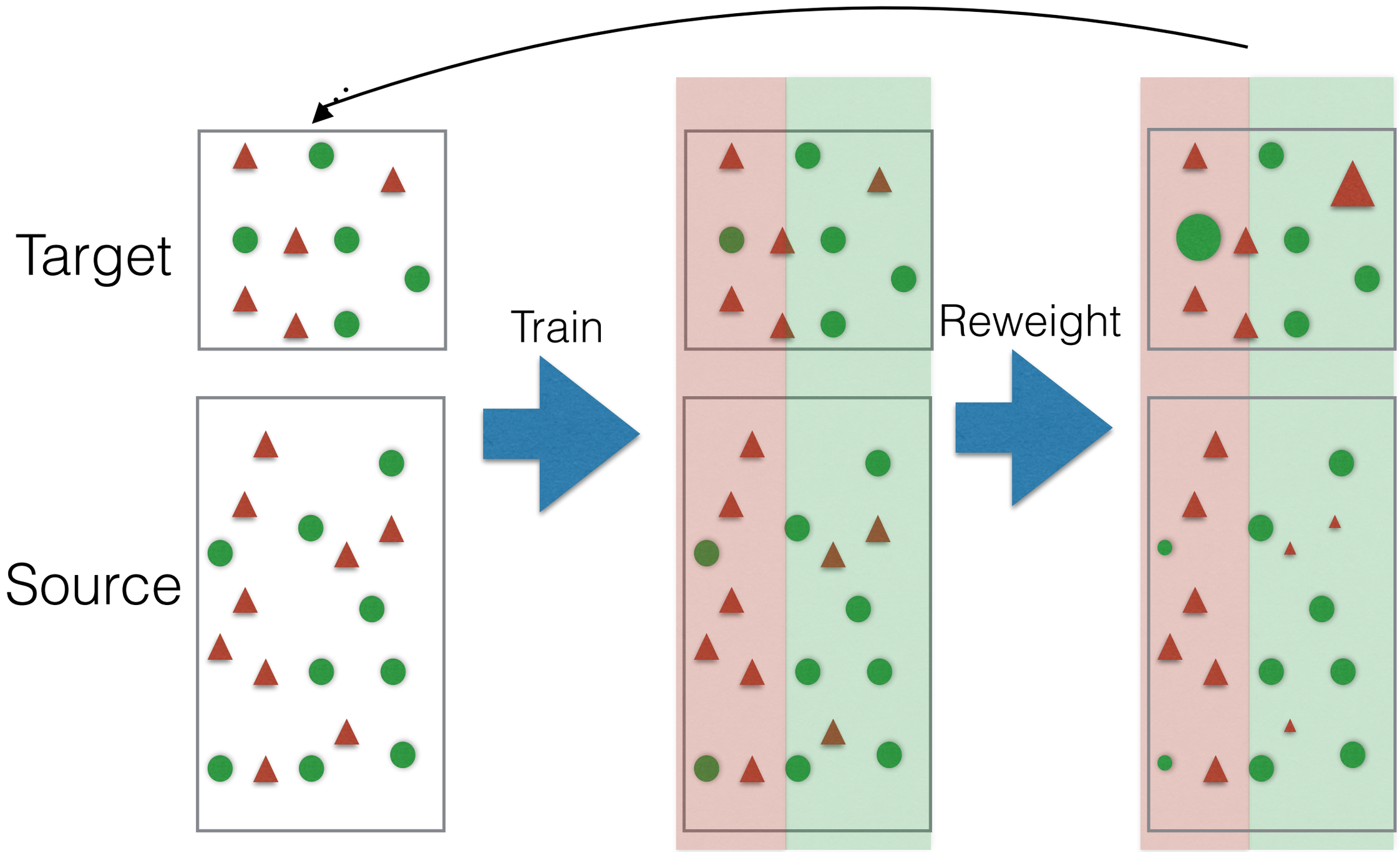
Book

Algorithm: TrAdaBoost

TrAdaBoost
(Dai et al, 2007, Boosting for transfer learning)

- A Transfer learning algorithm framework
- Extend *Adaboost*
- Iteratively reweigh instances

TrAdaBoost



Results

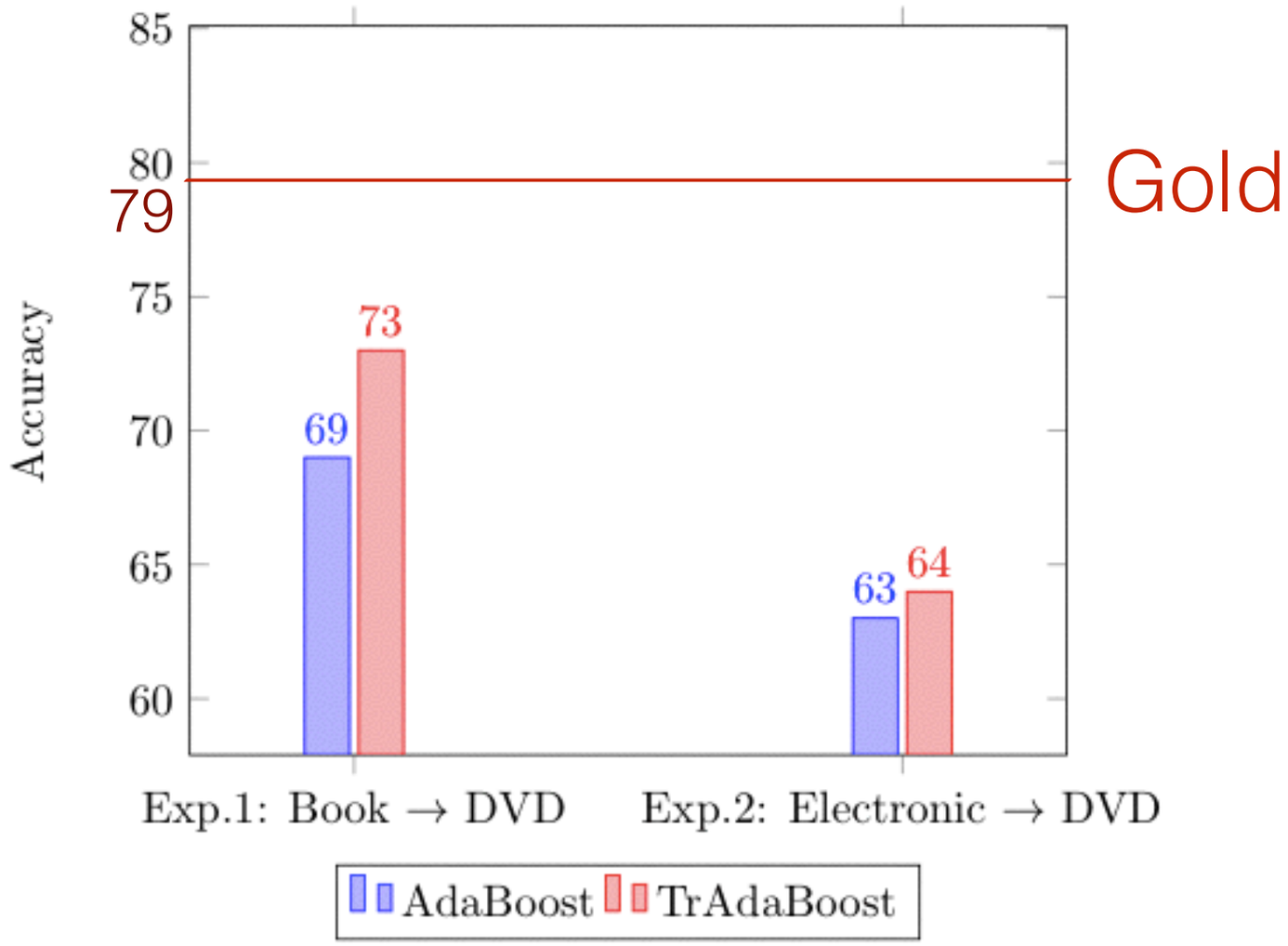


Figure 1: Accuracy results for domain adaptation between Book-DVD and Electronic-DVD pairs

Dataset

(Blitzer et al. 2007)

	DVD Review (Target)	Book Review (Source)	Elec. Review (Source)
Total	2000	2000	2,000
Exp 1.	300	1,600	0
Exp 2.	300	0	1,600
Gold	1,600	0	0
Test	100	0	0

Conclusion

- Apply transfer learning to Sentiment Analysis Task
- Implement TrAdaBoost
- The source domain affect results