TensorFlow Study

Domain Adaptation & Deep Learning Approach

Woojin Lee

This document is confidential and is intended solely for the use

Domain Adaptation

- ## Domain Adaptation

amazon.com

Customer Reviews

789 of 845 people found the following review helpful:

Gripping but amazingly incomplete, October 27, 2011

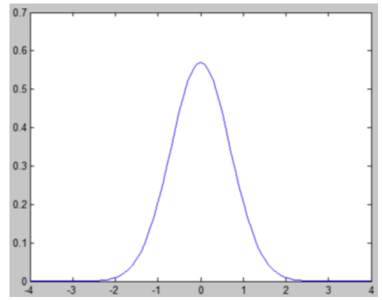
By **David Dennis** (West Palm Beach, FL United States) - See all my reviews

REAL NAME™

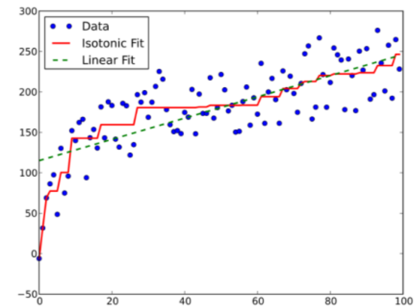
This review is from: **Steve Jobs (Hardcover)**

This is a gripping journey into the life of an amazing individual. Despite its girth of nearly 600 pages, the book zips along at a torrid pace.

data



model



prediction

amazon.com

Customer Reviews

789 of 845 people found the following review helpful:

Gripping but amazingly incomplete, October 27, 2011

By **David Dennis** (West Palm Beach, FL United States) - See all my reviews

REAL NAME™

This review is from: **Steve Jobs (Hardcover)**

This is a gripping journey into the life of an amazing individual. Despite its girth of nearly 600 pages, the book zips along at a torrid pace.

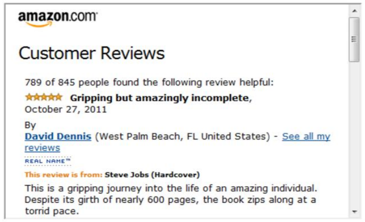
Book, Kitchen, DVD, Electronics

모든 Domain에 대하여 모델을 만드는 것은 비 효율적일 뿐만 아니라 data가 적거나 labeled data가 적은 domain에서는 적용이 힘들다.

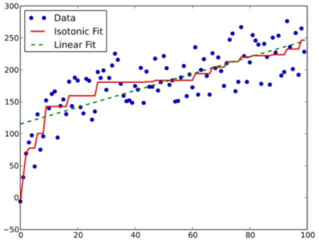
Domain Adaptation

Domain Adaptation

Given a source domain D_s *and a corresponding leaning task* T_s, *a target domain* D_T *and a corresponding learning task* T_T, *transductive transfer learning aims to improve the learning of the target predictive function* f_T *in* D_T *using the knowledge in* D_s *and* T_s, *where* $D_s \neq D_T$ *and* $T_s = T_T$.



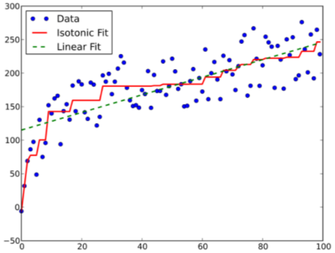
Source Domain (BOOK)



Source task (sentimental)

Knowledge Transfer

Target Domain (KITCHEN)



Target task (sentimental)

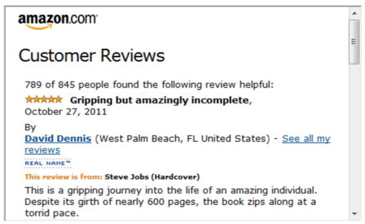
Structural Correspondence Learning

- **Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification (2007)**
  - Source Domain and Target Domain is drawn from different distribution.
  - Structural Correspondence Learning (SCL) domain adaptation algorithm
  - Amazon Review Data set
    - Books, DVD, Electronics, kitchen
  - Books – Read Half, headache
  - Kitchen – Defective, returned
  - Labeled data for source domain will help to build a good classifier for target domain

Structural Correspondence Learning

■ Step Learning Process

- **Step 1**



Source Domain (BOOK)

Target Domain (KITCHEN)

→ **Correspondence Mapping**

- **Step 2**
  - **Learning with labeled Source Domain**

SLCF
What You Imagine is What You Get

Structural Correspondence Learning

- BOOK
  - The book is so repetitive that I found my self yelling.... I will definitely not buy another
  - A disappointment ... Ender was talked about for 50 pages altogether
  - It's unclear ... It's repeated and boring

- Kitchen
  - Do not buy the steamer .... It is defective
  - The very nice lady assured me that I must have a defective set .... What a disappointment
  - Maybe mine was defective , the directions were unclear

SCL, not SCL-MI	SCL-MI, not SCL
book one <num> so all very about they like good when	*a_must a_wonderful loved_it weak don't_waste awful highly_recommended and_easy*

Structural Correspondence Learning

Dataset and Baseline

- Amazon product reviews for four different product types: books, DVDs, electronics and kitchen appliances. Each review consists of a rating (0-5 stars)
- Reviews with rating > 3 were labeled positive, those with rating < 3 were labeled negative
- We had 1000 positive and 1000 negative examples for each domain
- Each labeled dataset was split into a training set of 1600 instances and a test set of 400 instances.
- Linear Classifier

Structural Correspondence Learning

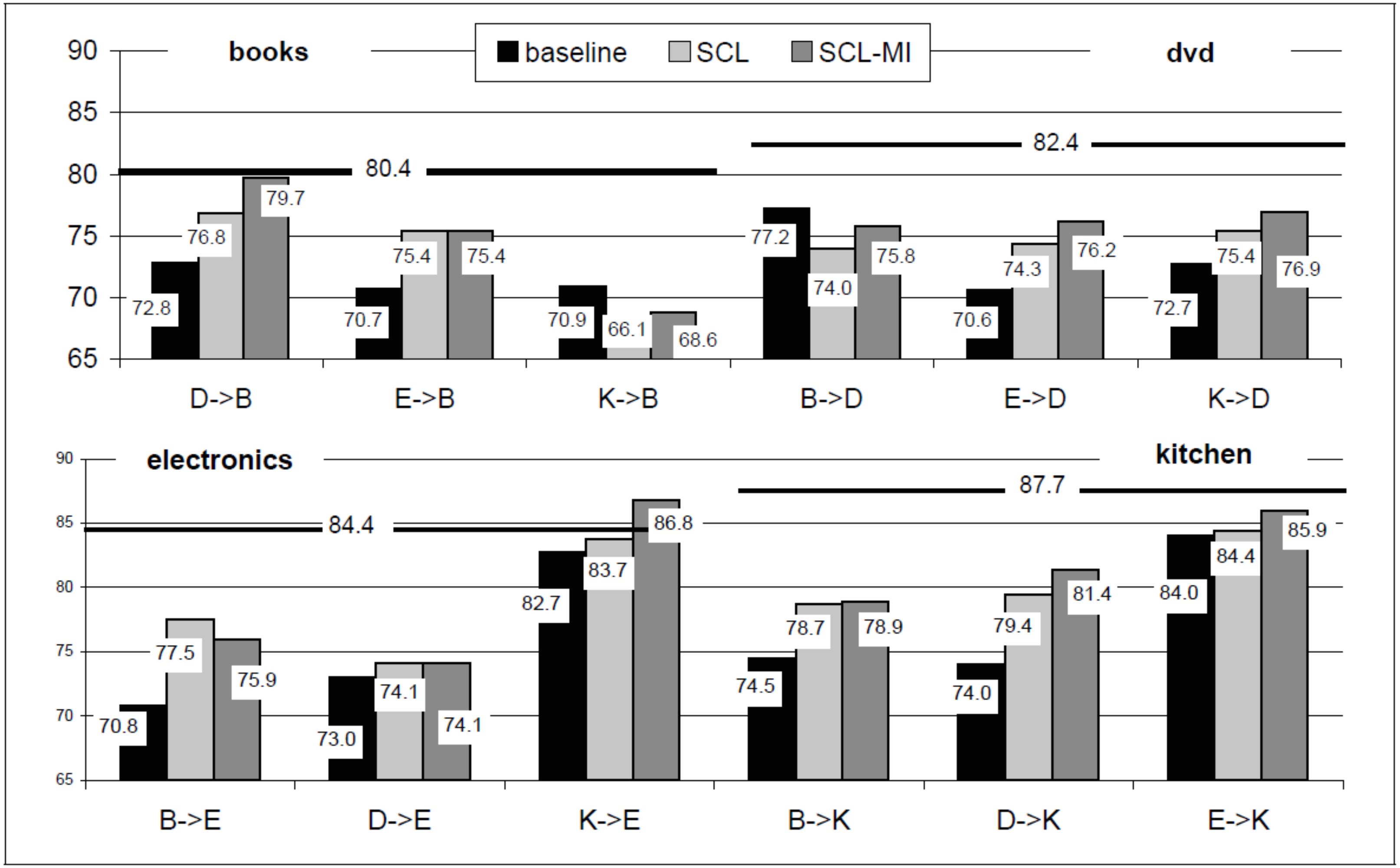


Figure 1: Accuracy results for domain adaptation between all pairs using SCL and SCL-MI. Thick black lines are the accuracies of in-domain classifiers.

Structural Correspondence Learning

Correcting Misalignments

- We now show how to use a small amount of target domain labeled data to learn to ignore misaligned projections from SCL-MI.

$$\min_{\mathbf{w},\mathbf{v}} \sum_i L\left(\mathbf{w}'\mathbf{x}_i + \mathbf{v}'\theta\mathbf{x}_i, y_i\right) + \lambda||\mathbf{w}||^2 + \mu||\mathbf{v}||^2 ,$$

- W = original features , v = projected features
- Suppose now that we have trained source model weight vectors w_s and v_s.
- A small amount of target domain data is probably insufficient to significantly change w, but we can correct v, which is much smaller.

$$\begin{aligned}\min_{\mathbf{w},\mathbf{v}} \textstyle\sum_j L\left(\mathbf{w}'\mathbf{x}_j + \mathbf{v}'\theta\mathbf{x}_j, y_j\right) + \lambda||\mathbf{w}||^2 \\ +\mu||\mathbf{v} - \mathbf{v}_s||^2 .\end{aligned}$$

Structural Correspondence Learning

Result

- As a baseline, we used the label of the source domain classifier as a feature in the target, but did not use any SCL features.

dom \ model	base	base +targ	scl	scl-mi	scl-mi +targ
books	8.9	9.0	7.4	5.8	**4.4**
dvd	8.9	8.9	7.8	6.1	**5.3**
electron	8.3	8.5	6.0	5.5	**4.8**
kitchen	10.2	9.9	7.0	5.6	**5.1**
average	9.1	9.1	7.1	5.8	**4.9**

Table 3: For each domain, we show the loss due to transfer for each method, averaged over all domains. The bottom row shows the average loss over all runs.

Domain Adaptation for Large–Scale Sentiment Classification

- **Domain Adaptation for Large–Scale Sentiment Classification : A Deep Learning Approach – Xavier Glorot , Antoine Bordes, Yoshua Bengio (2011)**
  - Source Domain and Target Domain is drawn from different distribution.
  - The learning problem consists in finding a function realizing a **good transfer** from S to T
  - Deep Learning algorithms learns **intermediate concepts** between raw input and target.
  - Our intuition for using it in this setting is that these **intermediate concepts could yield better transfer across domains.**
  - Amazon Review Data

Domain Adaptation for Large–Scale Sentiment Classification

- **Amazon Review Data**

Table 1. **Amazon data statistics.** This table depicts the number of training, testing and unlabeled examples for each domain, as well as the portion of negative training examples for both versions of the data set.

Domain	Train size	Test size	Unlab. size	% Neg. ex
Complete (large-scale) data set				
Toys	6318	2527	3791	19.63%
Software	1032	413	620	37.77%
Apparel	4470	1788	2682	14.49%
Video	8694	3478	5217	13.63%
Automotive	362	145	218	20.69%
Books	10625	10857	32845	12.08%
Jewelry	982	393	589	15.01%
Grocery	1238	495	743	13.54%
Camera	2652	1061	1591	16.31%
Baby	2046	818	1227	21.39%
Magazines	1195	478	717	22.59%
Cell	464	186	279	37.10%
Electronics	10196	4079	6118	21.94%
DVDs	10625	9218	26245	14.16%
Outdoor	729	292	437	20.55%
Health	3254	1301	1952	21.21%
Music	10625	24872	88865	8.33%
Videogame	720	288	432	17.01%
Kitchen	9233	3693	5540	20.96%
Beauty	1314	526	788	15.78%
Sports	2679	1072	1607	18.75%
Food	691	277	415	13.36%
(Smaller-scale) benchmark				
Books	1600	400	4465	50%
Kitchen	1600	400	5945	50%
Electronics	1600	400	5681	50%
DVDs	1600	400	3586	50%

Domain Adaptation for Large–Scale Sentiment Classification

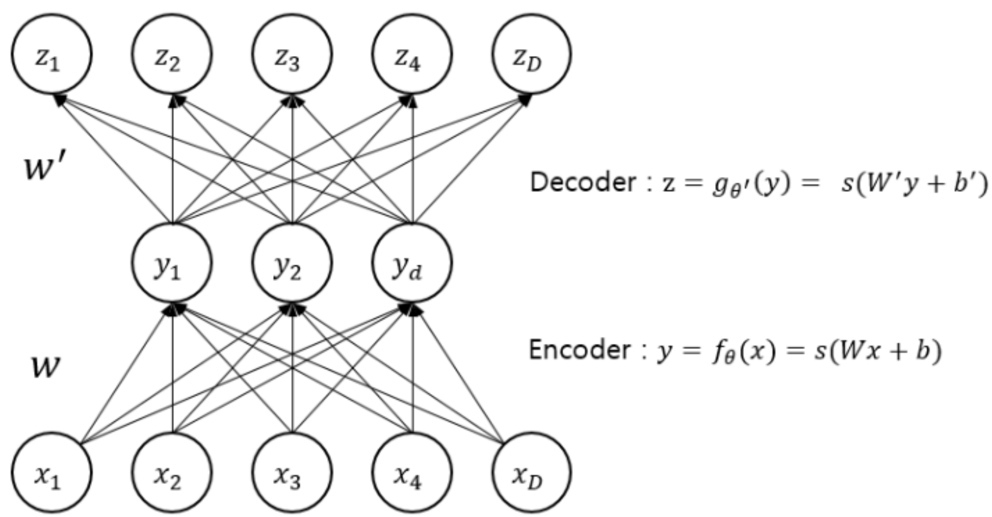
Applying Deep Learning

- First, they can **extract features** that somewhat disentangle the underlying factors of variation
  - This would likely help to perform transfer across domains, since we expect that there **exist generic concepts that characterize product reviews** across many domains.
- For our Amazon datasets, we know **some of these factors**(positive negative product)so we can use this knowledge to quantitatively check to what extent they are disentangled in the learned representation
- Finally, even though Deep Learning algorithms have not yet been evaluated for domain adaptation of sentiment classifiers several very **interesting results have been reported on other tasks involving textual data**

Domain Adaptation for Large–Scale Sentiment Classification

Stacked Denoising Auto-encoders

- An auto-encoder is comprised of an encoder function h() and a decoder function g(), typically with the dimension of h() smaller than that of its argument.
- The reconstruction of input x is given by r(x) = g(h(x)), and auto-encoders are typically trained to minimize a form of reconstruction error loss(x; r(x)).
- Once an auto-encoder has been trained, one can stack another auto-encoder on top of it, by training a second one which sees the encoded output of the first one as its training data.



Domain Adaptation for Large–Scale Sentiment Classification

Proposed Protocol

- In our setting we have access to unlabeled data from various domains, and to the labels for one source domain only
- First, a higher-level feature extraction is learnt in an unsupervised fashion from the text reviews of all the available domains using a **Stacked Denoising Autoencoder (SDA)**
  - Stochastic Gradient Descent
  - First layer = logistic sigmoid, training criterion = KL divergence
  - Upper layer = soft plus activation function, squared error, Gaussian corruption noise
- In a second step, a linear classier is trained on the transformed labeled data of the source domain.
  - **Linear SVM**

Domain Adaptation for Large–Scale Sentiment Classification

Experimental Setup

Books	1600	400	4465	50%
Kitchen	1600	400	5945	50%
Electronics	1600	400	5681	50%
DVDs	1600	400	3586	50%

Bag of Words

Most 5000 frequent words

- Baseline – linear SVM with raw data
  - Hyper parameter = by cross validation
- SDA
  - Hidden = 5000 – 2500 – 1000

Domain Adaptation for Large–Scale Sentiment Classification

Metrics

- e(S,T) = transfer error
  - Test error obtained by a method trained on the source domain S and tested on the Target domain T
- eb(T,T) = baseline in-domain error
  - test error obtained by the baseline method, i.e. a linear SVM on raw features trained and tested on the raw features of the target domain
- **Transfer Loss t**
  - t(S,T) = e(S,T) – eb(T,T)

Domain Adaptation for Large–Scale Sentiment Classification

Domain Adaptation for Sentiment Classification with Deep Learning

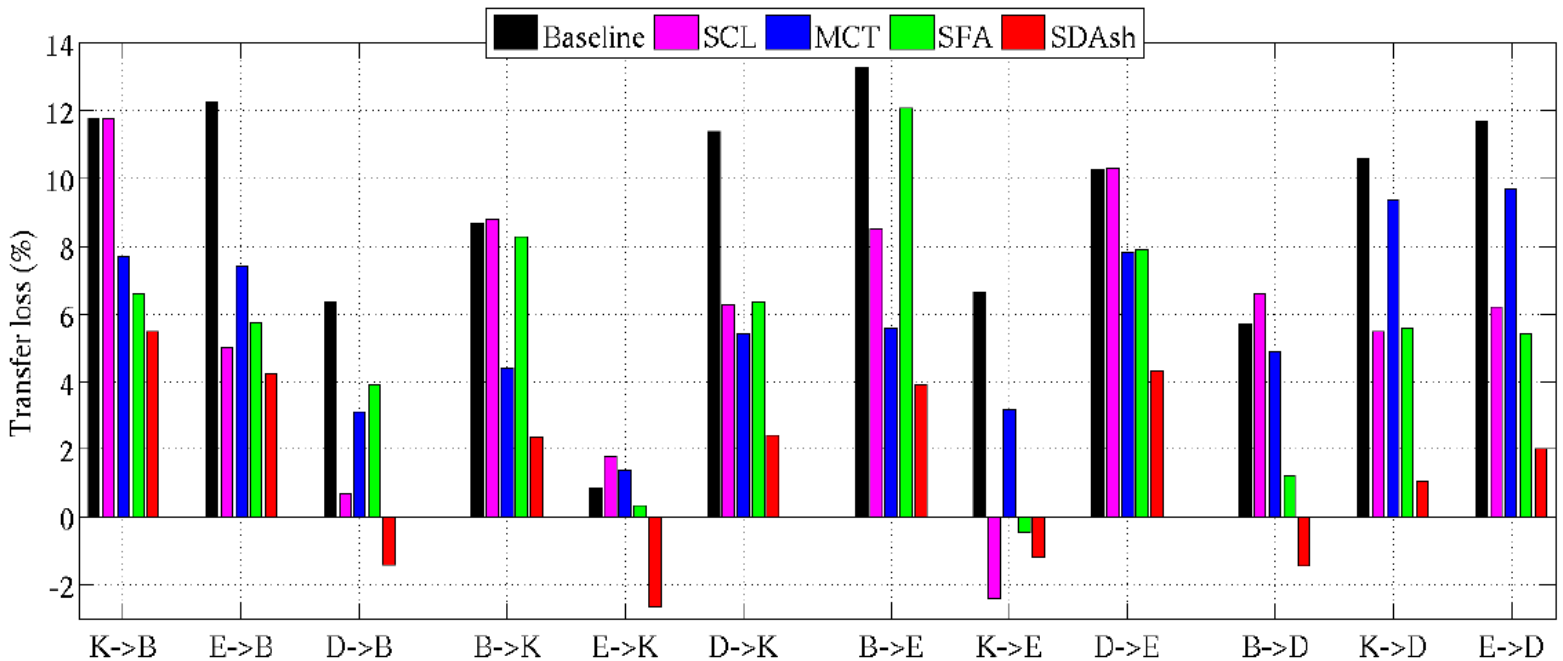


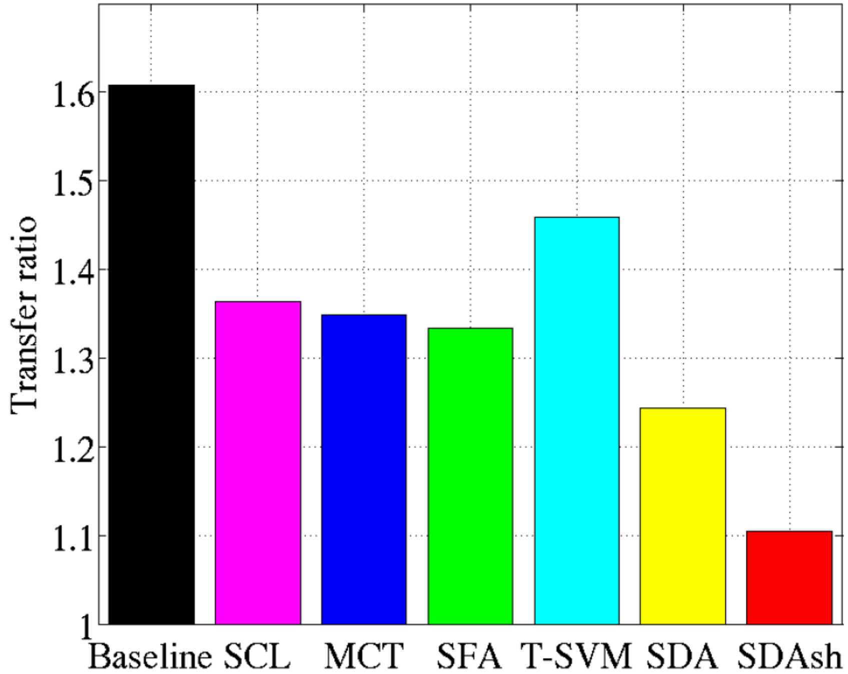
Figure 1. **Transfer losses on the Amazon benchmark** of 4 domains: *Kitchen*(K), *Electronics*(E), *DVDs*(D) and *Books*(B). All methods are trained on the labeled set of one domain and evaluated on the test sets of the others. SDA_{sh} outperforms all others on 11 out of 12 cases.

Domain Adaptation for Large–Scale Sentiment Classification

Metrics

- Transfer Ratio Q
- We report its mean over all source-target couples of the data set:

$$Q = \frac{1}{n} \sum_{(S,T)_{S \neq T}} \frac{e(S,T)}{e_b(T,T)}$$



Domain Adaptation for Large–Scale Sentiment Classification

- **Metrics**
  - **A-distance**
    - The A-distance is a measure of similarity between two probability distributions.