

基于“产生/判别”混合模型的分类器领域适应性问题研究

刘康, 赵军

中国科学院自动化研究所, 北京 100190
E-mail: {kliu, jzhao}@nlpr.ia.ac.cn

摘要: 本文针对自然语言理解中的领域适应性问题, 从领域概率分布的角度给出了一种基于混合模型的领域适应性学习方法, 它可以把判别式模型和产生式模型集成到一个模型当中。判别式模型比产生式模型有较好的分类效果, 但是产生式模型的优势在于在训练过程中可以较容易的引入非标注样本从而具有较好的推广性能。混合模型可以集中他们各自的优势。对于领域适应性, 我们的方法利用不同领域概率分布之间的差异性, 调节训练集中样本的权重, 从而使得训练得到的分类器更加适应于目标领域。在多个NLP任务应用本方法得到的实验结果可以证明本方法的有效性, 说明相对于传统监督学习以及半监督学习的方法, 本文方法有很大的优势, 同时也证明了混合模型的方法要优于单个模型的方法。

关键词: 领域适应性, 混合模型, 自然语言理解

Domain Adaptation in NLP based on Hybrid Generative and Discriminative Model

Kang Liu, Jun Zhao

Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
E-mail: {kliu, jzhao}@nlpr.ia.ac.cn

Abstract: This study investigates the domain adaptation problem for nature language processing tasks in the distributional view. A novel method is proposed for domain adaptation based on the hybrid model which combines the discriminative model with the generative model. The advantage of the discriminative model is to have lower asymptotic error, while the advantage of the generative model can easily incorporate the unlabeled data for better generalization performance. The hybrid model can integrate their advantages. For domain transfer, the proposed method exploits the difference of the distributions in different domains to adjust the weights of the instances in the training set so that the source labeled data is more adaptive to the target domain. Experimental results on several NLP tasks in different domains indicate that our method outperforms both the traditional supervised learning and the semi-supervised method.

Key Words: Domain adaptation, hybrid model, nature language processing

1. 引言

近些年, 统计学习方法在许多NLP任务中都得到了广泛的应用, 例如: 命名实体识别、句法分析、情感分析等。通常, 传统的统计学习方法都假设训练集与测试集具有相同的概率分布。但是在实际情况中, 这种假设往往不能成立, 训练集与测试集可能来源于不同的领域, 这种领域的差异性使得分类器的性能下降明显。我们期望分类器能够具有较好的领域适应性, 即就是在一个领域(源领域)上训练的分类器在另外一个领域(目标领域)上也能够得到较好的效果。这个问题称为统计分类器的领域适应性问题, 是本文研究的重点。

设计一个具有领域适应能力的统计分类器可以有两种途径: 一是设计一个产生式的分类器; 另一种是设计一个判别式分类器。产生式分类器学习样本 x 与标记

y 之间的联合概率 $p(x, y; \theta)$; 判别式分类器直接学习样本与标记直接的条件概率 $p(y/x; \theta)$ 。相对于产生式分类器, 判别式模型通常能够达到更好的分类效果[1], 但是在进行领域适应的时候往往需要目标领域内的非标注样本来指导训练, 判别式模型不容易引入非标注样本, 通常需要一些人为的假设, 例如最小熵原则[3]。而产生式模型在这一方面具有一定的优势, 比较容易的引入非标注样本, 因此产生式模型具有较强的推广性能[1]。

基于以上分析, 我们给出了一种基于产生/判别混合模型的学习方法, 能够利用两种模型各自的优势, 在达到较好的分类效果的同时也具有较强的推广性能。基于混合模型的监督学习、半监督学习方法近几年已经有如下研究工作[2] [4] [11] [12]。其中McCallum等给出了一种multi-conditional learning (MCL)方法, 它的目标函数

是多个条件目标函数的加权。可以有效地将产生式模型与判别式模型集成起来。McCallum也通过实验证明了这种方法的有效性。但是基于混合模型的分分类器适应性研究工作还很少。因此本文在MCL理论框架下给出了一种基于混合模型的领域适应学习方法，其由产生式以及判别式两个部分组成。在混合模型框架下，我们从领域概率分布的角度进行分类器的领域适应。通过计算目标领域与源领域之间的分布差异，我们的方法可以自动地调节源领域样本在学习过程中的权重。这样使得由源领域标注训练得到分类器更适应于目标领域。

同时，我们给出了该混合模型的训练方法以及相对应的目标函数，该目标函数是基于目标领域的概率分布，并由三个部分组成。第一部分是基于判别式模型的，由源领域中的样本训练得到；第二部分是基于产生式模型，也由源领域中的样本训练得到；最后一部分是基于边缘产生式模型，由目标领域中的非标注样本得到。三个部分根据领域的差异性得到不同的权重。同时，每个部分具有相同参数集。我们利用判别式训练方法可以对参数进行训练。在后面的章节中我们将给出的具体训练过程。

为了验证本文方法的有效性，本文方法应用于多个NLP任务上，例如实体识别、情感分类等。实验结果证明了我们的方法要优于传统的监督学习以及半监督学习方法，所训练的分类器更加适应于领域适应性问题。同时证明基于混合模型的方法也要优于单个模型（判别式或者产生式）的方法。我们也讨论了混合模型中每个部分权重的变化对参数的影响，从另一个角度也证明的混合模型的优势。

2. 基于混合模型的领域适应性方法

根据上文所述，本文方法是基于MCL框架的。所以在这一节我们首先介绍MCL的基本框架和理论。然后给出我们的方法，在MCL框架如何进行领域适应。并详细给出了训练方法以及参数设置的过程。

2.1 混合模型框架

多条件学习（MCL）是由McCallum[2]等人给出。它把产生式模型与判别式模型混合到一个单一模型中，所产生的多条件目标函数是由产生式模型与判别式模型加权组成的。目标函数可以写成如下式子：

$$\begin{aligned} L_{mcl} &= \alpha \log P(Y / X; \theta) + \beta \log P(X / Y; \theta) \\ &= \alpha L_{dis}(\theta) + \beta L_{gen}(\theta) \end{aligned} \quad (1)$$

这里 Y 指的是样本的标记， X 是指训练样本。 α 和 β 分别是判别式目标函数 $L_{dis}(\theta)$ 以及产生式目标函数 $L_{gen}(\theta)$ 的权重。要得到最优参数，必须使得 L_{mcl} 最大。即：

$$\hat{\theta} = \arg \max_{\theta} (\alpha L_{dis}(\theta) + \beta L_{gen}(\theta)) \quad (2)$$

在混合模型中，判别式部分与产生式部分具有相同的参数集，我们假设参数符合高斯分布，则有：

$$\log(p(\theta)) \propto -\frac{\|\theta\|^2}{2\sigma^2} \quad (3)$$

2.2 基于混合模型的领域适应性方法

在本节中，我们将详细描述如何利用混合模型使得分类器适应于另外一个领域。假设在源领域 \mathcal{D}_s 上有一个标注训练集 D_s ，源领域的概率分布为 $p_s(\cdot)$ ，同时在目标领域 \mathcal{D}_t 有未标注的测试集 D_t ，目标领域的概率分布为 $p_t(\cdot)$ 。我们定义联合分布函数 $p(X, Y; \theta)$ ，这里 X 是训练样本， Y 是标记， θ 是参数集。我们的目标是在目标领域的概率分布 $p_t(\cdot)$ 下求使得上面所示的联合分布最大的参数 θ 。则目标函数可写成：

$$\begin{aligned} L(\theta) &= \log(p_t(X, Y; \theta)) \\ &= \log(p_t(Y / X; \theta)) + \log(p_t(X; \theta)) + \log(p_t(\theta)) \end{aligned} \quad (4)$$

在MCL框架下，上式中的每个部分可以配以不同的权重，则有：

$$\begin{aligned} L(\theta) &= \alpha \log(p_t(Y_s / X_s; \theta)) \\ &\quad + \beta \log(p_t(X_s; \theta)) \\ &\quad + \beta \log(p_t(X_t; \theta)) + \log(p_t(\theta)) \end{aligned} \quad (5)$$

这里 X_s 是源领域中的标记样本， Y_s 是他们的标记； X_t 是目标领域内的非标记样本。从这个式子，我们看出通过混合模型，我们可以很容易的把为标记样本引入的训练中，而不需要任何的假设条件。

对于领域适应性，必须利用源领域内的标注样本。对于(5)式中的 $\log(p_t(Y_s / X_s; \theta))$ ，我们用 $p_s(Y_s / X_s; \theta)$ 去近似 $p_t(Y_s / X_s; \theta)$ 。对于 $\log(p_t(X_s; \theta))$ ，可以重写为

$$\log\left(\frac{p_t(X_s; \theta)}{p_s(X_s; \theta)} p_s(X_s; \theta)\right); p(X_s / D_s; \theta) \text{ 和 } p(X_t / D_t; \theta)$$

分别用来逼近 $p_s(X_s; \theta)$ 和 $p_t(X_t; \theta)$ 。对于源领域中的样本，每一个样本都已经被标记，所以我们用联合分布 $p_s(Y_s, X_s; \theta)$ 替代边缘分布 $p_s(X_s; \theta)$ ；对于目标领域内的样本，由于标记未知，边缘分布 $p_t(X_t; \theta)$ 可以写为联合分布对于标记的和，即 $\sum_Y p_t(Y, X_t; \theta)$ 。则原目标函数可以重写为如下形式：

$$\begin{aligned}
L(\theta) &= \alpha \sum_{x \in D_s} \log(p_s(y_s / x_s; \theta)) \\
&+ \beta \sum_{x \in D_s} \log(\lambda_s p(x_s, y_s / D_s; \theta)) \\
&+ \beta \sum_{x \in D_t} \log(\sum_y p(x_t, y / D_t; \theta)) \\
&- \|\theta\|^2 / \sigma^2
\end{aligned} \tag{6}$$

从上面的式子，我们可以看出源领域中的每一个训练样本的权重被改变为 $\lambda_s = \frac{p_t(x_s; \theta)}{p_s(x_s; \theta)}$ ，在下面的章节中我们讲如何计算这个参数。

2.3 参数学习

对于学习最优参数 $\hat{\theta}$ ，我们利用判别式学习方法得到最优参数使得(6)式中的目标函数最大。为了寻优，必须得到目标函数的梯度 $\nabla L(\theta)$ 。我们用一个朴素马尔可夫随机场来表示每一个样本，这里朴素马尔可夫随机场模型是一个无向图模型。则样本 x 与标记 y 的联合分布可以表示为：

$$p(x, y / \theta) = \frac{1}{Z} \exp(y^T \theta^T x) \tag{7}$$

这里 $Z = \sum_{x \in D} \sum_y \exp(y^T \theta^T x)$ 并且 $D = D_s \cup D_t$ 。

对于后验概率 $p(y / x; \theta)$ ，可以写为

$$p(y / x; \theta) = \frac{\exp(y^T \theta^T x)}{\sum_y \exp(y^T \theta^T x)} \tag{8}$$

在某个特定领域内，样本 x 与标记 y 的联合分布函数可以表示为

$$p(x, y / D_k; \theta) = \frac{\exp(y^T \theta^T x)}{\sum_{x \in D_k} \sum_y \exp(y^T \theta^T x)} \quad k \in \{s, t\} \tag{9}$$

基于公式(8)，式(6)中 $L(\theta)$ 的判别式模型可以写为如下式子：

$$\log(p_s(\bar{y}_s / \bar{x}_s)) = \bar{y}_s^T \theta^T \bar{x}_s - \log(\sum_y \exp(y^T \theta^T \bar{x}_s)) \tag{10}$$

这里 \bar{y} 和 \bar{x} 是对于源领域内所标记的样本的观察。它的梯度可以通过下式计算：

$$\nabla \log(p_s(\bar{y}_s / \bar{x}_s)) = \bar{x}_s \bar{y}_s^T - \sum_y p_s(y / \bar{x}_s) \bar{x}_s y^T \tag{11}$$

同样的， $L(\theta)$ 中的第二部分可以通过式(9)计算得：

$$\begin{aligned}
&\log(\lambda_s p(\bar{x}_s, \bar{y}_s / D_s; \theta)) \\
&= \log \lambda_s + \bar{y}_s^T \theta^T \bar{x}_s - \log(\sum_{x \in D_s} \sum_y \exp(y^T \theta^T x))
\end{aligned} \tag{12}$$

并且可以得到它的梯度表达式：

$$\begin{aligned}
&\nabla \log(\lambda_s p(\bar{x}_s, \bar{y}_s / D_s; \theta)) \\
&= \bar{x}_s \bar{y}_s^T - \frac{\sum_{x \in D_s} \sum_y \exp(y^T \theta^T x) x y^T}{\sum_{x \in D_s} \sum_y \exp(y^T \theta^T x)}
\end{aligned} \tag{13}$$

$L(\theta)$ 中的第三部分是由目标领域的未标注样本训练得到的。它可以被看做一个边缘产生式模型。则其似然目标函数表达式可以写作：

$$\begin{aligned}
&\log(\sum_y p(x_t, y / D_t; \theta)) \\
&= \log(\sum_y \exp(y^T \theta^T x_t)) - \log(\sum_{x \in D_s} \sum_y \exp(y^T \theta^T x))
\end{aligned} \tag{14}$$

它的梯度表达式为：

$$\begin{aligned}
&\nabla \log(\sum_y p(x_t, y / D_t; \theta)) \\
&= \frac{\sum_y \exp(y^T \theta^T x_t) x_t y^T}{\sum_y \exp(y^T \theta^T x_t)} - \frac{\sum_{x \in D_s} \sum_y \exp(y^T \theta^T x) x y^T}{\sum_{x \in D_s} \sum_y \exp(y^T \theta^T x)}
\end{aligned} \tag{15}$$

得到了目标函数的表达式以及它的梯度，我们利用 Limited-memory BFGS (L-BFGS) 算法来进行寻优。这种方法也广泛的应用于很多需要进行判别式训练的方法中，例如最大熵，条件随机场等。

2.4 设置参数 λ_s

在式(6)中， $\lambda_s = \frac{p_t(x_s; \theta)}{p_s(x_s; \theta)}$ 反映了源领域与目标领域的差异程度。为了计算这个参数，假设每篇文档中的每一个词都是相互独立的，则我们通过 unigram 语言模型去计算这个参数，则其可以表示为如下表达式：

$$\frac{p_t(x_s; \theta)}{p_s(x_s; \theta)} = \frac{\prod_{w \in x_s} p_t(w)}{\prod_{w \in x_s} p_s(w)} \approx \frac{\prod_{w \in x_s} p(w / D_t)}{\prod_{w \in x_s} p(w / D_s)} \tag{14}$$

这里 w 是每篇文档中的词； $p(w / D_i)$ 表示在每个特定领域内词的分布情况，可以有下式得到。

$$p(w / D_i) = N(w, D_i) / N_{D_i} \tag{15}$$

这里 $N(w, D_i)$ 表示词 w 在文档 $D_i, i \in \{s, t\}$ 中的数目，而 N_{D_i} 表示在文档 D_i 中所有词的数目。

2.5 判别式模型与产生式模型权重的设置

在式(6)中, α 和 β 分别表示了判别式模型与产生式模型的权重, 这两个参数的选择对于分类器产生巨大的影响。在 MCL 中, 如果 $\alpha > \beta$, 说明混合模型将偏向于判别式模型的结果, 相反当 $\alpha < \beta$, 混合模型将偏向于产生式模型的结果。但是在本文所给出的混合模型中, 判别式部分与产生式部分在不同的尺度上进行训练。其中判别式部分是基于类别的概率模型, 而产生式部分是基于各自领域内样本的概率模型。判别式概率模型的分母要小于产生式模型的分母, 这说明判别式模型的变化有可能会掩盖产生式模型的变化。如果设定 $\alpha = \beta$, 并不能说明本文的混合模型对于两类模型赋予了相同的权重。

McCallum[4] 利用交叉验证来估计最优的 α 和 β 。但是, 在目标领域内, 我们无法得到样本的标记, 这使得在源领域内的标记样本上训练得到的 α 和 β 对于目标领域不一定是最优的。所以 α 和 β 的设定是十分困难的。对于非监督学习, Gregory 等[4] 设定 $\alpha \gg \beta$ 来对判别式模型和产生式模型的权重进行归一化。但是在领域适应性问题中, 未标记的样本不一定比标记样本的数目多; 同时, 由于我们希望训练的分类能适应于目标领域, 所以需要给产生式模型赋予更大的权重。所以我们并没有设定 $\alpha \gg \beta$ 。在我们的实验中我们设定 $\alpha = 1$, 同时变化 $\beta \in [10^{-5}, 10^4]$, 以此来讨论参数对于模型性能的影响。

3. 实验

在这一节, 将给出实验及其结果, 我们主要进行如下实验: 1) 比较本文方法与传统监督学习以及半监督学习方法在领域适应性问题上的优劣; 2) 比较单一模型(判别式或者产生式)和混合模型的差异; 3) 讨论在混合模型中 α 和 β 对分类器性能的影响。在实验之前, 我们首先对数据集以及Baseline进行描述。

2.1 实验设置

我们选择两个NLP任务来测试我们方法的有效性。一是情感分类任务, 我们采用与Blitzer[9] 相同的数据集, 其中包含了来自于Amazon的产品评论, 语料来源于四个领域: 书(book)、DVD(dvd)、电子产品(electronic)、厨房用品(kitchen), 每个领域包含2000篇文档, 标记集为{正面, 反面}, 我们从每个领域中随机选取1600篇文档做为训练样本, 另外400篇做为测试样本。二是实体识别任务, 我们采用ACE2005年的语料, 语料分别来自于广播新闻(bn)、网页(wl)以及新闻专线(nw), 我们用一个领域训练, 在另外一个领域上测试。标记集为{B, I, O}。

为了和我们的方法进行比较, 我们选择传统的监督学习以及半监督学习方法做为baseline, 其中包括基于产生式模型的方法以及判别式模型的方法。对于监督学习方法, 我们选择Naive Bayes (NB)、支撑矢量机(SVM)以及最大熵(MaxEnt), 对于半监督学习方法, 我们选择EM-NB、以及Semi-supervised MCL (S-MCL)方法[4]。对于NB以及EM-NB, 我们采用Laplace平滑方法, SVM分类器采用LibSVM¹工具, 最大熵采用Mallet²工具包。所有方法采用相同的特征。

2.2 我们的方法vs.监督学习、半监督学习方法

在实验中我们的分类器只由在一个领域上进行训练, 在另外一个领域上进行测试, 表1和表2给出了实验结果, 其中表1是对情感分类任务上的分类结果; 表2是在实体识别任务上的分类结果。第一列是数据集, 第一个字母表示训练领域, 第二个字母表示测试领域。例如“B-D”表示分类器在书籍领域上进行训练, 在DVD领域上进行测试; 又“wl-nw”表示在wl领域上训练, 在nw领域上测试。

从实验结果中, 可以看出我们的方法在两个任务上的效果都要优于监督学习的方法(NB, SVM, MaxEnt)。这说明我们的方法在领域分布不同的条件下更加适应于目标领域, 因为在训练的时候我们的方法考虑了目标领域里的分布信息。同时, 把半监督方法的结果与我们方法的结果相比较, 可以看出我们的方法也要优于传统半监督学习方法(EM-NB, S-MCL)。这是由于我们的方法利用参数 λ_s 区别了不同领域的概率分布差异性。同时我们还注意到, 由于传统半监督学习没有区分领域之间的差异性, 在很多数据集上效果往往不如监督学习的方法, 比如说EM-NB相对于NB性能下降了很多, 尽管S-MCL平均性能高于MaxEnt, 但是在有些数据集上还略有下降。而我们的方法几乎在所有的数据集上都能达到最好的效果。

2.3 混合模型vs.单一模型

我们还比较了基于混合模型的方法和单一模型方法的差异。从表1、表2中NB、EM-NB是基于产生式模型的, SVM、MaxEnt都是基于判别式模型的。而S-MCL与我们的方法是基于混合模型的。从实验结果中, 我们可以看出S-MCL和我们的方法在平均精确率上要优于其它方法, 这说明基于混合模型的方法要优于单一模型的方法, 我们的方法在混合模型的基础上进而考虑了领域概率分布信息的差异性, 所以能够进一步的提高分类的精度。

¹ <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

² http://mallet.cs.umass.edu/index.php/Main_Page

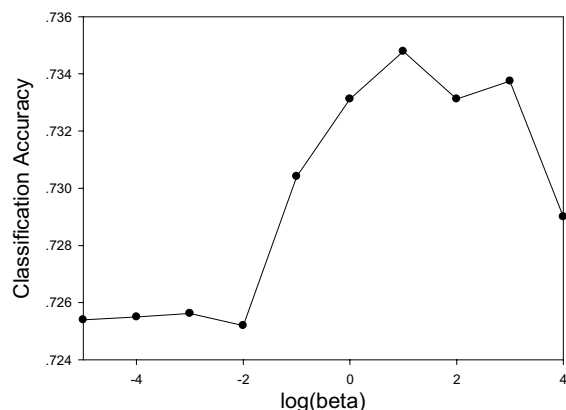
表 1. 情感分类结果

Data Set	NB	SVM	MaxEnt	EM-NB	S-MCL	Ours
D-B	0.715	0.7325	0.7125	0.6925	0.7375	0.7575
E-B	0.6675	0.675	0.68	0.675	0.675	0.715
K-B	0.6675	0.6725	0.6725	0.63	0.6725	0.7075
B-D	0.74	0.7425	0.7325	0.7475	0.735	0.7875
E-D	0.6875	0.6975	0.7025	0.70	0.70	0.7075
K-D	0.6975	0.695	0.695	0.7025	0.69	0.7025
B-K	0.6925	0.715	0.7275	0.56	0.7275	0.76
D-K	0.6725	0.7125	0.7225	0.56	0.7125	0.73
E-K	0.7625	0.82	0.81	0.7725	0.8225	0.8425
B-E	0.6275	0.6775	0.67	0.5375	0.6775	0.6875
D-E	0.63	0.73	0.72	0.5375	0.7225	0.74
K-E	0.7425	0.7875	0.7825	0.7325	0.7825	0.815
Mean	0.6919	0.7215	0.7190	0.6540	0.7213	0.7460

表2. 实体识别的结果

Data Set	NB	SVM	MaxEnt	EM-NB	S-MCL	Ours
ns-wl	0.7244	0.761	0.7412	0.6975	0.7504	0.774
ns-bn	0.7	0.7544	0.7186	0.6747	0.7209	0.7427
wl-bn	0.66	0.7368	0.718	0.6134	0.6842	0.744
wl-ns	0.6546	0.7329	0.716	0.631	0.7133	0.7341
bn-ns	0.7177	0.7291	0.7088	0.7202	0.7215	0.736
bn-wl	0.7491	0.7972	0.7412	0.7185	0.7446	0.7755
Mean	0.7009	0.7519	0.7239	0.6759	0.7224	0.7511

2.4 权重参数 α 和 β 对分类器性能的影响

图1. $\log(\beta)$ 对于分类器性能的影响

在这个实验中,我们讨论了模型权重参数 α 和 β 对于实验结果的影响。由前面章节可知设置 α 和 β 是十分困难的,所以在实验中我们固定 $\alpha = 1$, 同时变化 $\beta \in [10^{-5}, 10^4]$, 来讨论其对分类器的影响。当 β 增大, 我们的模型趋向于产生式模型, 当 β 减少, 我们的模型趋向于判别式模型。图1给出了实验结果, 由于空间的原因, 我们只给出了情感分类结果精确率平均值的变化曲线。从图中可以看出我们的方法在 $\beta = 10$ 附近达到最优, 这也进一步证明混合模型要优

于单一模型, 同时由于分类器要适应于目标领域, 所以我们的模型略微偏向于产生式模型。

4. 相关工作

与本文工作相关的方法包括两个方面, 一是在领域适应性方面的工作, 另一方面是基于混合模型的方法。

目前针对领域适应性的研究工作主要包含两个角度: 一是特征表示角度, 二是概率分布角度。从特征表示角度出发, 以往研究的重点放在对于目标领域如何选取有效的特征[6] [8] [9] [10] [13]。而本文方法是从概率分布的角度进行领域适应的, 通过计算不同领域之间的概率分布差异度来调教训练样本的权重, 并没有对特征进行领域层面的筛选。从概率分布角度出发的领域适应性方法包括[5] [7] [13] [15]。其中 Dai 等[15] 给出了一种 Transfer Learning Naïve Bayes 方法。在 EM 迭代的过程中加入两个领域的 KL 距离信息, 这种方法是基于产生式模型的。Tan 等 [13] 给出了一种 Similarity Ranking 算法, 利用源领域中标注样本从目标领域中筛选有用的目标样本, 共同来训练分类器, 但这种方法没有区别分布的差异性。Hal Daumé III 给出了一种产生式的图模型[5], 它是一个产生式模型, 包含领域特有的分布以及领域之间共有的分布信息。他们利用 CEM 算法来进行参数估计, 相

对传统方法可以达到不错的效果，但算法复杂度较高，同时需要先验的假设。

基于混合模型的工作主要集中在监督学习和半监督学习领域[1][2][4][11][12]。其中R. Raina等[12]将Naïve Bayes与Logistic回归混合起来进行文本分类，在他们的方法中一篇文本被分成很多个区域，然后利用判别式模型训练区域的权重，利用产生式模型训练各个区域内部的参数。McCallum等[2]给出了一种multi-conditional learning (MCL)方法，它的目标是函数是多个条件目标函数的加权。在MCL框架下，Gregory等给出了一种基于混合模型的半监督学习方法。多条件目标函数是由加权的判别式目标函数以及产生式目标函数组成的，两个部分具有相同的参数集。Julia A.等也给出了一种parameter coupling prior (PCP)方法[11]。在他们的方法中，判别式部分与产生式部分的参数是分开的，利用一个先验函数来表示这两类参数之间的关系。本文的工作是在MCL的框架下利用目标领域内的概率分布信息，使得训练的分类器更加适应于目标领域，从而达到领域适。

5. 结论

本文针对NLP任务中领域适应性问题，给出了一种基于混合产生式与判别式模型的方法，使得分类器在训练样本与测试样本分布不同的条件下能够适应于目标领域。通过计算源领域与训练领域概率分布的差异度，调节训练样本中每个样本的权重，使得分类器适应于目标领域。通过在两个NLP任务上的实验，对比传统监督学习和半监督学习方法，本文方法能够取得更好的分类效果。同时我们也证明了混合模型的方法要优于单一模型的方法。

在以后的工作中，我们将致力于抽取更有效的特征，建立源领域与目标领域特有特征之间的联系，使得分类器更加适应于目标领域。

致谢

本文受国家自然科学基金项目(60673042)以及北京市自然科学基金项目(4052027, 4073043)的资助。

参考文献

[1] Jun Suzuki, Akinori Fujino and Hideki Isozaki. *Semi-Supervised Structured Output Learning based on a Hybrid Generative and Discriminative Approach*. In *Proc. of EMNLP-CoNLL*, 2007, pages 791-800.

[2] Andrew McCallum, Chris Pal, Greg Druck and Xuerui Wang. *Multi-Conditional Learning: Generative/Discriminative Training for Clustering and Classification*. In *Proc. of AAAI*, 2006.

[3] Feng Jiao, Shaojun Wang, Chi-Hoon Lee, Russell Greiner and Dalse Schuurmans. *Semi-Supervised Conditional Random Fields for Improved Sequence Segmentation and Labeling*. In *Proc. of ACL*, 2006, pages 209-216.

[4] Gregory Druck, Chris Pal, Xiaojin Zhu and Andrew McCallum. *Semi-Supervised Classification with Hybrid Generative/Discriminative Methods*. In *Proc. of KDD*, 2007, pages 280-289.

[5] Hal Daumé III and Daniel Marcu. *Domain Adaptation for Statistical Classifiers*. *Journal of Artificial Intelligence Research* 26, 2006, pages 101-126.

[6] Hal Daumé III. *Frustratingly Easy Domain Adaptation*. In *Proc. of ACL*, 2007.

[7] Jin Jiang and ChengXiang Zhai. *Instance Weighting for Domain Adaptation in NLP*. In *Proc. of ACL*, 2007, pages 264-271.

[8] Jin Jiang and ChengXiang Zhai. *A Two-stage Approach to Domain Adaptation for Statistical Classifiers*. In *Proc. of CIKM*, 2007, pages 401-410.

[9] John Blitzer, Mark Dredze and Fernando Pereira. *Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification*. In *Proc. of ACL*, 2007, pages 440-447.

[10] John Blitzer, Ryan McDonald and Fernando Pereira. *Domain with Structural Corresponding Learning*. In *Empirical Methods in Natural Language Processing (EMNLP)*, 2006, pages 120-128.

[11] Julia A. Lasserre, Christopher M. Bishop and Thomas P. Minka. *Principled Hybrids of Generative and Discriminative Models*. In *Proc. of CVPR*, 2006, pages 87-94.

[12] R. Raina, Y. Shen, A. Y. Ng, and Andrew McCallum. *Classification with Hybrid Generative/Discriminative Models*. In *NIPS*, 2003.

[13] Sandeepkumar Satpal and Sunita Sarawagi. *Domain Adaptation of Conditional Probability Models via Feature Subsetting*. In *Proc. of PKDD*, 2007.

[14] Songbo Tan, Gaowei Wu, Huifeng Tang and Zueqi Cheng. *A Novel Scheme for Domain-transfer Problem in the context of Sentiment Analysis*. In *Proc. of CIKM 2007*.

[15] Wenyan Dai, Gui-Rong Xue, Qiang Yang and Yong Yu. *Transferring Naïve Bayes Classifiers for Text Classification*. In *Proc. of AAAI*, 2007.