

科研训练2017论文阅读报告

汪蓉. 句子级情绪分类方法研究[D]. 苏州大学, 2016.

科研训练2017论文阅读报告

[1 基于整数线性规划的情感和情绪联合学习](#)

[1.1 方法](#)

[1.2 实验](#)

[2 基于标签关系的句子级情绪分类方法](#)

[2.1 多标签间的依赖关系](#)

[2.2 基于标签关系的句子级情绪分类](#)

[2.3 实验](#)

[3 基于上下文关系的句子级情绪分类](#)

[3.1 基准系统](#)

[3.2 准备工作](#)

[3.3 因子图模型](#)

[3.4 模型学习和预测](#)

[3.5 基于上下文关系的句子级情绪分类方法的实现](#)

[4 参考文献](#)

1 基于整数线性规划的情感和情绪联合学习

单标签的句子级情绪分类问题 multi-class emotion classification

多次二分类

1.1 方法

- 1) 首先作者分开训练情感分类器和情绪的分类器；然后将这两类分类器分别作用于同一个测试数据，分别得到初始的情感分类和情绪分类结果；
- 2) 最后作者用**整数线性规划方法**作用于初始的情感分类和情绪分类结果，从而得到优化后的情感分类和情绪分类结果。为了充分利用情感分类和情绪分类任务之间的联系，作者将情感倾向和个人情绪之间的联系转化为整数线性规划方法中的一系列限制条件。

特点：

- 1) 联合学习
- 2) 不需要额外的标注数据

整数线性规划方法是一种**基于限制条件**来寻找一组**最优变量**的数学方法。具体表现为，在满足特定数量的限制条件的基础上，最小化目标损失函数（Loss Function）。

作者定义 \mathbf{x} 为测试样本的特征向量。由于在情感分类和情绪分类中，文本的表示是相同的，例如，词袋（Bag-of-words）表示，所以句子级情感分类器或句子级情绪分类器都可以对测试样本 \mathbf{x} 进行分类。

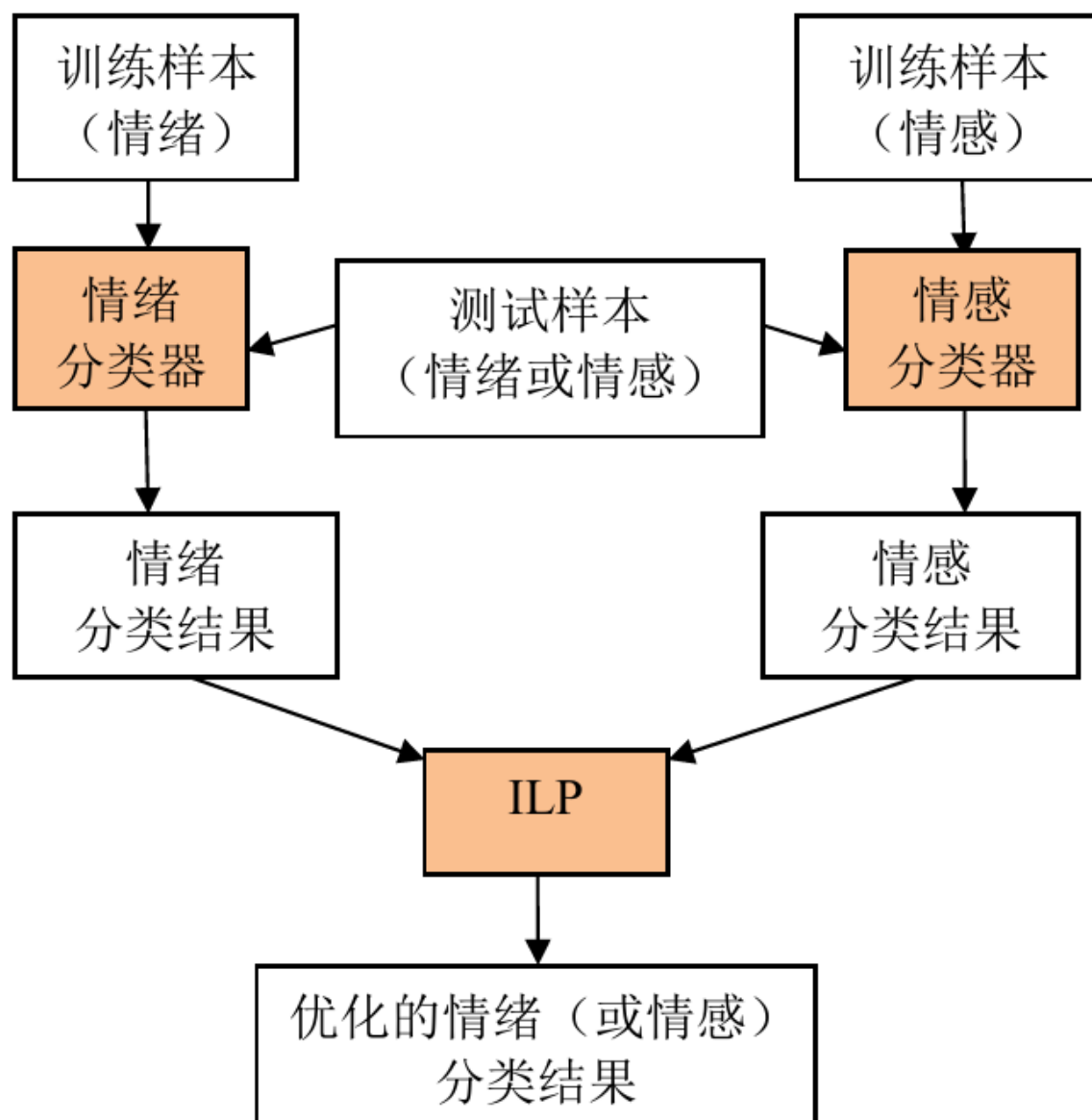


图 3-3 联合学习框架

在情感分类中，测试样本被分类为情感类别，例如，正面或者负面。假设，第 i 个测试样本的情感类别为 y_i ，且 $y_i \in \{0,1,2\}$ ，其中0代表中立类别，1代表正面类别，2代表负面类别。为了方便使用布尔值表示结果，们进一步定义了三个不同的变量，例如， $y_{0i} \in \{0,1\}$ ， $y_{1i} \in \{0,1\}$ ， $y_{2i} \in \{0,1\}$ ，分别代表了分类的结果是否中立，正面或者负面类别。因此，第 i 个测试样本的情感分类结果可以用一个向量表示，例如 $\langle y_{0i}, y_{1i}, y_{2i} \rangle$ 。另外，基于统计学的分类器会给出每个分类标签的预测概率，例如：

$$P_y = \langle P_{y_{0i}=1}, P_{y_{0i}=0}, P_{y_{1i}=1}, P_{y_{1i}=0}, P_{y_{2i}=1}, P_{y_{2i}=0} \rangle$$

其中， $P_{y_{0i}=1}$ 代表了第 i 个测试样本的标签变量 $y_{0i}=1$ 的概率。

相应的标签信息表示为：

$$Y = \langle y_{0i}, 1-y_{0i}, y_{1i}, 1-y_{1i}, y_{2i}, 1-y_{2i} \rangle$$

在情绪分类中，测试样本被分类为情绪类别。假设第 i 个测试样本的情绪类别为 Z_i ，且 $Z_i \in \{0,1,2,3,4,5,6,7,8\}$ ，其中0-8 分别代表neutral、expect、joy、love、surprise、anxiety、sorrow、angry 和hate 等情绪类别。同样的，为了方便使用布尔值表示结果，作者定义了九种不同的变量， $z_{0i-8i} \in \{0,1\}$ 。

相应的概率和标签结果表示为：

$$P_z = \langle P_{z_{0i}=1}, P_{z_{0i}=0}, P_{z_{1i}=1}, P_{z_{1i}=0}, \dots, P_{z_{8i}=1}, P_{z_{8i}=0} \rangle$$

以及

$$Z = \langle z_{0i}, 1-z_{0i}, z_{1i}, 1-z_{1i}, \dots, z_{8i}, 1-z_{8i} \rangle$$

基于整数线性规划的联合学习的目的是根据一系列限制条件，在使目标函数最小的前提下调整情感分类和情绪分类的结果。在本章中，作者设置的目标函数，目的是使标签结果和概率结果尽可能相似。作者使用 Cosine 方法来计算标签向量和概率向量之间的相似度。因此，作者的目标转化为最大化以下公式：

$$Max \frac{\langle P_y, P_z \rangle \cdot \langle Y, Z \rangle}{|\langle P_y, P_z \rangle| |\langle Y, Z \rangle|}$$

其中，概率向量的结果是两个初始分类器给定的，所以 $|\langle P_y, P_z \rangle|$ 是定值。另外，作者发现，在假设每个测试样本只有一个情感标签和一个情绪标签的前提下， $|\langle Y, Z \rangle|$ 也是定值。因此，作者的目标函数也转化为：

$$Max \langle P_y, P_z \rangle \cdot \langle Y, Z \rangle$$

以上问题可看作整数线性规划的问题。目标函数可以详细表示为：

$$Max \sum_{k=0}^2 P_{y_{ki}=1} \cdot y_{ki} + \sum_{k=0}^2 P_{y_{ki}=0} \cdot (1 - y_{ki}) + \sum_{l=0}^8 P_{z_{li}=1} \cdot z_{li} + \sum_{l=0}^8 P_{z_{li}=0} \cdot (1 - z_{li}) \quad (3.7)$$

目标函数确定后，接下来就是设置一些列的限制条件：

(C1) 整数限制：

$$y_{ki} \in \{0,1\} \text{ 以及 } z_{li} \in \{0,1\}$$

(C2) 单标签限制：

$$\sum_{k=0}^2 y_{ki} = 1, \text{ and } \sum_{l=0}^8 z_{li} = 1$$

(C3) 中立情绪限制：

当测试样本在情感分类中被分类为中立时，情绪标签也必须是中立：

$$y_{0i} = z_{0i}$$

(C4) 正面情绪限制：

当测试样本在情感分类中被分类为正面时，情绪标签必须是集合 { joy, love, expect, surprise } 中的一个：

$$y_{1i} = Z_{1i} + Z_{2i} + Z_{3i} + Z_{4i}$$

(C5) 负面情绪限制:

当测试样本在情感分类中被分类为负面时, 情绪标签必须是集合 { anger, anxiety, hate, sorrow } 中的一个:

$$y_{2i} = Z_{5i} + Z_{6i} + Z_{7i} + Z_{8i}$$

1.2 实验

(1) 实验设置

在本章实验中, 作者只考虑情绪主要情绪。另外, 作者在原有的 8 种情绪基础上, 加入了中立情绪 neutral, 即当一句话不含有这 8 种情绪的任意一种情绪时, 作者将此句子的情绪标签定义为 neutral。

实验使用的初始分类算法是基于公共工具 **Mallet 工具包** 的最大熵分类方法。选用每个句子的词袋特征, 并且根据词的出现与否, 将特征处理成布尔值向量的形式。线性规划方法, 作者选用了 **lp_solve5.5.2.0 工具**。

作者从语料库中随机地选取 3500 个句子作为情感分类和情绪分类共同的测试样本。然后在剩下的语料中, 作者随机地选取 2000、3500 和 7000 个句子作为情感分类的训练样本, 最后再选取另外的 2000、3500 和 7000 个句子作为情绪分类的训练样本。

(2) 实验结果

名称解释:

Single_sentiment: 只使用最大熵分类器获得情感分类结果 Single_emotion: 只使用最大熵分类器获得情绪分类结果 Joint_sentiment: 基于整数线性规划的情感分类结果 Joint_emotion: 基于整数线性规划的情绪分类结果

Baseline (基准线): 只使用最大熵分类器的分类结果 ILP-中立: 只加入中立情绪限制的整数线性规划方法结果 ILP-正面: 只加入正面情绪限制的整数线性规划方法结果 ILP-负面: 只加入负面情绪限制的整数线性规划方法结果

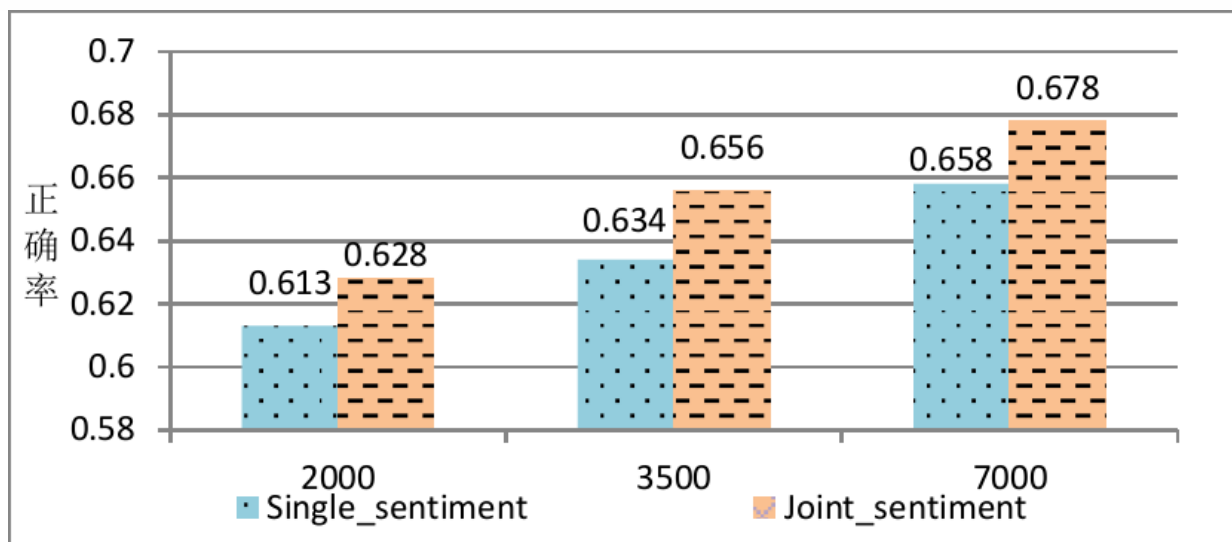


图 3-4 最大熵分类器和基于整数线性规划的情感分类结果

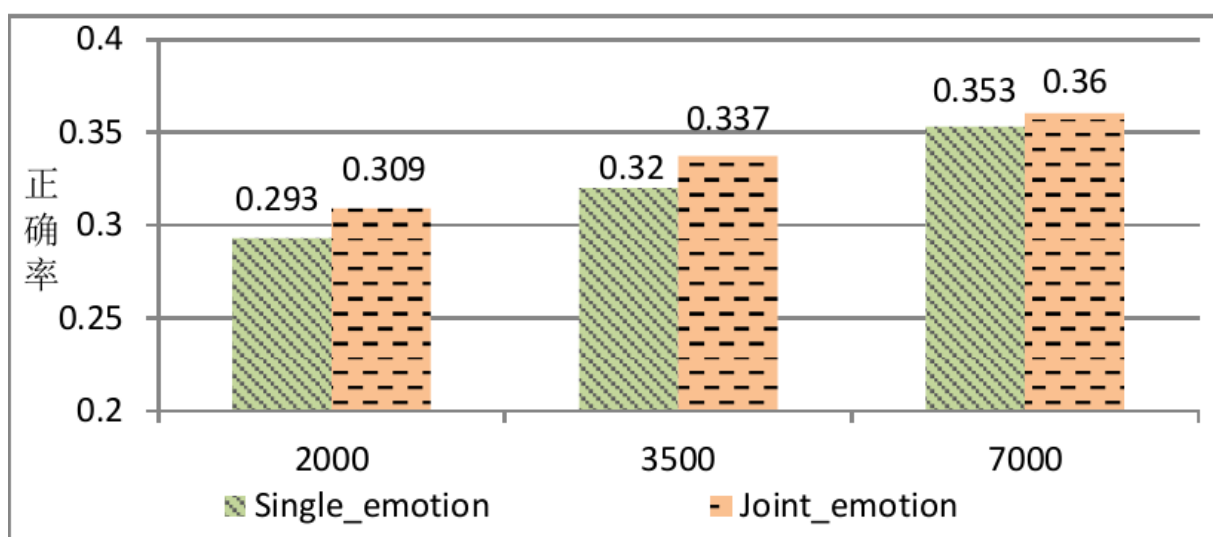


图 3-5 最大熵分类器和基于整数线性规划的情绪分类结果

基于整数线性规划的联合学习方法，在不同标注样本规模下的结果都优于单一的情感分类或者情绪分类。平均的，情感分类的准确率从0.635 上升到0.645，情绪分类的准确率从0.322 上升到0.335。

在联合学习中加入不同限制的分类性能都高于基准线的结果。其中，只加入正面情绪限制和负面情绪限制的分类性能相比基准线结果提高明显，而只加入中立情绪限制的分类性能也有一定程度的提高，但是提高的幅度较小。造成此现象的原因可能是由于样本中，中立的样本比例较小。

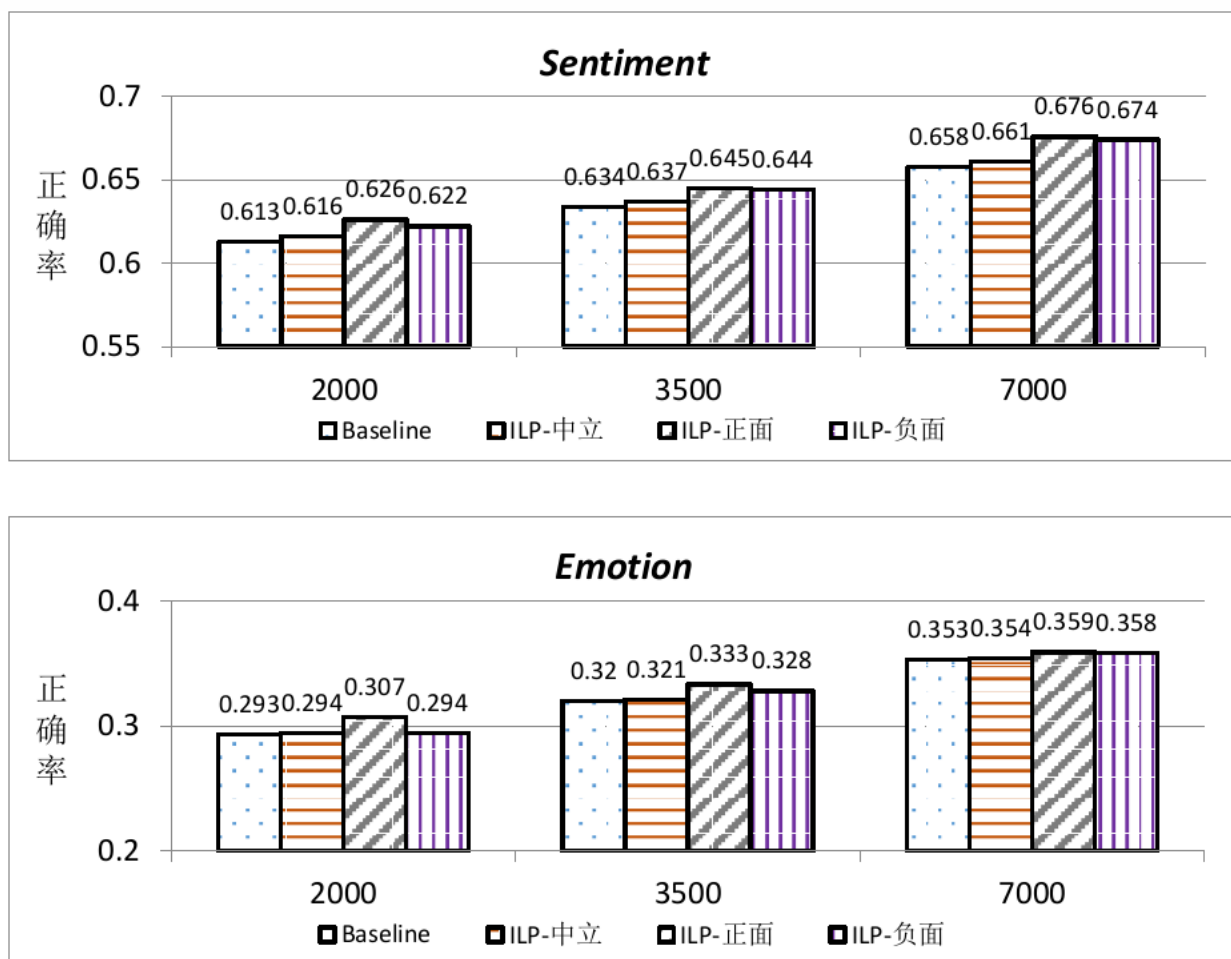


图 3-6 加入不同限制条件的情感分类和情绪分类结果

2 基于标签关系的句子级情绪分类方法

多标签句子级情绪分类

常用的方法是基于标签的独立性假设，将多标签分类问题转化为多个单标签分类问题。因为这种方法忽略了**标签之间的联系**，所以被称为**标签独立的学习方法**，此方法的基本思路是针对不同的标签构建特征空间（简称特征学习）。

目前常用的方法存在一定的缺陷，抛弃了标签之间的联系是不符合实际情况的，不同的情绪之间是相互关联的，所以如何发现并利用多种情绪之间的联系是至关重要的。

论文中给出的方法：（1）根据二元关联方法，针对每个样本构造其对应每个情绪类别的伪样本，（2）然后利用初始分类器对伪样本进行二元分类，（3）最后根据得到的标签关系，设置相应的限制条件，并且将整数线性规划方法作用于初始的分类结果，得到优化的句子级多标签情绪分类结果。

2.1 多标签间的依赖关系

从中文情绪语料库中随机抽取100篇文章，共2751个句子。作者可以计算实例中同时含有标签 l_i 和 l_j 的概率，标记为 $p(l_i, l_j)$ ，其计算公式如下：

$$p(l_i, l_j) = \frac{n_{ij}}{N}$$

实际上，这个概率可能会受到抽取的语料的不同而发生变化，因此应该进行多次的抽样调查，以减小不同情绪分布而带来的影响。文章中也提出了，不同情绪的句子分布并不平衡，不同句子含有标签的个数也有可能有很大差距。

2.2 基于标签关系的句子级情绪分类

基于标签关系的句子级情绪分类方法的基本思想是利用从语料统计中总结的的标签关系以及标签数量范围约束信息来帮助多标签的句子级情绪分类。特别地，作者将标签间的关系转化为相应的限制条件，利用整数线性规划方法作用于初始的句子级情绪分类器输出结果，从而获得优化的句子级多标签情绪分类结果。

在情绪分类器的选择上选择了最大熵分类器，将训练语料放入最大熵模型进行训练，再作用于测试语料，得到初始结果，通过之前得到的标签关系，可以帮助得到ILP的限制条件，得到更加准确的分类结果。

在该方法中，作者还使用了二元关联的方法。作为一种流行的多标签分类方法，它将每一个标签转化为二元分类的问题，得到一组独立的二元分类问题的结果。本文将二元关联作为语料的初始化步骤，作者将每个原始实例转化为K个伪样本，其中K是所用语料中情绪类别的总个数。

对于每个测试样本，定义K个伪样本，通过二元分类确定样本是否包含每种情绪。

采用的中文情绪语料库来共包含8种情绪类

别joy, hate, love, sorrow, anxiety, surprise, anger, expect分别表示为 $\{l_1, l_2, l_3, l_4, l_5, l_6\}$ 。作者定义第*i*个样本为 X_i , $i \in D$ 其中*D*为样本空间大小，则 $X_i^1, X_i^2, X_i^3, X_i^4, X_i^5, X_i^6, X_i^7, X_i^8$ 分别代表对应情绪标签的伪样本。作者将伪样本 $X_i^j (1 \leq j \leq 8)$ 的情绪类别定义为 y_i^j , 且 $y_i^j \in \{0, 1\}$ 。当样本 X_i 含有情绪 l_j 时, $y_i^j = 1$, 否则, $y_i^j = 0$ 。并且定义 $y_i^j = 1$ 时的概率为 $p_{<i,j>}^1$ 。

整数线性规划方法所用到的目标损失函数表示:

$$\text{Min} \sum_{i \in D} \sum_{j \in [1,8]} (c^1 y_i^j + c^0 (1 - y_i^j))$$

其中, $c_{<i,j>}^1 = -\log(p_{<i,j>}^1)$, $c_{<i,j>}^0 = -\log(1 - p_{<i,j>}^1)$ 。

*i*表示伪样本序号, *j*表示对应情绪序号, 所以 y_i^j 为二元分类的结果, $p_{<i,j>}^1$ 为第*i*个样本包含第*j*种情绪的概率, 这个概率由条件概率得到, 通过之前的对情绪的贡献统计得出。由于 $p_{<i,j>}^1 \in [0, 1]$, 取负对数后是一个非负值, 由于式中 y_i^j 与 $1 - y_i^j$ 系数的存在, $c^1 y_i^j + c^0 (1 - y_i^j)$ 得到的结果是第*i*个样本是否对第*j*种包含情绪概率的一种评估, 而且使用负对数的处理也方便了之后的极大似然处理的过程。

之后的极大似然估计中设置的限制条件是通过开始对情绪标签共现的统计分析结果得到的。

(补充限制条件)

2.3 实验

(1) 实验设置

本章所用的语料包含了从中文情绪语料库随机选取的2000句作为测试样本，在剩下的语料中，作者随机的选取2000、4000、6000、8000作为不同规模的训练语料。

在特征选择上，作者采用了每个样本的词袋特征¹，并且将一元词的存在与否转化成二进制向量的形式。

初始的分类算法使用了最大熵分类器的公共工具Mallet Toolkits²。整数线性规划方法使用了lp_solve5.5.2.0工具³。

(2) 实验结果

这一节使用初始情绪分类器作为基准线，对比了基于贝叶斯网络的多标签情绪分类方法和基于标签关系的句子级情绪分类方法的优劣。评价标准使用2.4.2章节提出的标准，即从汉明损失(*Hamming loss*)⁴、正确率(*Accuracy*)⁵、F1值(*F1 - measure*)⁶三个方面进行评价。

论文之后对单个的限制条件对分类效果的影响进行了评估，发现无论是加入哪一种限制条件分类效果都比基准线的结果好，只是对分类的帮助效果不同，其中对一句话中所含的情绪个数信息对分类的结果帮助最大。

3 基于上下文关系的句子级情绪分类

multi-label emotion classification

基于上下文关系的句子级多标签情绪分类方法

为了克服句子级情绪分类中数据稀疏的问题

目的：为了解决句子级情绪分类中的数据稀疏问题，在考虑在特征学习的基础上，加入上下文关系来帮助句子级情绪分类。

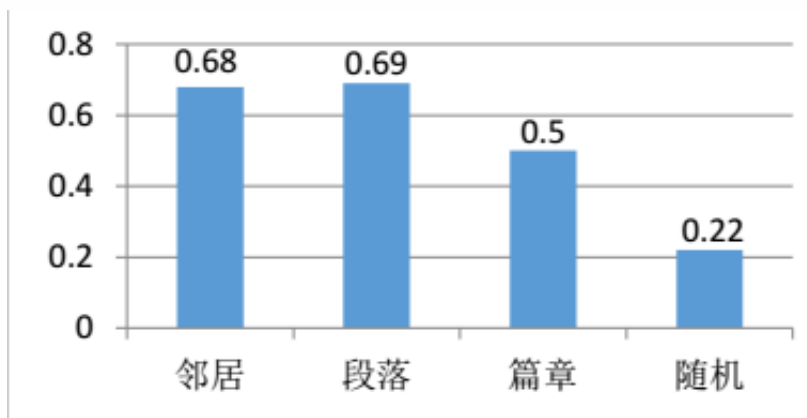
方法：（1）首先针对每一个含有多种情绪标签的样本，构造其在所有情绪标签下的伪样本；（2）然后构建一个伪样本网络用于描述上下文之间的联系。（3）最后构建依存因子图模型进行特征学习和上下文关系的联合学习，从而实现句子级的多标签情绪分类。

3.1 基准系统

采用了目前的最好的基于上下文关系的句子级情绪分类方法[2]作为基准系统。该方法的基本思想是利用训练样本中相邻句子的相似度来进行句子级的情绪分类。首先，通过特征学习来获取训练样本中与每一句最相似的 k 个句子的标签集的统计概率信息。然后，利用相邻句子间的情绪标签转移概率来优化句子的情绪标签结果。

3.2 准备工作

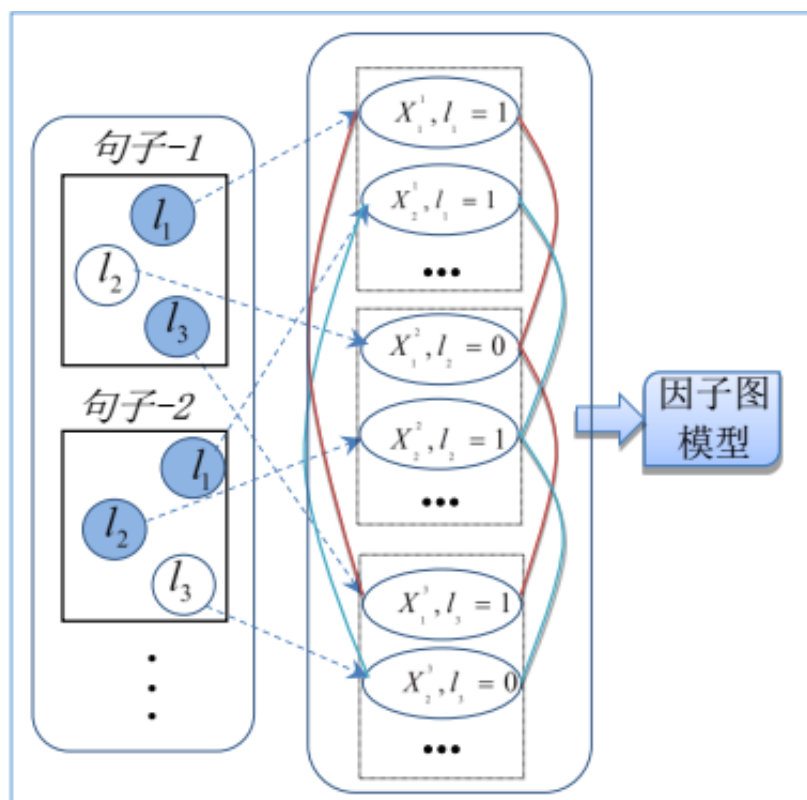
为了更好地说明作者对上下文关系建模的动机，作者系统地调查了所用预料中的上下文依赖关系。



作者通过计算两个实例至少含有一个共同情绪的概率来研究上下文的依赖关系。上图显示了不同情况下，不同的两个实例含有相同情绪标签的概率。其中，“邻居”代表了两个实例为相邻的句子。“段落”代表两个实例来自同一段落。“篇章”代表两个实例来自同一篇文档。“随机”代表两个实例是随机选取的。从图中可以看出，来自同一上下文语境的两个实例含有相同情绪标签的概率远远大于随机的两个实例。

3.3 因子图模型

下面将详细介绍如何利用因子图模型对文本特征和上下文关系进行建模。

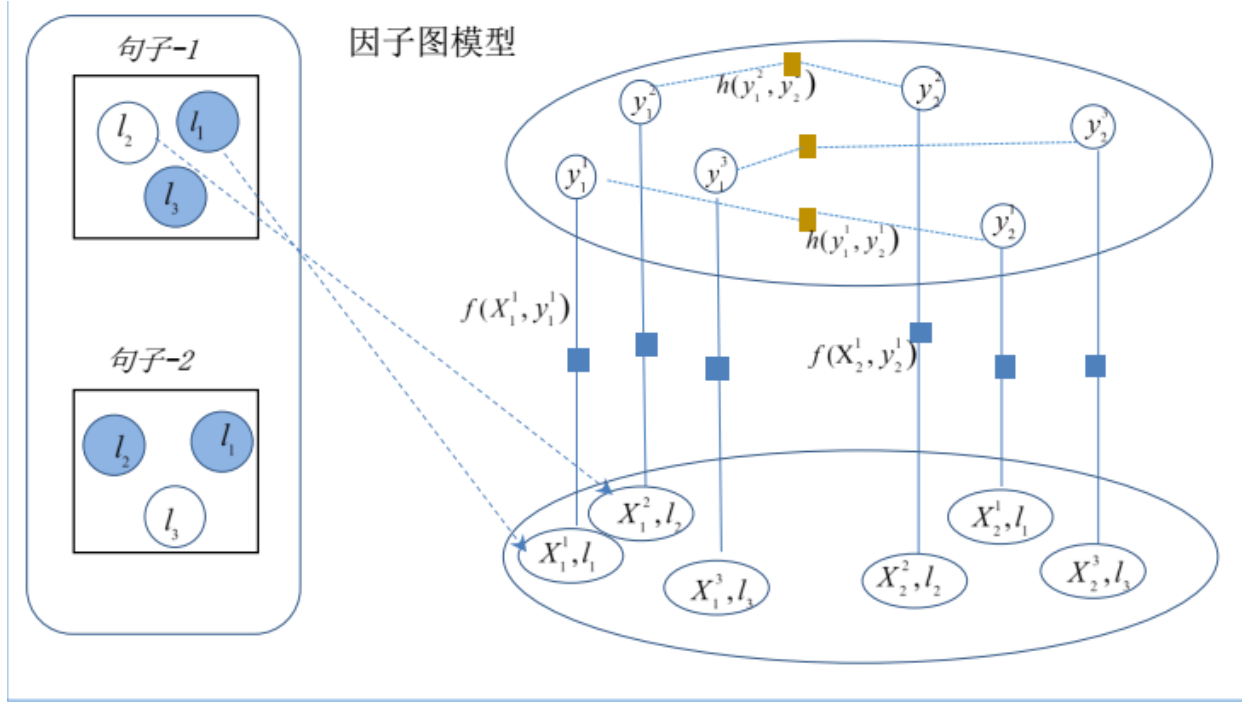


在这个研究中采用**联合转换**的方法。首先假设 K 为一组互斥的情绪标签集合，然后针对每种情绪类别构造样本对应的伪样本。上图给出了本文因子图模型的样本转换方法流程。其中， K 含有三种情绪标签，即 $K=\{l_1, l_2, l_3\}$ 。上图可以分为三个部分：

(1) 左边的部分代表还有不同情绪标签的样本。其中每个方形代表可一个样本，每个蓝色圆圈代表文本还有次情绪，而白色与安全代表文本不含有此情绪。

(2) 中间的部分表示伪样本网络，其中方形代表针对每种情绪类别所构造的伪样本集。椭圆代表每个样本在对应情绪类别下的温杨波，每个伪样本包含文本内容和相应的情绪标签。

(3) 右边部分表示因子图模型。该模型如下图：



作者将 $G=(V,E,X)$ 定义为一个图网络，其中 V 代表一组， E 代表句子间的关系， X 指句子对应文本特征向量。一个因子图有两层节点组成。所有变量的联合分布可以被分解为因子的结果。而该研究的目标就是学习联合分布来预测实例的情绪类别。其中联合分布如下：

$$P(Y|G) = \prod_k \prod_i f(X_i^k, y_i^k) h(y_i^k, H(y_i^k))$$

此公式包含两种因子函数：

(1) 文本特征因子函数。 $f(X_i^k, y_i^k)$ 表示与每个文本 X_i^k 相关的传统文本特征的因子函数。此因子函数的初始化如下：

$$f(X_i^k, y_i^k) = \frac{1}{Z_1} \exp \left(\sum_j \alpha_{kj} \Phi(x_{ij}^k, y_i^k) \right)$$

(2) 上下文关系因子函数。 $h(y_i^k, H(y_i^k))$ 表示实例间的上下文关系，其中 $H(y_i^k)$ 是和 y_i^k 相连接的实例集合。该函数初始化如下：

$$h(y_i^k, H(y_i^k)) = \frac{1}{Z_3} \exp \left\{ \sum_{y_j^k \in H(y_i^k)} \beta_{ijk} (y_i^k - y_j^k)^2 \right\}$$

其中， β_{ijk} 是函数的权重，它代表了两个实例 y_i^k 和 y_j^k 的影响力。

3.4 模型学习和预测

学习因子图模型就是估计最佳的参数配置 $q = (\{a\}, \{b\})$ ，用来最大化对数似然目标函数。即：

$$\theta^* = \arg \max L(\theta)$$

接下来利用梯度下降法来优化目标函数。举例说明，作者可以根据目标函数写出每个 α_{kj} 的梯度：

$$\frac{\partial L(\theta)}{\partial \alpha_{kj}} = E\left[\Phi(x_{ij}, y_i^k)\right] - E_{P_{\alpha_{kj}}(Y|G)}\left[\Phi(x_{ij}, y_i^k)\right]$$

其中 $E[\Phi(x_{ij}, y_i^k)]$ 代表特征函数 $\Phi(x_{ij}, y_i^k)$ 的基于数据分布的期望， $\Phi(x_{ij}, y_i^k)$ 由数据分布给出。下图描述了学习参数 α 的详细算法。利用 LBP 方法来近似的计算边缘分布。注意点，在每次迭代过程中炫耀进行两次 LBP 算法。最后，根据得到的梯度和学习率 η 来更新参数 θ 。其他参数的梯度可以类似的被推到出来。

输入：

样本网 G ，学习率 η

输出：

参数 α 的估计值

初始化 $\alpha \leftarrow 0$

重复：

- 1) 执行 LBP 算法来计算 $E\left[\Phi(x_{ij}, y_i^k)\right]$
- 2) 执行 LBP 来计算 $E_{P_{\alpha_{kj}}(Y|G)}\left[\Phi(x_{ij}, y_i^k)\right]$
- 3) 根据公式 (5.5) 计算梯度 α
- 4) 根据学习率 η 更新参数 α

$$\alpha_{\text{new}} = \alpha_{\text{old}} + \eta \frac{L(\alpha)}{\alpha}$$

直到收敛

有了学习好的参数配置，预测任务就变成了寻找 Y^{U*} 来优化目标函数，即：

$$Y^{U*} = \arg \max P(Y^U | Y^L, G, \theta)$$

其中， Y^{U*} 代表测试样本中实例的情绪标签。在此，作者再一次利用 LBP 算法来计算每一个实例的边界概率以及预测根据最大的边界概率预测实例的标签。上诉预测是一个迭代的过程，知道结果收敛为止。

3.5 基于上下文关系的句子级情绪分类方法的实现

上文介绍了方法中加入上下文关系的动机，以及详细介绍了因子图模型的因子分解过程和模型的学习和预测过程。通过构建每个实例对应于所有情绪标签的伪样本以及由这些伪样本构成的样本网络，本文使用了特征-上下文因子图（FCFG）模型进行特征学习和上下文关系的联合学习，通过对训练样本的学习实现对测试样本的多标签情绪预测。下图描述了方法实现过程。

输入：

已标注的训练样本 L

测试样本 U

输出：

测试样本的情绪分类结果

过程：

1) 样本处理：

对每个样本构造每种情绪类别对应的伪样本；

如果两个伪样本来自同一个原始样本，则在伪样本之间建立连接，构造样本网络；

2) 利用特征-上下文因子图（FCFG）模型来进行特征学习和上下文关系的联合学习，对测试样本进行情绪标签预测，得到最终的分类结果。

4 参考文献

[1] 汪蓉.句子级情绪分类方法研究[D].苏州：苏州大学计算机科学与技术学院, 2016：36.

[2]Xu J., R. Xu, Q. Lu and X. Wang. Coarse-to-fine Sentence-level Emotion Classification based on the Intra-sentence Features and Sentential Context [C]. In Proceeding of CIKM-12, 2012, 2455-2458.

1. 词袋模型假设一个文本或文档可以看做是一袋子单词，不考虑其语法和词序关系，每个词都是独立的。使用向量表示文本，词序信息已经丢失，每个文档看做一系列不相关的词的集合。[↩](#)

2. <http://mallet.cs.umass.edu/>[↩](#)

3. http://web.mit.edu/lpsolve_v5520/ (forbidden) http://web.mit.edu/lpsolve_v5520/java/docs/api/lpsolve/LpSolve.html[↩](#)

4. 针对所有待测实例，衡量样本预测标签与样本实际标签之间的不一致程度，即样本具有某标签但未被识别出来或者不具有某标签却被误判的可能性。[↩](#)

5. 针对所有待测实例，衡量样本的预测标签和样本实际标签之间的平均相似程度。[↩](#)

6. F1 值用于综合评价系统的性能，即准确率和召回率的调和平均值。[↩](#)