

条件随机场读书报告

2017 年 4 月 27 日

这篇报告基于两篇文章，一篇是 Lafferty 的 Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, 是 CRF 的开山之作，主要内容是 CRF 相较于 HMM 和 MEMM 的优点及原理，第二篇是中文信息学报上的“基于层叠 CRFs 模型的句子褒贬度分析研究”，主要是 CRF 的应用。

第一篇如下：在条件随机场之前，隐马尔科夫存在两个强假设，1 当前状态只依赖前一个状态，2 输出观察值之间严格独立。最大熵马尔科夫模型结合最大熵和 HMM 是生成模型，不要求输出观察值强独立，

$$p(Y_{1:n}|X_{1:n}) = \prod_{i=1}^n p(Y_i|Y_{i-1}, X_{1:n}) = \prod_{i=1}^n \frac{\exp(w^T f(Y_i, Y_{i-1}, X_{1:n}))}{Z(Y_{i-1}, X_{1:n})}$$

其中 Z 为归一化因子，最大熵马尔科夫不再要求输出观察值之间严格独立，但是从公式可以得到是在每个节点在前一个状态 Y_{i-1} 确定的前提下，考虑当前所有可能的状态，做局部归一化。当前状态仅依赖前一状态使得最大熵 MM 存在标注偏置问题，当各节点分支数不同时，在前一个节点分支数少时，得到的当前节点归一化概率大，概率分布不均衡，状态转移不公平。基于以上缺点，提出 CRF

X 指观察序列， Y 指标注序列， $G = (V, E)$ ，当满足

$$p(Y_v|X, Y_w, w \neq v) = p(Y_v|X, Y_w, w = v)$$

即 Y_v 只依赖与 Y_v 相连的边，则认为是条件随机场，文章中介绍了线性链条件随机场，即 $P(Y_i|X, Y_1, Y_2, \dots, Y_n) = P(Y_i|X, Y_{i-1}, Y_{i+1})$,

只和与其相连的边有关. 当观察序列取值为 x , Y 取值为 y 的概率为:

$$p(y|x) \propto \exp\left(\sum_{e \in E, k} \lambda_k f_k(e, y|_e, x) + \sum_{v \in V, k} \mu_k g_k(v, y|_v, x)\right)$$

$y|_S$ 表示 S 中与 y 相连的边, 其中 f_k, g_k 是特征函数, 分别表示转移特征和状态特征, 是给定的也是固定的, λ, μ 是权重

对观测序列 x 的每一个位置 $i = 1, 2, \dots, n+1$, 定义一个 m 阶矩阵, 其中 m 为所有可能的状态数, M_i 表示第 i 个时间的概率 (没有归一化之前的),

$$M_i(y_i - 1, y_i|x) = \exp(W_i(y_i - 1, y_i|x))$$

$$W_i(y_i - 1, y_i|x) = \sum_k \lambda_k f_k(e_i, Y|_{e_i} = (y', y), x) + \sum_k \mu_k g_k(v_i, Y|_{v_i} = y, x)$$

CRF 没有在像 HMM 一样建模联合概率, 所以不需要遍历所有可能的观察序列 x , 只需要在给定 x 和参数下计算条件概率。

给定观察序列 x , 标记序列 y 的非归一化概率可以通过 $n+1$ 个矩阵的乘积表示, $\prod_{i=1}^{n+1} M_i(y_{i-1}, y_i|x)$, 这个地方 $M_i(y_{i-1}, y_i|x)$ 不能理解为在给定 x 下 y_{i-1}, y_i 的概率, 而是在给定 x 下, 由 y_{i-1} 转移且表现状态为 y_i 的概率。归一化因子指以 start 为起点, stop 为终点通过的所有可能状态路径的非归一化概率和, 表示成下面的形式:

$$Z_w(x) = (M_1(x)M_2(x)\dots M_{n+1}(x))_{start, stop}$$

类比最大熵马尔科夫, CRF 可以写成下面的形式:

$$p(y_{1:n}|x_{1:n}) = \frac{1}{Z_w(x_{1:n})} \sum_{i=1}^n \exp(w^T f(y_i, y_{i-1}, x_{1:n}))$$

相对于 MEMM, CRF 是在全局归一化, 给定需要标记的观察序列的条件下, 计算整个标记序列的联合概率分布, 而不是在给定当前状态条件下, 计算下一个状态的状态分布。CRF 相对于前两个效

果更优，参数也更多，使得计算更复杂，利用迭代尺度法，通过极大化训练数据的对数似然函数求模型参数。

第二篇讲 CRF 在句子褒贬度分析的应用，一个句子表达出的情感又可以分为强烈贬抑、一般贬抑、客观、一般褒扬、强烈褒扬五个类别。通常把句子褒贬度分析看作分类任务，将每个句子看作独立样本，利用词典等做褒贬度打分，但是语言上下文具有连贯性，一个段落中的句子可以看作一个序列，序列间关系紧密，因此这不是独立句子分类任务，而是序列标注任务。在类别的独立性方面，孤立的看成 5 分类任务也不合理，强烈褒扬和一般褒扬具有相同的倾向性，褒贬较容易判别，但是强度不好分别，因此可以看成层级任务，第一层用以标注褒扬和贬抑，第二次用于区分强度。文章在层叠 CRFs 框架下，利用多个 CRFs 模型从粗到细分步地计算句子的褒贬度以及褒贬强度。

标记	人工答案	模型结果	匹配标记	Precision %	Recall %	F1 值 %
强烈褒扬	51	74	33	44.59	64.71	52.80
一般褒扬	166	226	74	32.74	44.58	37.76
客观	1 190	1 101	971	88.19	81.60	84.77
一般贬抑	105	102	29	28.43	27.62	28.02
强烈贬抑	19	28	10	35.71	52.63	42.55
合计	1 531	1 531	1 031	72.96	72.96	72.96

图 1: 层叠 CRF 褒贬度分析结果对比

上图一为文章的实验结果。文章选择了三类特征：倾向性表示特征、主题词特征、转移词特征。对于观察 X 上的点特征 $f_i(y_i = y_t, X)$ ，选择表示倾向性的单词、主题词作为特征。表示倾向性的词包括形容词、副词等以及这些词的先验性褒贬度打分。如果在形容词所在位置的 3 个窗口内出现 {not ,no ,can't ,merely ,never} 等表示否定的词语，对词语的先验性打分乘以 -1。主题词包括从训练语料中统计高频的名词，对于影评英文文章，主题词包括 {film ,story ,life ,movie ,scene} 等词语，这些表示电影属性的关键词对于句子褒贬度分析有着重要的作用。文中考虑上一句的标记与两类特征。

一是表示转折或者递进等关系的连词和副词，二是句子在篇章的位置。

对于第一层模型和第二层模型使用同样的特征，只是对于第二层褒贬强度分类任务中没有考虑句子位置特征。第一层中，忽略褒贬强度的影响，只考虑句子褒贬度的判别，把段落中的句子分为褒扬、客观、贬抑三类。在第二层，把第一层的结果作为第二层的输入，对于第一层中标为客观的句子不输入第二层，对于标为褒扬或贬抑而且相邻的句子作为独立的样本，分别利用两个 CRFs 模型确定每个句子的强度；后一层的结果不对前一层的结果产生影响，两层分别寻优。