

面向医学文本的 PICOT 多标签分类系统文献综述

奉捷 张睦婕 周嘉莹 梅磊

指导老师 张祥

摘要：医学文献中蕴含着海量的人类医学知识结晶。通常，一篇医学文献（与临床相关的医学文献）中论述的医学问题可以表达为五个要素：分别为 P（医学实验设计的人群和医学问题）、I（医学实验中用到的治疗方案或药物）、C（与其他治疗方案或药物进行的对比）、O（医学实验的结论）、T（时间）。医学文献中的大部分内容将围绕一个医学问题的 PICOT 这五个元素展开。

本项目尝试通过改进的多标签分类算法 ML-KNN 和多标记学习的反向传播神经网络 BP-MLL 设计并实现一个多标签分类系统，将医学文献中的句子（或段落、篇章等不同粒度的文本）进行多标签分类，判断句子属于 PICOT 中的哪个或哪几个要素。本项目主要使用监督学习算法，通过人工构建的训练集来学习分类模型。同时尝试使用无监督、半监督或远监督方法来学习分类模型。

关键词：医学文献 PICOT 多标签分类 ML-KNN BP-MLL

1.引言：医学文献中蕴含着海量的人类医学知识结晶。现代循证医学（EBM）要求医疗决策应在现有的最好的临床研究依据基础上作出，这使得医学文献中的论点常常基于对大量临床数据的分析和检验。

通常，一篇医学文献（与临床相关的医学文献）中论述的医学问题可以表达为五个要素：分别为 P（医学实验设计的人群和医学问题）、I（医学实验中用到的治疗方案或药物）、C（与其他治疗方案或药物进行的对比）、O（医学实验的结论）、T（时间）。通常，一篇医学文献中的大部分内容将围绕一个医学问题的 PICOT 这五个元素展开。

医学文献中的信息抽取、知识图谱构建目前是人工智能与生命科学交叉学科中受到广泛关注的领域。医学文献中的 PICOT 句子分类目前还未有完全成熟的工作。而本课题从这方面入手，涉及机器学习中的多标签分类、半监督或远监督方法等等。

2. 医学文本的 PICOT 多标签分类系统

2.1 PICOT 模型理解

PICOT 是基于循证医学（EBM）理论的一种将信息格式化的检索方式。为 participants, interventions, comparisons, outcomes, time 的缩写。

PICO 将每个问题分为四个部分：participants（对象）、interventions（干预）、comparisons（对照）、outcomes（预后），time（时间）。

举例：

文献 [Buchhorn 2011](#) 的摘要中给出了简单的标注。

Abstract

problem

intervention

outcome

outcome

Aim: Infants with congenital heart disease and left-to-right shunts may develop significant clinical symptoms of congestive heart failure in spite of therapy with digoxin and diuretics. We investigated the effects of β -blockade in infants with severe heart failure. *Methods and results:* We performed a prospective, randomized, open monocenter trial in infants treated with digoxin and diuretics ($n=10$) in comparison to 10 infants receiving additional β -blocker therapy. After 17 days on average β -blocker treated infants (propranolol 1.6 mg/kg/day) improved significantly with respect to Ross heart failure score (3.3 ± 2.3 vs. 8.3 ± 1.9 , $P=0.002$), lower renin levels (338 ± 236 vs. 704 ± 490 $\mu\text{U/l}$, $P=0.008$) and lower mean heart rates in Holter ECG (118 ± 10 vs. 142 ± 11 beats/min, $P<0.001$). While digoxin and diuretic treated infants had unchanged mean heart rate (149 ± 8 vs. 148 ± 10 beats/min), less decrease of symptoms (Ross Score: 8.5 ± 1.7 vs. 6.8 ± 2.3 , $P=0.02$) but a significant increase of renin levels (139 ± 102 vs. 938 ± 607 $\mu\text{U/l}$, $P=0.001$). *Conclusion:* Additional propranolol treatment but not digoxin and diuretics alone can effectively reduce clinical symptoms of heart failure in infants with congenital heart disease, who suffer from increased neurohormonal activation. © 2001 Elsevier Science Ireland Ltd. All rights reserved.

其中，每个标签还具有相应细粒标识模板，在此处可分别指对象，干预（治疗手段、检查方法），对照（与其他治疗或检查的区别），预后（期望达到的目的）。

对于 PICO 的研究已较为成熟，形成了较为完善的标签定义模板：

细粒度 PICO 模板：

Problem

1. [Disease or Syndrome] 【pr.Disease】“panic disorder” 文献研究主题中的疾病
2. [Behavior :Social Behavior ; Individual Behavior] 【pr.Behavior】 患者主观感受
3. [Sign or Symptom] 【pr.SS】 文献研究主题中病症的表现，医生描述的症状

Population

4. [AGE][Age Group][quantitative concept] 【po.Age】 研究对象的年龄
5. [sample size] 【po.size】 研究样本大小
6. [GENDER][organism attribute] 【po.Gender】 , e.g., “male”性别
7. [TREATMENT STATUS] 【po.Treatment】 , e.g., “delayed treatment” 治疗状态
8. [PHYSICAL CONDITION] 【po.PhyConditon】 , e.g., “healthy”[Qualitative Concept] 健康状态
9. [MEDICAL HISTORY] 【po.MedHistory】 , e.g., “with prior attacks” 用药历史，药物过敏史
10. [Therapeutic or Preventive Procedure & Pharmacologic Substance] 【po.Procedure】 , e.g., “taking hormone replacement therapy”接受过的治疗方式 服用的药物
11. [Disease or Syndrome] 【po.Disease】 , e.g., “nonvalvular atrial fibrillation” 临床试验中患者的疾病，包括不能患有什么病
12. [Sign or Symptom] 【po.SS】 , e.g., “chronic cough”患者的病症
13. [Clinical Attribute] 【po.Clinical】 (Diastolic blood pressure,Systolic blood pressure,Heart rate)包括血压，心率等等

Intervention & Comparison

14. [Therapeutic or Preventive Procedure & Pharmacologic Substance][Hazardous or Poisonous Substance,Organic Chemical] 【i.Procedure】 , e.g., “warfarin（华法林）” 研究的治疗方法/对照组的治療方法 和使用的药物
15. [Diagnostic Procedure] 【i.Diagnostic】 , e.g., “ultrasound（超声波）” 诊断过程
16. [DIAGNOSTIC TEST] 【i.DiagnosticTest】 , e.g., “Pap smear 巴氏涂片” 诊断测试的对象
17. [EXPOSURE] 【i.Exposure】 , e.g., “maternal smoking” 暴露或接触，例如暴露在低浓度一氧化碳环境、接触细菌 Phenomenon or Process
18. [Disease or Syndrome] 【i.Disease】 , e.g., “a flare-up of the Crohn’s” 对照组患者患有的疾病
19. [Sign or Symptom][finding] 【i.SS】 , e.g., “a very low serum iron”临床实验后的症状

Outcome

20. [TREATMENT OUTCOME] 【o.Treatment】 , e.g., “fibroid volume reduction 肿瘤数量减少” 治疗结果，从病症/医学指标角度
21. [PATIENT OUTCOME] 【o.Patient】 , e.g., “decreased mortality 死亡率降低” 从患者角度或群体角度

备注：[] 中是属性名称，除全部大写的属性外都是 UMLS 中的语义类型，【】 中是属性的缩写

然而，该模型对于 T 标签的定义尚不明确，对于 T 标签相关研究较为稀少。我们参照 PICO 的定义方法，给出 T 标签的粗细粒度定义，用于 T 标签的分类依据：

Time

22. [Sign or Symptom] 【t.SS】患者的出现病症时间（与 P 相关）/临床试验后出现症状的时间（与 I 相关）
23. [MEDICAL HISTORY] 【t.MedHistory】用药时间（与 P 相关）
24. [Therapeutic or Preventive Procedure & Pharmacologic Substance] 【t.Procedure】实验持续的时间（与 I 相关）
25. [TREATMENT OUTCOME] 【t.Treatment】治疗后获得效果所用时间或者治疗效果维持的时间（与 O 相关）

2.2 多标签分类

2.2.1 多标签分类问题的定义

多标签分类问题很常见，比如一部电影可以同时被分为动作片和犯罪片，一则新闻可以同时

属于政治和法律，还有生物学中的基因功能预测问题，场景识别问题，疾病诊断等。

用 X 代表样本空间， $L=\{\lambda_1, \lambda_2, \lambda_3 \cdots \lambda_m\}$ 为有限标签集合，我们假设 X 中的样本实例 $x \in X$ 和 L 的一个子集 $I \in 2^L$ 相关，这个子集称作相关标签集。同时补集被认为与 x 不相关。相关标签集 L 用向量 $y = (y_1, y_2, y_3 \cdots y_m)$ 标识，其中 $y_i = 1 \Leftrightarrow \lambda_i \in L$ 。用 $y = \{0, 1\}^m$ 表示可能的标签集。一个多标签分类器 h 是一个映射 $X \rightarrow Y$ ，对每一个实例 $x \in X$ 分配一个标签子集。因此分类器 h 的输出是一个向量 $h(x) = (h_1(x), h_2(x) \cdots h_m(x))$ 。

本项目试图设计并实现一个多标签分类系统，能够将医学文献中的句子（或段落、篇章等不同粒度的文本）进行多标签分类，判断句子属于 PICOT 中的哪个或哪几个要素。本项目主要使用监督学习算法，通过人工构建的训练集来学习分类模型。同时尝试使用无监督、半监督或远监督方法来学习分类模型，可以充分利用医学中的系统综述（systematic review）来自动生成训练集。

2.2.2 多标签分类的方法

方法基本上分为两种，一种是将问题转化为传统的分类问题，二是调整现有的算法来适应多标签的分类

将问题转化为传统分类：不扩展基础分类器的本来算法，只通过转换原始问题来解决多标签问题。比如对每个实例确定或随机的分配一个标签，或只保留只有一个标签的训练样本，然后用传统的多分类方法去训练。如 BR，LP 等。转化方法往往存在训练样本的特征损失，或未考虑多个标签之间的关联，分类效率低下等问题。

调整现有算法适应分类：比如通过调整 boost kNN SVM 等，改编算法来直接执行多标签分类，而不是将问题转化为不同的问题子集。例如，kNN 的多标签版本是由 MLkNN 表示的。这些方法考虑到了不同标签之间的依赖。

2.3 医学文本多标签分类

2.3.1 医学文本的多标签特点

医学文献是文本的一个特殊领域，涉及到诸多专有名词，特别是专有短语。在分析文本时，必须根据医学领域知识进行单词的边界划分以及词义的消歧，以使之适用于医学文本的分类。

医学文本的标签属于无层次标签，标签之间无明显的层次关系。同时医学文本标签数量较并且固定。并且标签之间具有相关性。

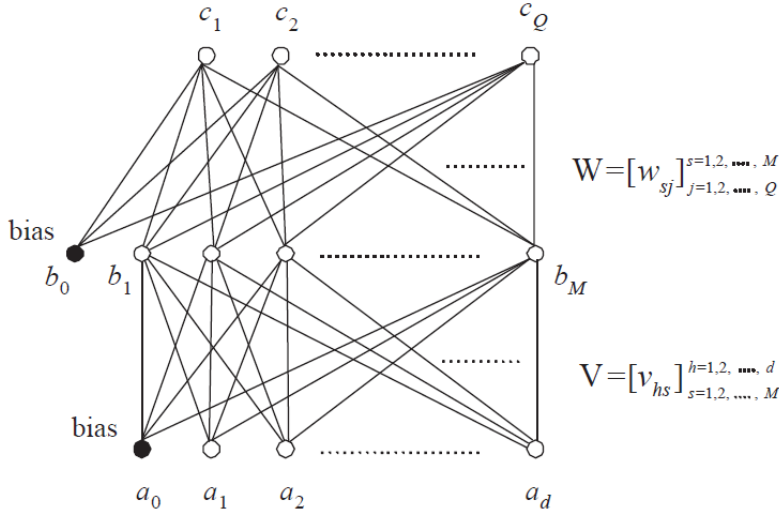
2.3.2 医学文本多标签分类

针对医学文本的标签特点，宜利用 BERT 进行文本的预训练。同时借助 UMLS 中收录的医学概念和医学词汇建立字典，将文本中的词语与 UMLS 中的词语进行匹配，以便后期分类。分类算法宜使用调整算法直接执行多标签分类的方法。我们暂定使用 ML-KNN、BP-ML 和 Bi-LSTM 方法，我们比较了前种方法的优缺点，并且通过实现来验证它，并尝试采用 Bi-LSTM 进行更好的模型构建。

MLKNN是由传统的K近邻 (K-nearest neighbor, KNN) 发展而来的，对于每一个测试样本，在训练集中找到它的K近邻。然后，基于邻居样本的统计信息，如属于相同类别的邻居的个数，用最大后验概率原则 (MAP) 决定测试样本的标签集合。它的优点是：基于实例的学习，思路较为清晰和简单；该方法所构建的模型是局部模型；目标变量是离散的或实值的。缺点是：该方法所构建的模型是局部模型，预测结果必须通过实例的引入来得出，是一个懒惰的学习方法；同时由于是基于实例的学习，计算量偏大。

BP-MLL的神经网络算法，即多标记学习的反向传播，这是第一个多标签神经网络算法。顾名思义，BP-MLL源自流行的反向传播算法，通过用定义的新误差函数替换其错误函数来捕获多标记学习的特征，也就是说，属于实例的标签的排名应高于不属于该实例的标签。传统的多层前馈神经网络适用于从多标签示例中学习。前馈网络将神经元排列成层，第一层采用输入，最后一层采用输出。中间层与外部世界没有联系，称为隐藏层。通过反向传播，从后向前调整每一层神经网络的参数权值，最终获得能够准确预测标签的全局模型。该方法用于多标签分类的核心是设计一些特定的误差函数，而不是简单的比较方差，以捕捉多标记学习的特征。其次必须考虑多个标签之间的关联性，利用此重要关系优化算法，以减少计算量，提高分类效率。与此同时，需要设计特定的评估标准，以检测多标签分类的准确性，进一步提升算法的学习能力。

2.3.3 BP-MLL 算法对于文本的分类



$c_j^i = c_j(x_i)$ 是神经网络的输出第 j 个标签的实际输出, d_j^i 是神经网络的输出第 j 个标签的期望输出, $d_j^i = +1$ 表示 j 属于 Y_i , $d_j^i = -1$ 表示 j 不属于 Y_i .

对于一个训练集，误差函数为：

$$E = \sum_{i=1}^m E_i$$

$$E_i = \sum_{j=1}^Q (c_j^i - d_j^i)^2$$

其中

$c_j^i = c_j(x_i)$ 是神经网络的输出第 j 个标签的实际输出, d_j^i 是神经网络的输出第 j 个标签的期望输出, $d_j^i = +1$ 表示 j 属于 Y_i , $d_j^i = -1$ 表示 j 不属于 Y_i .

若考虑标签之间的相关性, 则应对 E 加以修正:

$$E = \sum_{i=1}^m E_i = \sum_{i=1}^m \frac{1}{|Y_i| |\bar{Y}_i|} \sum_{(k,l) \in Y_i \times \bar{Y}_i} \exp(-(c_k^i - c_l^i))$$

其中 \bar{Y}_i 是 Y_i 的补集, $|\cdot|$ 表示集合中的元素数量。 c_k 表示标签属于 x_i , c_l 表示标签不属于 x_i

E 的值越大, 表示标签属于 x_i 的可能性越大, 反之标签不属于 x_i 的可能性越大。

然后采用梯度下降法减小 E 。

3.总结

参考文献

- [1] [Riva J J , Malik K M P , Burnie S J , et al. What is your research question? An introduction to the PICOT format for clinicians\[J\]. Journal of the Canadian Chiropractic Association, 2012, 56\(3\):167-171.](#)
- [2] [Building Evidence Graph for Clinical Decision Support Evidence graph](#)
- [3] [PICO 原则--循证医学](#)
- [4] [A Unified Architecture for Semantic Role Labeling and Relation Classification](#)
- [5] [An introduction to the PICOT format for clinicians](#)
- [6] [Evaluation of PICO as a Knowledge Representation for Clinical Questions](#)
- [7] [Evolving PICOT for the Digital Age PICOT-D](#)
- [8] [Multi-Label Neural Networks with Applications to Functional Genomics and Text Categorization](#)
- [9] [A Review on Multi-Label Learning Algorithms](#)
- [10] [Explainable Prediction of Medical Codes from Clinical Text James Mullenbach, Sarah Wiegrefe, Jon Duke, Jimeng Sun, Jacob Eisenstein](#)