**临床领域的临时信息标注**

**William F. Styler IV1, Steven Bethard2, Sean Finan3, Martha Palmer1,**

**Sameer Pradhan3, Piet C de Groen4, Brad Erickson4, Timothy Miller3,**

**Chen Lin3, Guergana Savova3 and James Pustejovsky5**

**1 Department of Linguistics, University of Colorado at Boulder**

**2 Department of Computer and Information Sciences, University of Alabama at Birmingham**

**3 Children’s Hospital Boston Informatics Program and Harvard Medical School**

**4 Mayo Clinic College of Medicine, Mayo Clinic, Rochester, MN**

**5 Department of Computer Science, Brandeis University**

**摘要:**本文讨论临床叙述中临时信息标注的正式规范的要求。 我们讨论ISO-TimeML的实现和扩展，用于标注临床备注语料库（称为THYME语料库）。 为了反映信息任务和领域内基于推理的推理需求，制定了一个新的标注指南，即“THYME ISO-TimeML指南（THYME-TimeML）”。 为了澄清哪些关系值得标注，我们区分了文本中语言派生和推论派生的时间顺序。 我们还针对这一新资源应用了性能最佳的Temp-Eval 2013系统来衡量系统适应临床领域的难度。 该语料库可供社区使用，并已提议用于2015年下半年任务。

# 1 引言

生物医学界对时间推理有着长期兴趣（Savova等，2009; Hripcsak等，2009; Meystre等，2008; Bramsen等，2006; Combi等，1997; Keravnou，1997; Dolin，1995; Irvine等，2008; Sullivan等，2008）。 这种兴趣延伸到自动提取和解释来自医疗文本的时间信息，例如电子出院摘要和患者病例摘要。 有效利用这些叙述中的时间信息是医学研究人员对信息学智能分析的关键一步，而对于许多数据挖掘任务而言，对文本中的时间信息（包括隐式和显式两种）的了解也是必不可少的。

目前已证明临床叙述中的时间信息可以有效地挖掘，以提供一些更高层次的时间推理信息（Zhao et al。，2005）。 然而，由于确定事件之间的时间关系的复杂性，时间表达的多样性以及与更广泛的计算语言学问题的相互作用，所以对这种叙述的强大的时间理解很难实现。

最近关于电子健康记录（EHR）的工作指出了利用和挖掘其中所含信息的新方法（Savova等，2009; Roberts等，2009; Zheng等，2011; Turchin等，2009） 。我们着重于提取数据的两个主要用例。首先，我们希望能够在访问时向医生提供交互式显示和患者记录摘要，对患者的病史进行全面审查，既快速又不易发生疏漏。其次，我们希望能够跨越医疗记录的大型数据库进行时间感知的二级研究（例如，“接受手术X的患者中有多少百分比在Z个月内发展副作用？”）。这两种应用都需要提取关键事件的时间和日期关联以及病人护理期间事件的相对顺序，这些都来自组成病人EHR的各种记录。尽管我们考虑了这两个特定的应用，但我们开发的模式是可扩展的，可能会嵌入各种各样的生物医学用例中。

电子病历中的叙述文本文中有着丰富的时间表达，其中经常包含关于医疗事件时间的断言，如访问，实验室值，症状，体征，诊断和程序（Bramsen等，2006; Hripcsak等，2009; Zhou 等人，2008）。 由于以下原因，病历中的时间表示和推理是困难的：（1）时间表达的多样性; （2）确定事件之间的时间关系的复杂性（通常是推断）; （3）处理事件时间粒度的困难; 和（4）自然语言处理中的一般问题（如歧义，回指，省略，连词）。 因此，用于重构时间轴的信号既可以是特定领域的也可以是复杂的，并且通常是隐含的，需要重要的领域知识才能准确检测和解释。

在本文中，我们讨论在临床笔记中准确注释这些时间信息的要求。 我们描述了专门为临床领域开发的ISO-TimeML（Pustejovsky等人，2010）的实现和扩展，我们称之为“ISO-TimeML的THYME指南”（“THYME-TimeML”），其中THYME 代表“您的医疗事件的时间历史（Temporal Histories of Your Medical Events）”。 这些指南的简化版本形成了2012年i2b2医学领域时间关系挑战的基础（Sun等人，2013a）。

这是在THYME项目的背景下开发的，其目标是为临床笔记中的语义信息创建稳健的黄金标准，并开发最先进的算法来训练和测试该数据集。

从新闻文本中提取时间线需要具体实现与时间间隔，排序和组织有关的上下文相关假设，这些假设是文本中标记的显式信号的基础（Pustejovsky and Stubbs，2011）。 从临床笔记中提取患者病史的时间表也涉及这些类型的假设，但是临床叙述的特点带来了特殊的要求。 由于医学速记练习和一般领域知识，许多事件-事件关系根本没有在文本中标明，并且依赖于医疗过程进展的共同理解和共同概念模型，仅供熟悉语言使用的医学界读者阅读。

识别这些隐含的关系和时间属性给注解过程带来沉重的负担。因此，在THYME-TimeML指南中，在只能通过特定领域的时间知识才能推断出的描述和禁止的时间顺序的注释上花费了很大的努力。

尽管为了使得注释更加便利，THYME指南中有些描述偏离了ISO-TimeML标准，但本文将着重讨论由临床领域需求特别激发的差异，以及为提取时态数据而构建系统在临床和一般领域的意义。

# 2 临床文件的类型

在THYME语料库中，我们一直在检查来自大型医疗保健实践（Mayo Clinic）的1,254个未识别1记录，这些记录代表肿瘤学中两个不同的领域：脑癌和结肠癌。 迄今为止，我们主要在我们的EHR中检查了两种不同类型的临床叙述：临床说明和病理报告。

临床记录是医生与病人交互的记录，通常包括多个明确划分的部分，详细描述病人护理和现患病的不同方面。 这些笔记在各个机构和专业中都是相当通用的，虽然某些术语和推论可能特定于特定类型的实践（如肿瘤学），但它们具有统一的结构和模式。 例如，“现患病史”总结了患者主诉的过程以及迄今为止尝试的干预和诊断。 在其他部分中，医生可以概述她目前的病人治疗计划，然后描述病人的具体病史，过敏症，护理指示等等。

对于时间推理最关键的是，每个临床记录都反映了患者治疗历史中的所有医生的陈述都准确记录的确切时间（DOCTIME），并且每个部分倾向于描述特定时间段的事件。 例如，“现代病史”主要描述在DOCTIME之前发生的事件，而“药物”在DOCTIME中提供快照，而“持续护理订单”则讨论尚未发生的事件.2

临床笔记包含丰富的时间信息和背景，从以前的治疗和症状中流畅地移动，为未来的干预提供条件。 他们通常也富有假设性陈述（“如果肿瘤复发，我们可以......”），每一种都可以形成自己独立的时间表。

相比之下，病理笔记则完全不同。 这些注释由医学病理学家在接收和分析样本（从活组织检查的组织样本到肿瘤或器官的切除部分）范围内生成。 病理笔记为患者的医生提供了至关重要的信息，用于确认样本中的恶性肿瘤（癌症），描述手术边缘（指示肿瘤是否完全切除）以及分类和“分期”肿瘤，描述癌症的严重程度和扩散情况。 由于这些注释中的信息属于在单个时刻采集的样本，因此它们在时间上很稀疏，很少涉及样本检查之前或之后的事件。 但是，它们包含关于患者疾病状态和癌症本身的重要信息，并且必须解释为了解患者疾病的历史。

最重要的是，在所有的电子健康档案中，我们都必须与现代医疗记录中的根本性紧张结果相抗衡：超详细的记录为防范医疗事故诉讼提供了重要的防御措施，但记录这些细节需要很长时间，因而医生很少这样做。 鉴于这些笔记是由医学专业人员撰写的（他们构成了一个相对孤立的言语社区），使用了许多非标准的表达，缩写和共享知识的假设，这些知识同时对于有相似背景的其他人来说是简洁和细节丰富的。

这些节省时间的标记的范围可以从暂时使用的缩略词（例如，'qid'，拉丁文中的四分之一，'每日四次'）到假定的排序（在采取治疗措施之前即假定一种疾病的诊断测试），甚至是完全隐含的事件和时间细节。 例如，考虑（1）中的句子。

（1）结肠镜检查3/12/10，结节活检阴性 我们必须了解，在结肠镜检查过程中，医生获得了结节的活检标本，将其打包并送至病理科医师处，对其进行检查并确定其为“阴性”（非癌症）。

在这些文件中，我们必须尽可能多地恢复时间细节，尽管它可能以医学界以外难以理解的方式表达，更不用说语言学家或自动化系统。 我们还必须了解某些事件的法律相关性（例如，“我们讨论了可能的副作用”），即使这些事件似乎与病人的实际护理似乎无关。

最后，每个专业和笔记类型都有单独的约定。 在结肠癌癌症评论中，美国癌症联合委员会（AJCC）的分期代码（例如，T4N1，指示肿瘤的性质，淋巴结和转移受累）被精确记录，但在脑癌注中很大程度上缺失 我们项目中的第二个语料库。 因此，虽然临床笔记有许多相似之处，但没有足够领域专业知识的注释者可能需要额外的培训来适应新临床子领域的推论和细微差别。

# 3 标注临床领域中的事件和时间表达

在标准化文本中的事件和时间表达的注释方面已经做了大量的工作。 最广泛使用的方法是ISOTimeML规范（Pustejovsky 等，2010），ISO标准为注释和分析时间，事件和事件关系提供了一个通用框架。 正如ISO-TimeML所定义的，事件指的是可以说“获得或保持真实，发生或发生”的任何事情。 这是一个广泛的事件概念，与巴赫使用术语“可能性”（Bach，1986）以及AI中的流体概念（McCarthy，2002）一致。

由于THYME项目的目标涉及从临床记录中自动识别患者的临床时间表，所以采用的纳入事件领域的范围比ISO-TimeML3更广泛。在THYME-TimeML指南中，EVENT与临床时间表相关，即任何会在患者护理或生活的详细时间表上显示的事物。然后将用于EVENT的最好的单词语法头用作其跨度。 例如，诊断肯定会出现在这样的时间线上，就像肿瘤，疾病或手术一样。 另一方面，在整个临床时间表的相关时间段内持续存在的实体（本体论圈中的终结者）不会被视为事件状。 这包括病人，其他人类（病人的岳母或医生），组织（急诊室），非解剖对象（病人的车）或病人解剖体的个别部分（手臂不是 一个事件，除非缺失或其他值得注意）。

为了达到我们明确的目标，THYME-TimeML指南引入了ISO-TimeML规定的另外两种解释级别：（i）明确的任务; 和（ii）明确标识的域。 通过专注于从临床叙述中创建临床时间表，指南强加了一个约束条件，该约束条件不能被广泛定义和独立于领域的注释模式所假设。

根据我们的指南注释的一些EVENT被认为是有意义和均衡的，主要是凭借特定的临床或法律价值。 例如，AJCC分期代码（在第2节中讨论）只是在患者护理的特定时刻将代码分配给肿瘤的意义上是均匀的。 但是，它们对医生来说非常重要且具有信息价值，因此我们选择专门对其进行注释，以便他们能够在临床环境中显示患者的时间表。

同样，由于建立知情同意和病人的风险知识的法律压力，整段临床笔记都致力于记录医生对风险，计划和替代策略的讨论。 因此，我们注释讨论动词（“我们谈论了这种药物的风险”），同意（“她同意当前的计划”）和理解（“拉森夫人重复了潜在的副作用”）。 尽管它们与法律防卫相比，更多地与医疗相关。

也正因为临床语言的这种基础，实体和其他非事件经常被解释为它们相关的均衡性质。 有两种主要类型，这是语义解释的重大转变：

（2）a作为事件的药物：

订单：Lariam每日两次。

b作为事件的障碍：

左肺肿瘤。

在这两种情况下，通常没有标记为事件的实体都是这样标识的，因为它们为正在构建的临床时间线提供重要信息。 在（2a）中，例如，“每日两次”的TIMEX3被解释为针对服用药物的患者的可能性而不是处方事件的范围。 在句子（2b）中，“肿瘤”被解释为具有位于解剖区域内的肿瘤的患者的静态可能性，而不是实体内的实体。

在医学领域，这些对药物，增长和状态代码的解释是明确而一致的。 临床笔记中的医生（与生物医学研究文献不同）不会在没有相关（隐式）管理EVENT的情况下讨论药物（尽管有些提及可能是假设的，通用的或否定的）。 类似地，提及症状或障碍反映患者生活中的事件，而不是抽象实体。 考虑到这些解释，我们可以安全地推断出，例如，所有类型为紊乱，化学/药物，程序和征兆/症状的UMLS（统一医学语言系统，（Bodenreider，2004））实体都将是事件。

一般来说，在医学领域中，阅读医生使用的速记表达的“线条之间”是重要的，并且识别特定解剖部位或药物所指的隐含事件。

# 4 对临床领域ISO-TimeML的修改

总体而言，我们发现临床领域中临时注释所需的规范不需要对现有的通用领域规范进行实质性修改。 临床领域包括不缺乏推断，短语和不寻常的语言使用，但潜在时间表的结构并不是唯一的。

因此，我们已经能够采用ISO-TimeML的大部分框架，根据需要调整指导方针，并重新标注注释的重点。 这体现在一个全面的指南中，纳入了临床数据中所看到的事件和时间表达的特定模式和用途。 这种方法可使得到的注释与现有的解决方案互操作，同时仍能适应文本性质的主要差异。 我们的指导方针以及注释数据可在网站http://thyme.healthnlp.org4查阅。

我们将ISO-TimeML规范扩展到临床领域旨在解决医学文本中的特定构造，含义和现象。 我们的模式与ISO-TimeML有一些不同之处。

**事件属性** 我们既简化了EVENTs的ISO-TimeML编码，又扩展了它以满足临床领域的需求和临床叙述的特定语言目标。

例如，考虑如何在ISO-TimeML中处理模态从属关系。 这涉及事件的语义表征为“likely”，“possible”，或者由观察，证据或传闻所呈现。 所有这些都在SLINK（从属链接）关系（Pustejovsky 等，2005）中用ISO-TimeML进行了组合分析。 在接受ISO-TimeML对事件模态的定义的同时，我们简化了当前指南中的注释任务，以便EVENT现在具有“上下文模式”，“上下文方面”和“持久性”的属性。

上下文模态允许实际值，假设值，被隐藏值和一般值。 ACTUAL涵盖了实际发生的事件，例如“我们注意到肿瘤”。 HYPOTHETICAL涵盖条件和可能性，例如，“如果她发展成肿瘤”。 HEDGED适用于医生提供诊断但谨慎行事的情况，以避免不正确诊断或忽视正确诊断的法律责任。 例如：

(3) a. The signal in the MRI is not inconsistent with a tumor in the spleen.

b. The rash appears to be measles, awaiting antibody test to confirm.

（3）a. MRI中的信号与脾脏中的肿瘤不矛盾。

b. 皮疹似乎是麻疹，等待抗体测试来确认。

这些HEDGED EVENT比假设的诊断更真实，并且可能值得作为诊断历史的一部分包含在时间表中，但不能与确认的事实混淆。 这些（以及医学领域中的其他形式的不确定性）在（Vincze等，2008）中进行了广泛讨论。 相反，GENERIC EVENTs并不是指患者的疾病或治疗，而是通常讨论疾病或治疗（通常在患者的特定人群中）。 例如：

(4) In other patients without significant **comorbidity** that can **tolerate** adjuvant **chemotherapy**, there is a **benefit** to systemic adjuvant **chemotherapy**.

（4）对于无明显**并发症**的其他可**耐受**辅助**化疗**的患者，全身辅助**化疗有益**。

如果粘贴在任何患者的便条中，这些部分都是正确的，并且经常是相同的文本块，反复用于证明行动或治疗的方法以及防范责任。

语境方面（区别于语法方面）允许临床必需的类别INTERMITTENT。 这有助于将间歇性事件（如呕吐或癫痫发作）与持续不断的更加静止的事件（如发烧或酸痛）区分开来。 例如，（5a）中的粗体EVENT将被标记为INTERMITTENT，而（5b）中的粗体不会：

(5) a She has been **vomiting** since June.

b She has had **swelling** since June.

（5）a 自6月起她一直呕吐。

b 自6月以来，她发生了肿胀。、

在第一种情况下，我们假设她的呕吐是间歇性的，即自6月以来有几次不呕吐。 在第二种情况下，除非另有明确说明（“她偶尔有肿胀”），否则我们认为肿胀是一个恒定状态。 当某个EVENT的特定实例处于间歇状态时，也会使用此属性，即使它通常不是：

(6) Since starting her new regime, she has had occasional bouts of fever, but is feeling much better.

（6）自从开始新措施以来，她偶尔会发高烧，但感觉好多了。

持久性属性有两个值，有限和永久。持久性是疾病本身的一种属性，大致与“慢性”与“急性”疾病的医学概念相对应，它表明诊断后疾病是否持续存在。例如，像多发性硬化症这样的（目前）不可治愈的疾病将被归类为永久性的，因此，一旦在患者的便笺中提及，就会被假定为在患者的时间表结束时持续存在。这与“流感”或“发热”等有限性疾病相比较，如果在随后的说明中未提及，应认为其治愈，不再属于患者的时间表。因为它需要领域特定的知识，尽管在规范中存在，永久性目前没有注释。然而，注释者接受了基本思想方面的培训，并讲述了随后的公理分配。这个属性添加到我们的模式旨在减轻注释者的任何责任感，以某种其他方式表达这种推断信息。

**TIMEX3类型** 临床领域中的时间表达（TIMEX3s）的功能与一般语言社区中的功能相同，除了两个明显的例外。 ISO-TimeML SET（频率声明）在医疗领域中发生得相当频繁，特别是在药物和治疗方面。 注释中的用药部分通常包含很长的药物清单，每个药物都有一个特定的相关组（“Claritin 30mg每日两次”），进一步的时间规格并不少见（例如，“每天三次，每餐一次”，“每周一次睡前”）。

医疗领域的第二大变革是一种新型的TIMEX3，我们称之为PREPOSTEXP。 这涵盖了“术前”，“术后”和“术中”等时间复杂的术语。 这些时间表达式指定了一个时间跨度，通常仅在一侧，通过并入的事件（一个操作，在先前的EVENT中）。 在很多情况下，所指对象是明确的：

(7) She underwent **hemicolectomy** last week, and had some postoperative **bleeding**.

（7）她上周接受了半结肠切除术，术后出血。

在此我们了解到“术后”是指“半结肠切除术后的一段时间”。 在这些情况下，PREPOSTEXP明确了出血和血肿切除术之间的时间联系。 在其他情况下，不存在明确的指称：

(8) Patient shows some **post-procedure** scarring.

（8）患者出现一些手术后疤痕。

在这些情况下，如果没有提到任何过程（或者从未明确解析过参考），我们将PREPOSTEXP视为一个叙述容器（参见第5节），涵盖未命名过程之后的时间跨度。

最后，值得注意的是，相对于一般领域，标准化TIMEX3的过程要复杂得多，因为许多时间表达式不是固定在日期或时间上，而是固定在其他EVENT上（其日期通常不被医生提及或未知）。 随着我们朝着一个完整的系统迈进，我们正在努力扩展TIMEX3标准化的ISO-TimeML系统，以允许在没有参考日期存在的情况下，将某些值分配给“在她的半切除术后数月内”。 与ISO TC 37SC 4讨论的ISO-TimeML计划在未来发布的标准中引用此类TIMEX3。

# 5 时间顺序和叙述容器

时间线的语义内容和信息影响被编码在临床笔记中存在的临时表达式和事件表达式之间所确定的排序关系中。 ISO-TimeML从区间演算（Allen，1983）中指定了标准的十三个“艾伦关系”，它被称为TLINK值。 对于无导向的通用注释，可以注释的关系的数量随着事件和时间的数量呈二次增长，并且该任务很快变得难以管理。 然而，我们可以采取一些策略来使这个标签任务更易于处理。 文本中的时间顺序关系有三种：

1.两个事件之间的关系

2.两次之间的关系

3.时间和事件之间的关系。

ISO-TimeML作为语言传达的时间信息的正式规范，不会对这些排序类型加以区分。但是，人类根据当地的时间标记和叙述中建立的话语关系（Miltsakaki等，2004; Poesio，2004）确实做出了区分。

由于人类很难捕捉到记录中存在的每种关系（以及注释者试图这样做时产生的分歧），因此注释准则描述必须考虑减少关系数量的方法就至关重要了，同时仍要做到最大限度的信息时间链接。 我们发现，之前诠释方法中的许多不足都源于两个相互竞争的目标之间的相互作用：

•准则要规定应执行的某些类型的注释;

•准则不应强制在不需要时执行注释。

第一个目标的失败会导致诠释不足和忽视为推断和分析提供必要信息的关系。 第二个目标的失败会导致过度诠释，从而产生复杂的时间关系网络，这些网络会产生大多数可推断的信息，但这会使注释和裁决变得相当复杂。

我们在时间关系诠释中处理两个目标的方法是在Pustejovsky和Stubbs（2011）中讨论的叙述性容器。 叙事容器可以被认为是EVENT或一系列EVENT可能落入其中的时间桶，或者是在给定时间或情况下的自然聚类EVENT。 这些叙述容器通常由日期或其他时间表达式（其中发生各种不同事件）表示（或“锚定”），尽管它们也可以被锚定到更抽象的概念（可能涉及各种EVENT的“恢复” ）或甚至是持续性事件（在手术期间可能发生许多其他EVENT）。 我们不是在每个EVENT之间标记每个可能的TLINK，而是尝试将所有EVENT链接到它们的叙述性容器，然后链接这些容器，以便可以通过推断将所包含的EVENT链接起来。

首先，注释者将每个事件分配到四个广义叙述容器中的一个：在DOCTIME之前，在DOCTIME之前且覆盖DOCTIME，仅覆盖DOCTIME或在DOCTIME之后。 此叙述容器由EVENT属性Doc-TimeRel标识。 在分配DocTimeRel之后，必须使用时间链接（TLINK）指定叙述容器关系的其余部分。 这种TLINK有五种不同的时间关系：BEFORE，OVERLAP，BEGINS-ON，ENDS-ON和CONTAINS5。 由于我们的叙述性容器方法，CONTAINS是最常见的关系。

作为叙述性容器锚点的事件不被标记为容器本身。 相反，注释者使用叙述容器的思想来帮助他们可视化文档中的时间关系，然后制作一系列CONTAINS TLINK注释，将EVENT和TIMEX3建立为锚点并指定其内容。 如果注释者正确地完成他们的工作，正确地实施DocTimeRel并创建精确的TLINK，那么对于文档中出现的叙述性容器的理解将自然从注释文本中出现。

叙事容器引入的主要优点是：叙述事件被放置在文本中明确提到的边界时间区间内。 这允许通过事后推理，时间推理和领域知识而不是通过明确的（并且耗时的）一个接一个的时间关系注释将单独容器内的EVENT链接起来。

第二个优点是，这种方法可以很好地处理常规和临床领域的一般记叙性结构，并且为解释时间表提供了引人注目的和有用的比喻。 通常，特别是在临床病史中，医生会围绕给定日期（例如，从2009年6月开始的整段），特定住院（“她1月在Mercy

住院期间”）的症状，干预和诊断，或者参考疾病或治疗时间（“她在接受化疗时”）。 即使特定EVENT没有在集群内明确排序（通常是因为可以根据领域知识轻松推断顺序），将EVENT放入容器通常很容易，并且只有少数TLINK可以将容器相对于另一个排序具有足够的细节来创建对整个时间表的临床有用的理解。

叙述容器也允许推断嵌套容器内的子事件之间的关系：

(9) December 19th: The patient underwent an **MRI** and **EKG** as well as emergency surgery. During the **surgery**, the patient experienced mild **tachycardia**, and she also **bled** significantly during the initial **incision**.

1. December 19th CONTAINS **MRI**

2. December 19th CONTAINS **EKG**

3. December 19th CONTAINS **surgery**

a. **surgery** CONTAINS **tachycardia**

b. **surgery** CONTAINS **incision**

c. **incision** CONTAINS **bled**

（9）12月19日：患者接受了**MRI**和**EKG**紧急手术。 **手术**过程中，患者出现轻度**心动过速**，并且在**初始切口**时也出现明显的**出血**。

1. 12月19日CONTAINS MRI

2. 12月19日包含EKG

3. 12月19日CONTAINS手术一个。

a. 手术包含心动过速

b. 手术CONTAINS切口

C. 切口包含流血

通过我们的容器嵌套，我们可以自动推断出'流血'发生在12月19日（因为'第19'包含'手术'包含'切口'包含'流血'）。 这也可以捕获事件/子事件关系，以及复杂时间交互的快速表达。

# 6 显式与隐式标注

给定一个规范语言，基本上有两种方法将元素引入到被注释的文档（数据源）中：6

•手动注释：元素直接由人类注释器按照指南引入文档。

•自动（推断）注释：元素是通过应用自动过程创建的，该过程引入了可从人类注释导出的新元素。

在某些情况下，这些添加的注释是从人类注释逻辑推导出来的。 显式标记的时间关系可用于推断未标记但通过闭包而隐含存在的其他关系。 例如，考虑到事件A，B和C以及TLINK'A BEFORE B'和'B BEFORE C'，TLINK'A BEFORE C'可以被自动推断。 反复应用这样的推理规则可以生成所有可导出的TLINK（Verhagen，2005）。 我们可以使用这种闭包的想法来向我们的注释者显示哪些注释不需要明确标记，从而节省时间和精力。 我们还将这些关闭规则纳入了我们的时间关系计算机间协议（IAA）计算中，详见7.2节。

文本注释后的规则自动应用不限于逻辑上可推断的关系或EVENT的标记。 在临床领域，团体内共享的知识和对简洁写作的压力的结合导致了许多常见的推断关系。 举例来说，这个句子：

(10) Jan 2013: **Colonoscopy, biopsies. Pathology** showed **adenocarcinoma, resected** at Mercy**. Diagnosis T3N1 Adenocarcinoma.**

（10）2013年1月：**结肠镜检查，活检。 病理学**显示在Mercy**切除的腺癌。 诊断T3N1腺癌。**

在这句话中，只明确说明“2013年1月”与事件（粗体）之间的CONTAINS关系。 然而，根据已知的结肠癌治疗进展，我们可以推断结肠镜检查首先发生，在结肠镜检查期间发生活检，病理发生后，诊断（这里是腺癌）在病理学检验后产生，切除肿瘤在诊断后发生。 最后一句中AJCC分期信息的存在（以及确诊腺癌诊断）意味着对切除标本进行术后病理检查，因为如果没有这种额外的检查，AJCC分期信息就无法确定。

这些推论自然而然地提供给了领域专家，但是如果没有相当多的注释者培训，医学界以外的人很难获得这些推论。 明确我们对这些“理解顺序”的理解是至关重要的; 虽然它们在我们的模式中没有被人类注释者所标记，但注释者经常发现它最初令人沮丧的是将这些（纯粹推论）的关系置之不理。 尽管我们的许多（主要是受过语言训练的）注释者学会了理解这些模式，但我们选择将它们从手动任务中排除，因为具有不同程度的领域知识的新注释者可能会被要求手动注释它们。

在整个临床领域中发现了类似的没提过但能理解的顺序。 如第3节所述，永久性和情境性方面：间歇性是症状和疾病本身的特性，而不是患者的特殊情况。 因此，这些属性可以很容易地通过医学本体进行识别和标记，然后自动分配给被视为特定医疗命名实体的事件。

最后，由于电子病历系统的特殊性，一些注释必须以编程方式完成。 患者就诊（或病理学/放射学咨询）的准确日期通常记录为EHR自身而不是文本内的元数据，从而使得标识的DOCTIME（或自动区段修改时间）难以在去识别的明文数据中访问 ，但很容易自动找到。

# 7 结果

我们报告了THYME结肠癌语料库中已发布的子集的注释结果，其中包括35个诊断患有结肠癌的患者的临床笔记和病理报告，总共107个文件。 每张笔记都由科罗拉多大学语言学专业的一对研究生或本科生注释，然后由领域专家进行审核。 这些临床叙述是从主要医疗中心（Mayo Clinic）的EHR采样的。所有患者的敏感信息除原始日期外都被替换了。

|  |  |
| --- | --- |
| 标记类型 | 原词数量 |
| 事件 | 15769 |
| 时间 | 1426 |
| 关系 | 7935 |
| 总计 | 25130 |

表 1： 不同标记类型的词的频率

|  |  |  |
| --- | --- | --- |
| 关系类型 | 原词数量 | 比例 |
| CONTAINS | 5,112 | 64.42% |
| OVERLAP | 1,205 | 15.19% |
| BEFORE | 1,004 | 12.65% |
| BEGINS-ON | 488 | 6.15% |
| ENDS-ON | 126 | 1.59% |
| 总计 | 7,935 | 100.00% |

表 2： 不同关系类型的关系频率

**7.1 描述性统计**

表1列出了裁决黄金注释中事件，时间表达和链接的原始计数。 表2列出了在判定关系黄金注释中TLINK类型的数量和百分比。

**7.2 内部注释者协议**

我们在THYME语料库上报告内部注释者协议（IAA）结果。每个笔记都由两位独立的注释员注释。 最终的黄金标准是在第三位注释员进行了分歧裁决后产生的。

|  |  |  |
| --- | --- | --- |
| 标记类型 | F1-Score | Alpha |
| 事件 | 5,112 | 64.42% |
| 时间 | 1,205 | 15.19% |
| LINK: Participants only | 1,004 | 12.65% |
| LINK: Participants+type | 488 | 6.15% |
| LINK: CONTAINS | 126 | 1.59% |

表 3： 不同标记类型的IAA

我们通过使用明确标记的时间关系来识别未标记但隐含存在的其他人，我们计算IAA为F1分数和Krippendorff's Alpha（Krippendorff，2012）。 在计算IAA时，推断的TLINK对分数无贡献，匹配或不匹配。 例如，如果两个注释器在B之前标记A BEFORE B且B BEFORE C，为了防止人为地夸大协议评分，推断的A BEFORE C将被忽略。 同样，如果一个注释者在B之前标注了A BEFORE B且B BEFORE C，而另一个注释者没有，则推断的A BEFORE C不被计数。 但是，如果一个注释者明确标记了A BEFORE C，那么将使用等效推断的TLINK来匹配它。 EVENT和TIMEX3 IAA分别基于精确和重叠跨度生成。 这些结果报告在表3中。

|  |  |  |  |
| --- | --- | --- | --- |
| 事件属性 | F1-Score | Alpha | |
| DocTimeRel | 0.7189 | 0.6889 |
| Cont.Aspect | 0.9947 | 0.9930 |
| Cont.Modality | 0.9547 | 0.9420 |

表 4： 不同事件属性的IAA

THYME语料库在EVENT属性方面也与ISOTimeML不同，增加了DocTimeRel，ContextualModality和ContextualAspect。 这些性能的IAA在表4中。

**7.3 基线系统**

为了了解将现有的临时信息提取系统应用于临床领域需要做多少工作，我们采用了免费的ClearTK-TimeML系统（Bethard，2013），该系统是TempEval 2013（UzZaman）中表现最佳的系统之一 2013年），估其表现在THYME语料库上得到了验证。

ClearTK-TimeML使用在TempEval 2013培训数据上训练的支持向量机分类器，使用一小组功能，包括字符模式，令牌，词干，词性标记，选区树中的附近节点以及小时间字地名词典。 对于EVENT和TIMEX3，ClearTK-TimeML系统可直接应用于THYME语料库。 对于DocTimeRels，在将INCLUDES映射到OVERLAP之后，EVENT的关系取自该EVENT和文档创建时间之间的TLINK。 没有这种TLINK的事件被假定为具有OVERLAP的Doc-TimeRel。 对于其他时间关系，INCLUDES被映射到CONTAINS。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | TemEval 2013 | | | THYME Corpus | | |
|  | P | R | F1 | P | R | F1 |
| 时间 | 83.2 | 71.7 | 77.0 | 59.3 | 42.8 | 49.7 |
| 事件 | 81.4 | 76.4 | 78.8 | 78.9 | 23.9 | 36.6 |
| 文档创建时间 | - | - | - | 47.4 | 47.4 | 47.4 |
| 关系7 | 28.6 | 30.9 | 26.6 | 22.7 | 18.6 | 20.4 |
| 事件-时间 | - | - | - | 32.3 | 60.7 | 42.1 |
| 事件-事件 | - | - | - | 7.0 | 3.0 | 4.2 |

表5：2013年TempEval竞赛中报告的ClearTK-TimeML模型的性能，以及适用于THYME语料库开发集的性能。

TempEval 2013和THYME语料库的系统结果如表5所示。对于时间表达式，转向临床数据时的性能从77.0下降到49.7，降低了大约25％。 对于事件，下降比78.8至36.6更大，约为40％，很可能是因为有大量临床症状，疾病，失调等词汇在系统训练过程中从未出现过。 时间关系比较难以比较，因为TempEval将DocTimeRel和其他时间关系集中在一起，并且在其评估指标7中存在若干差异。 但是，我们至少可以看到ClearTK-TimeML系统在时间关系上的表现在临床文本上很低，仅达到20.4的F1。

这些结果表明，临床叙述确实对时态信息提取系统提出了新的挑战，并且获得特定领域的训练数据对于准确提取临床领域是至关重要的。 同时，尽管对临床叙述所需的ISO-TimeML进行了几次扩展，但我们仍然能够将现有基于ISO-TimeML的系统应用到我们的语料库，这是令人鼓舞的。

# 8 讨论

CONTAINS在THYME语料库中扮演着重要角色，占TLINK注释量的66％，而第二种最常见类型的OVERLAP仅为14.6％。 我们还看到，BEFORE链接比OVERLAP和CONTAINS相对不常见，说明时间轴上的大部分时间顺序是通过使用许多垂直链接（CONTAINS，OVERLAP）来构建容器以及很少的横向链接（BEFORE，BEGINS- ON，ENDS-ON）来为它们排序。

关于EVENT和时间表达式的IAA很强大，但是将隐含的EVENT（不应标记）与明确的可标记EVENT区分开，仍然是最大的分歧来源之一。 与2012年i2b2挑战的数据（Sun等人，2013b）相比，我们的IAA数据非常相似。 即使有了我们更复杂的模式，我们获得了EVENTs的F1得分为0.8038（与部分匹配的i2b2得分为0.87相比）。 对于TIMEX3s，我们的F1分数为0.8047，而i2b2的F1分数为0.89。

TLINKing医疗事件仍然是一项非常艰巨的任务。通过使用我们的叙述性容器方法来约束必要注释的数量，并通过消除经常混淆的反向关系（比如“之后”和“期间”）（这两个都不是针对i2b2数据完成的），我们能够显着提高i2b2 TLINK标签协议F1得分为0.39，我们的语料库中所有LINK的协议得分为0.5012。大多数剩余的注释者分歧来自于关于任何两个EVENT是否需要在它们之间或推断的TLINK之间的明确TLINK而不是什么类型的TLINK（例如，BEFORE与CONTAINS）的不同意见。尽管我们的结果仍然高于i2b2报告的结果，并且与之前公布的一般新闻数据一致，但我们并不满意。改进IAA是未来工作的一个重要目标，通过进一步的培训，规范，经验和标准化，我们希望澄清明确的TLINKS的背景。

新闻训练的时间信息提取系统在应用于THYME语料库的临床文本时，性能显着下降。 但由于语料库是ISO-TimeML的扩展，未来的工作将能够在THYME语料库的注释中训练符合ISO-TimeML的系统，以减少或消除这种性能差距。

我们的工作可能实现的一些应用包括：（1）更好地理解事件语义，比如疾病是慢性的还是急性的，以及它通常的自然历史，（2）这些事件的典型事件持续时间，（3） 特定领域的事件及其在最终时间轴上的重要性，以及更一般地说，（4）粗糙的时间性和叙事容器的重要性，作为实现更细粒度时间表的一个步骤。

我们有几种正在进行和未来的工作方向。首先，我们正在努力展示THYME语料库用于培训机器学习模型的实用性。我们设计了带有选区树核的支持向量机模型，能够在EVENT-TIMEX3叙事容器识别任务中达到0.737的F1分数（Miller 等，2013），并且我们正在开发训练模型以识别事件，时间和其他类型的时间关系。其次，根据我们激励的用例，我们正在努力将这些注释数据与时间线可视化工具集成在一起，并将这些注释用于质量保健研究。例如，我们使用基于这项工作的时间推理来研究甲氨喋呤在EHRs大型语料库中的肝脏毒性（Lin 等，审查中）]。最后，我们计划试图将我们的事件概念（任何应该在域适当的时间线上可见的内容）应用到其他域。它应该自然地转移到关于其他（非癌症）状况的临床记录，甚至是其他类型的临床记录中，因为某些基本事件应始终包含在病人的时间表中。将我们的事件概念应用于更远距离的领域（如法律意见），首先需要确定事件必须出现在时间轴上的领域内的共识。

# 9 结论

医生使用的简明速记隐含了大量关于建立详细时间表的关键临床信息。 许多事件只能通过诸如“肿瘤”之类的术语来提及，而事件本身的属性（例如“间歇性”）可能未被指定。 另外，时间线上的事件排序通常留给读者根据领域特定的知识来推断。 注释准则有义务表明只应注释信息事件顺序，而将特定领域的顺序留给后注释推理。 本文详细介绍了我们的方法，即将现有的ISO-TimeML标准修改为隐性信息的恢复，并定义支持在这个复杂领域内进行注释的指南。 我们的指导方针以及注释数据可在http://thyme.healthnlp.org上获取，并且已经提出将全部语料库用于SemEval 2015共享任务。

# 致谢

所述项目得到国家医学图书馆的拨款号码R01LM010090和U54LM008748的支持。 内容完全是作者的责任，不一定代表国家医学图书馆或国立卫生研究院的官方观点。

我们还要感谢Mayo诊所的Piet C. de Groen博士和Brad Erickson博士以及William F. Styler III博士对模式的贡献以及我们对临床语言错综复杂的理解。

References

# 参考文献

[1] James F Allen. 1983. Maintaining knowledge about temporal intervals. Communications of the ACM, 26(11):832–843.

[2] Emmon Bach. 1986. The algebra of events. Linguistics and philosophy, 9(1):5–16.

[3] Steven Bethard. 2013. Cleartk-timeml: A minimalist approach to tempeval 2013. In Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), pages 10–14, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

[4] Olivier Bodenreider. 2004. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic acids research, 32(Database issue):D267–D270, January.

[5] Philip Bramsen, Pawan Deshpande, Yoong Keok Lee, and Regina Barzilay. 2006. Finding temporal order in discharge summaries. In AMIA Annual Symposium Proceedings, volume 2006, page 81. American Medical Informatics Association.

[6] Carlo Combi, Yuval Shahar, et al. 1997. Temporal reasoning and temporal data maintenance in medicine: issues and challenges. Computers in biology and medicine, 27(5):353–368.

[7] Robert H Dolin. 1995. Modeling the temporal complexities of symptoms. Journal of the American Medical Informatics Association, 2(5):323–331.

[8] George Hripcsak, Nicholas D Soulakis, Li Li, Frances P Morrison, Albert M Lai, Carol Friedman, Neil S Calman, and Farzad Mostashari. 2009. Syndromic surveillance using ambulatory electronic health records. Journal of the American Medical Informatics Association, 16(3):354–361.

[9] Ann K Irvine, Stephanie W Haas, and Tessa Sullivan. 2008. Tn-ties: A system for extracting temporal information from emergency department triage notes. In AMIA Annual Symposium proceedings, volume 2008, page 328. American Medical Informatics Association.

[10] Elpida T Keravnou. 1997. Temporal abstraction of medical data: Deriving periodicity. In Intelligent Data Analysis in Medicine and Pharmacology, pages 61–79. Springer.

[11] Elpida T Keravnou. 1997. Temporal abstraction of medical data: Deriving periodicity. In Intelligent Data Analysis in Medicine and Pharmacology, pages 61–79. Springer.

[12] Chen Lin, Elizabeth Karlson, Dmitriy Dligach, Monica Ramirez, Timothy Miller, Huan Mo, Natalie Braggs, Andrew Cagan, Joshua Denny, and Guergana. Savova. under review. Automatic identification of methotrexade-induced liver toxicity in rheumatoid arthritis patients from the electronic medical records. Journal of the Medical Informatics Association.

[13] John McCarthy. 2002. Actions and other events in situation calculus. In Proceedings of the International conference on Principles of Knowledge Representation and Reasoning, pages 615–628. Morgan Kaufmann Publishers; 1998.

[14] St´ephaneMMeystre, Guergana K Savova, Karin C Kipper- Schuler, John F Hurdle, et al. 2008. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform, 35:128–44.

[15] Timothy Miller, Steven Bethard, Dmitriy Dligach, Sameer Pradhan, Chen Lin, and Guergana Savova. 2013. Discovering temporal narrative containers in clinical text. In Proceedings of the 2013 Workshop on Biomedical Natural Langua ge Processing, pages 18–26, Sofia, Bulgaria, August. Association for Computational Linguistics.

[16] Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. 2004. The penn discourse treebank. In In Proceedings of LREC 2004.

[17] Massimo Poesio. 2004. Discourse annotation and semantic annotation in the gnome corpus. In In Proceedings of the ACL Workshop on Discourse Annotation.

[18] James Pustejovsky and Amber Stubbs. 2011. Increasing informativeness in temporal annotation. In Proceedings of the 5th Linguistic Annotation Workshop, pages 152– 160. Association for Computational Linguistics.

[19] James Pustejovsky, Robert Knippen, Jessica Littman, and Roser Sauri. 2005. Temporal and event information in natural language text. Language Resources and Evaluation, 39(2-3):123–164.

[20] James Pustejovsky, Kiyong Lee, Harry Bunt, and Laurent Romary. 2010. Iso-timeml: An international standard for semantic annotation. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta.

[21] Angus Roberts, Robert Gaizauskas, Mark Hepple, George Demetriou, Yikun Guo, and Ian Roberts. 2009. Building a semantically annotated corpus of clinical texts. Journal of biomedical informatics, 42(5):950–966.

[22] Guergana Savova, Steven Bethard,Will Styler, James Martin, Martha Palmer, James Masanz, and Wayne Ward. 2009. Towards temporal relation discovery from the clinical narrative. In AMIA Annual Symposium Proceedings, volume 2009, page 568. American Medical Informatics Association.

[23] Tessa Sullivan, Ann Irvine, and Stephanie W Haas. 2008. It’s all relative: usage of relative temporal expressions in triage notes. Proceedings of the American Society for Information Science and Technology, 45(1):1–8.

[24] Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013a. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. Journal of the American Medical Informatics Association.

[25] Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013b. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. Journal of the American Medical Informatics Association, 20(5):806–813.

[26] Alexander Turchin, Maria Shubina, Eugene Breydo, Merri L Pendergrass, and Jonathan S Einbinder. 2009. Comparison of information content of structured and narrative text data sources on the example of medication intensification. Journal of the American Medical Informatics Association, 16(3):362–370.

[27] Naushad UzZaman, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. 2013. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), pages 1–9, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

[28] Marc Verhagen. 2005. Temporal Closure in an Annotation Environment. Language Resources and Evaluation, 39(2):211–241.

[29] Veronika Vincze, Gyrgy Szarvas, Richrd Farkas, Gyrgy Mra, and Jnos Csirik. 2008. The bioscope corpus: biomedical texts annotated for uncertainty, negation and their scopes. BMC Bioinformatics, 9(Suppl 11):1– 9.

[30] Ying Zhao, George Karypis, and Usama M. Fayyad. 2005. Hierarchical clustering algorithms for document datasets. Data Mining and Knowledge Discovery, 10:141–168.

[31] Jiaping Zheng, Wendy W Chapman, Rebecca S Crowley, and Guergana K Savova. 2011. Coreference resolution: A review of general methodologies and applications in the clinical domain. Journal of biomedical informatics, 44(6):1113–1122.

[32] Li Zhou, Simon Parsons, and George Hripcsak. 2008. The evaluation of a temporal reasoning system in processing clinical discharge summaries. Journal of the American Medical Informatics Association, 15(1):99–106.