

# 基于多模态知识图谱的中医临床诊断辅助系统

刘帅 2020212267 2020219111班

## 摘要:

中医学是中华民族的传统科学，在漫长的发展与演变过程中，不断进行总结发展，并在长期医疗实践中形成的辨证论治的理论体系。在新冠疫情的预防与疾病传播过程中，中医的治疗方法与治疗体系得到了国内外医学专家的广泛关注和高度认可。中医在诊治过程中，以四诊五行八纲以及医疗经验作为理论基础，通过结合病位、病因、病性、病势等多个模态的信息，对病人状态进行诊断。然而，中医临床诊断有别于西医的最大特点在于其“形而上”的诊治观念，从系统整体性推至局部性诊断，进行疾病的根治；相比之下，西医的观念则是从局部性、单位性出发，从围观角度进行局部性诊治，最终达到疾病缓解的目的。在医学智能诊断的范畴上，由于西医诊断呈现“单元性”的特点，因此对于医学图像的辅助识别、诊断、分割等图像类任务已经取得了可观的成果；然而，中医的“望闻问切”四诊方法中，包含着视觉、听觉、语义、脉搏等多个模态的信息，从而给智能诊断带来了一定挑战。 本文将重点讨论中医对于多模态信息的智能诊断方法，利用知识图谱的方法描述中医诊断的不同环节及其载体，通过知识图谱的建立为诊断提供有价值的参考，同时对于集成学习与多模态技术在临床学的应用进行可行性分析和系统构建。

**关键词：**中医临床诊断、系统论治、四诊五行八纲、知识图谱、集成学习、多模态

## 第一章 绪论

### 1.1 研究背景与意义

随着信息社会发展和传统中医学的兴起，中医医学诊断的智能化是信息时代发展的必然趋势。围绕中医临床诊断进行信息化的构建过程，从现象背后解释关于本体性的成因，已经成为人工智能在医学领域的重要研究方向。以本体论、系统论为代表的多模态研究技术为中医临床诊断提供了理想化的研究方向。使得关于视觉、听觉、嗅觉、语义、脉搏等多维度诊断的实践经验能够进行信息融合，从而提供智能化的诊断辅助，为中医临床诊断的丰富和发展提供了全新方向和可能。

### 1.2 知识图谱技术导览

21世纪以来，智能诊断与医学数据库的建立一直是业内热点研究问题之一。2012年，随着知识图谱的问世，在文本和结构化的数据表示和图片、音频等视觉听觉数据表征上有了进一步发展。如阿里巴巴为加强对电商的监管管控建立了电商专业知识图谱，以CPV的方式对商品进行表征。

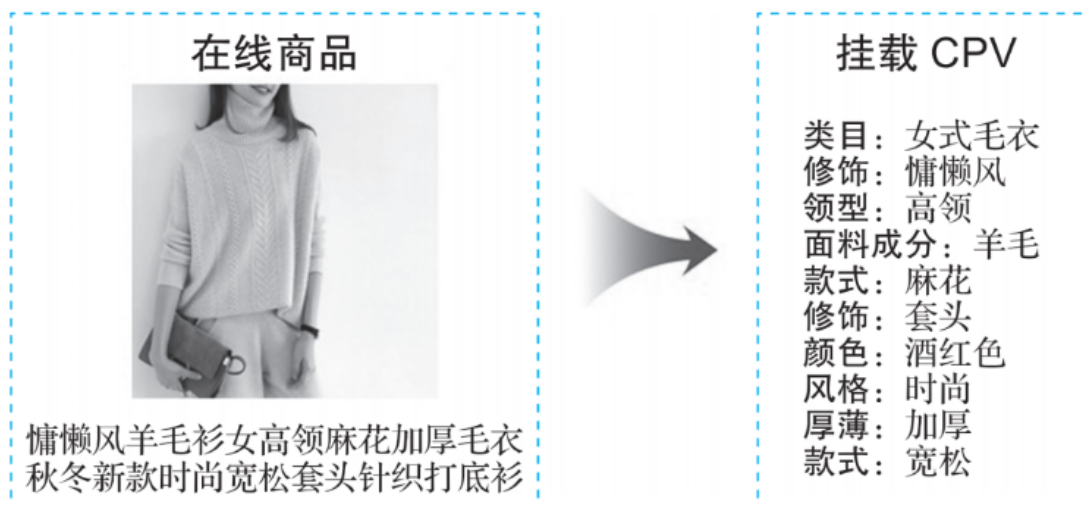
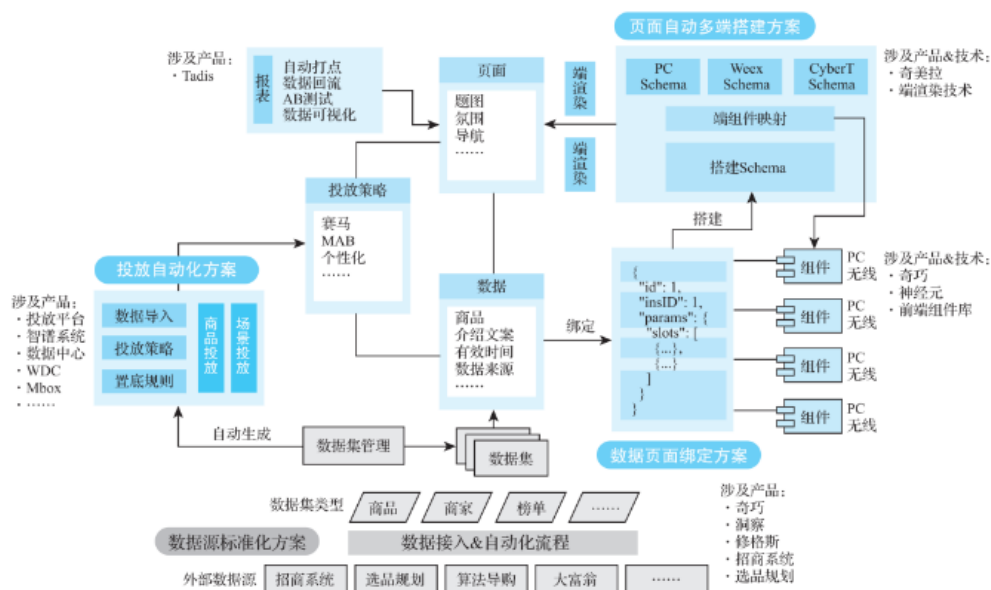


图7 商品CPV表征示意图

随着数据来源和维度增加，通过系统打通，基于数据源实现了产品体系的标准化搭建。



在医疗知识图谱方面，腾讯天衍实验室提出了基于知识图谱的真实场景医疗数据集和评估方式（**An Industry Evaluation of Embedding-based Entity Alignment**），并针对多知识图谱的对齐问题进行了深入探讨。通过提出基于本体指导的知识图谱对齐技术OntoEA，建模类别冲突关系从而丰富嵌入语义表示，避免错误映射。

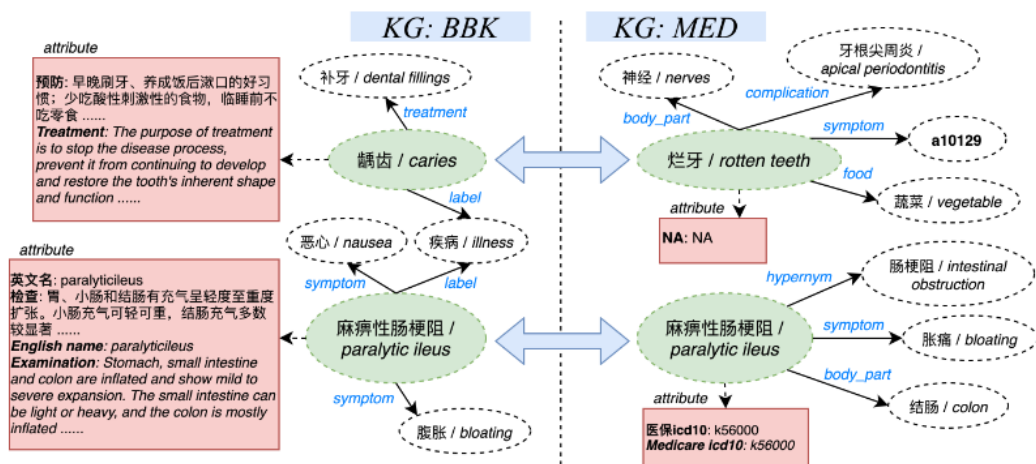
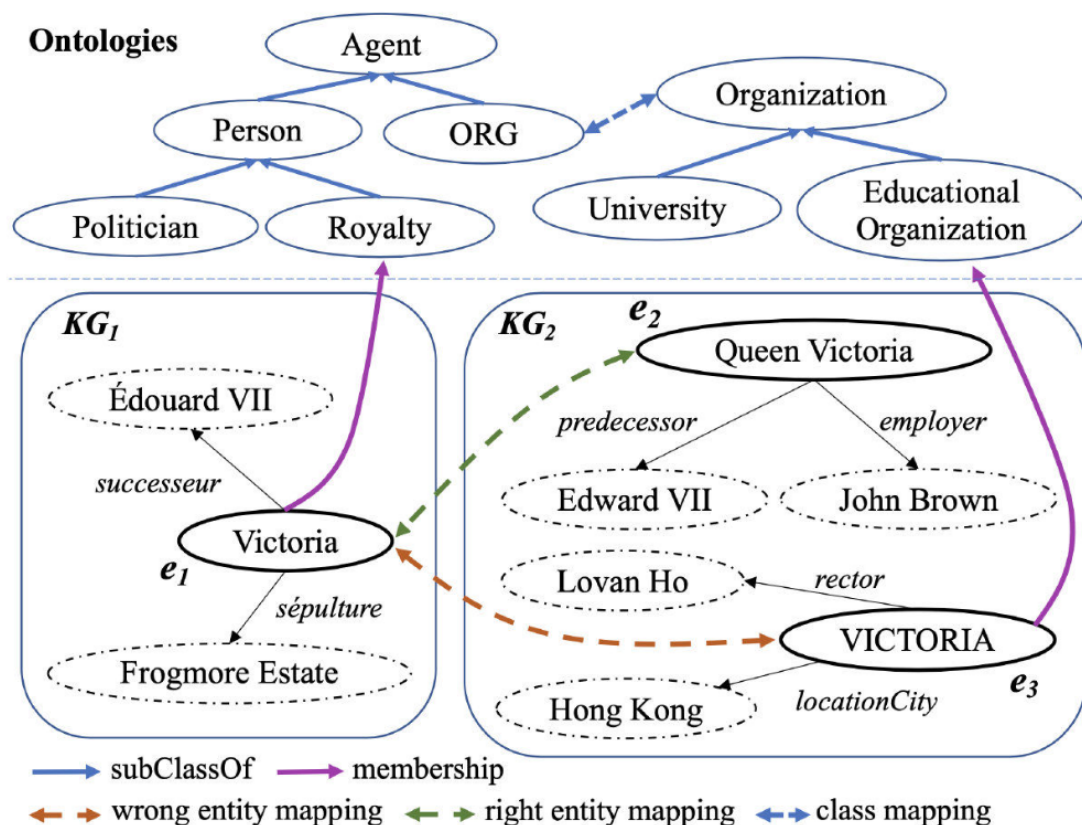


Figure 2: Two mapping examples from MED-BBK-9K with English translations.

注: 利用本体信息修正, 可以将不同域的信息进行分离 (如  $KG_1$  所表示的 Victoria 代表一个皇室成员, 而  $KG_2$  中的 VICTORIA 表示教育机构!), 从而避免信息的错误对齐。



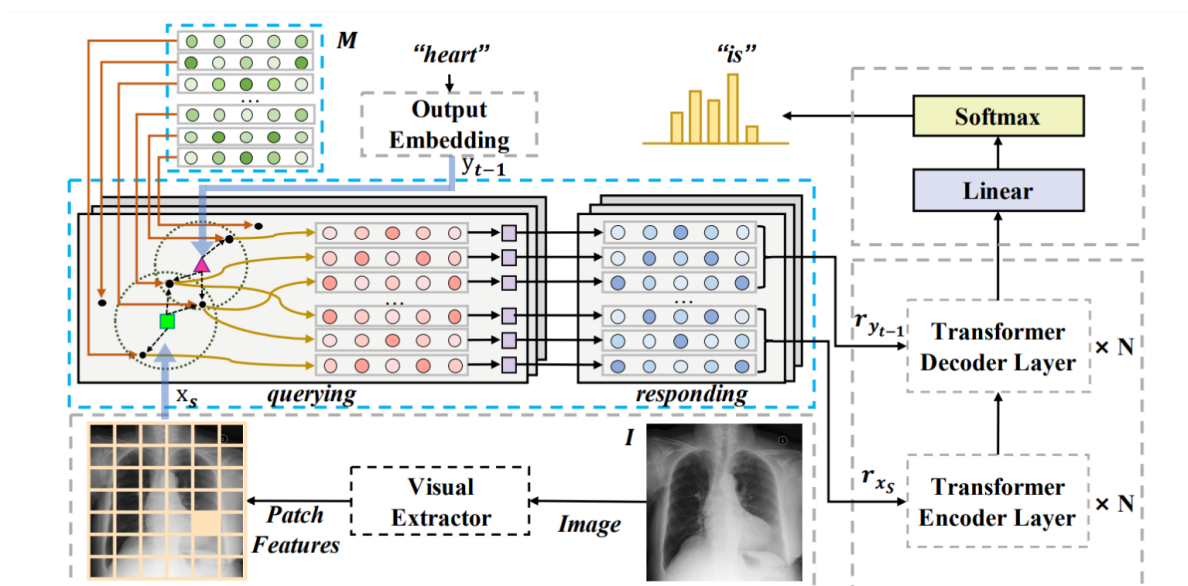
尽管天行实验室提出的知识图谱偏向于解决西医中的医学症状诊断, 然而, 其对于本体概念在知识图谱中的引入为中医智能诊断领域提供了重要尺度。

### 1.3 多模态信息融合与集成学习导览

**模态 (Modality)** 是德国理学家赫尔姆霍茨提出的一种生物学概念, 即生物凭借感知器官与经验来接收信息的通道。例如中医望闻问切诊断方法, 对于患者形体、动态进行望诊; 对于语言气息进行问诊; 对于既往病史、发病经过、痛苦所在等进行问诊; 以及患者脉象的切诊环节。中医诊断中的多模态便是利用如上多维度感官信息融合的方式, 对于患者主体进行病理诊断。

这样利用多种维度信息进行判别、分类的方法更早可追溯到**集成学习 (ensemble learning)** 的传统机器学习方法之中，Boosting系列与Bagging方法为代表，通过定义基学习器并在学习过程中赋予不同学习器权重的思想，来进行多维度信息预测，本文并不会针对集成学习等传统机器学习方法进行探究，只是透过集成学习体会“博采众长”的思想。

在ACL2021会议中，**Cross-modal Memory Networks for Radiology Report Generation**论文提出了跨模态存储器网络，引入了用于对齐文本和语义信息对齐的共享存储器。对于给定图像源序列 $\{x_1, x_2, \dots, x_S\}$ 及文本序列 $\{y_1, y_2, \dots, y_t - 1\}$ ，将二者送入存储器模块。以便在存储器中记录视觉和文本特征的共享信息，以便整个学习过程能够在图像和文本之间显式地映射。



## 第二章 系统设计

引言：本单元将针对系统的可行性，给出系统整体设计架构，对结果的评估体系进行确立，并给出多模态知识图谱的构建方案。

### 2.1 系统可行性分析

2.2.1 **数据基础**：中医长期诊治于临床的过程中，已经形成了系统的诊断方法，这些知识存在于各类中医学典籍、研究文献、临床数据中，具有较为可靠的数据来源。

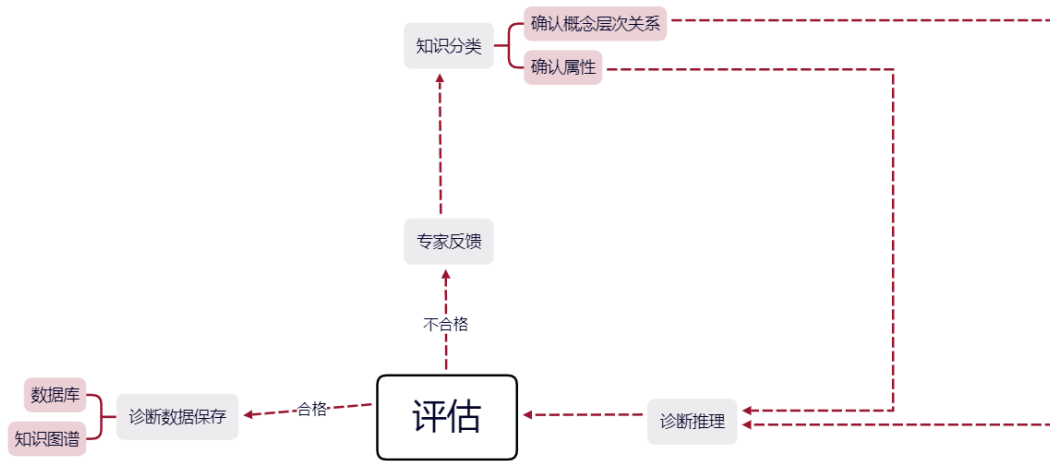
2.2.2 **教育基础**：国内对于中医人才的培养逐步体系化，接触中医学的学生与爱好者众多，从而使得中医理论的建成具有时代性与进步性。

2.2.3 **应用基础**：中医与智能信息结合仍然处于起步阶段，同时，国内外研究者对于西医范畴的诊断已经有了初步进展，部分临床经验具有可借鉴性与可迁移性。

2.2.4 **商用潜力**：中医智能系统诊断可以作为中医诊断的重要辅助工具，可对患者情况进行初步诊断，对“线上问诊”提供了较大便利，并可应用于中医教学、临床、研究等多个领域。

2.2.5 **算法基础**：多模态和知识图谱相关算法的不断建成与完善，涌现出许多有关开源的知识图谱系统，例如依赖专家知识构建完成的知识图谱：**OpenCyc**, **wordnet**, **yago**等，对于智能系统的建立起到了关键性的帮助，通过其他医学方向的迁移与域适应构建中医诊断系统。

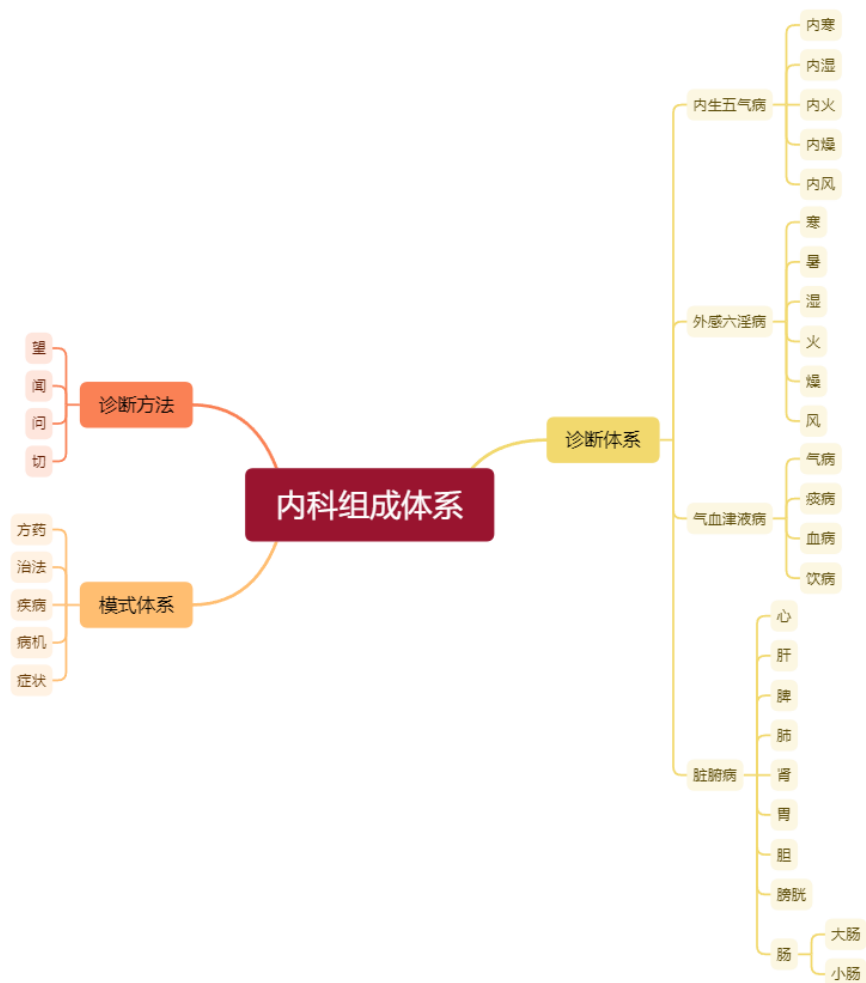




## 2.3 类别确立方法

本文以中医内科为例，对内科领域相关典籍的语义类内容进行划分，从而确定内科诊断种类的从属关系划分。

经整理，中医内科体系可简单分为模式层与实例层。通过中医诊断方法，对于病机和症状进行推断论证，从而指明疾病成因，给出方药与治法。而模式体系的五种要素又根据诊断体系的分类与细化产生其实例化对象，从而进行疾病的诊断过程。



2.4 基于多模态的知识图谱构建方法

多模态信息融合和知识图谱构建的首要任务是分别收集并处理单模态信息，在中医内科学之中，可对面部气色、舌相的视觉信息；气味信息；声音信息；脉象进行单类信息处理，以下针对不同信息的收集与处理提供可行性方案。

2.4.1 视觉信息分类和处理

视觉信息主要分为气色与舌相信息，此处以舌诊为例。借助辅助问诊对于舌相信息进行初步建库，基于专家对于舌相信息的诊断，首先对于舌相进行识别，现有的语义分割模型已经能够较好的识别出舌相信息，对于舌相进行精细定位和分类；其次，分类模型的完备也便于识别舌相的整体信息，从而产生舌相与疾病的映射关系，减少人工建立舌诊库的重复工作。如下图例

证素	舌神	舌色	舌形	舌态	苔色	苔质	润燥	舌下络脉 颜色	舌下络脉 异常程度
阴虚		红/红绛	嫩/瘦小			少苔/剥落/无苔	干燥		
阳虚; 虚寒<		淡白/紫	嫩/胖l	齿痕		薄	润/滑	淡紫	
气虚		淡白/淡红	嫩		白/B				络脉短细淡
~ 虚		淡白	嫩			薄			络脉短细淡
虚热									
实热		红/绛/舌衄		吐弄		厚腻/裂纹		紫红	紫红
表寒								青紫	
表热							干燥		
表虚							润		
表实		淡红/淡白					润		
里寒									淡紫
里热			红点/肿胀/芒刺 + 老	吐弄	B		燥		紫红
里虚							润		
里实									
痰湿		淡紫	胖l	齿痕	白				
寒湿									
湿热	暗	红							
瘀~									
阳l					B				
气l		紫暗/淡红							青紫
热		红							
气陷		淡	嫩 + 软萎			少、薄			
气不固		淡	嫩 + 软萎						
水饮		淡红/淡红							
津亏									络脉短细淡
精亏		枯白							络脉短细淡
食积									
内风									

2.4.2 气味信息分类和处理

闻诊中既包含了气息的语义信息，也包含了气味信息。而分子结构的复杂性，使其难以通过图像或一维矢量进行标识，为体现分子间的关联关系，采用图神经网络可以更好的进行图表征。通过将分子结构的节点矩阵转化为图，可以对于味觉种类进行预测。

味觉信息的收集是目前研究当中的一大障碍，尤其是“远程诊断”这一环节，如何将患者的气味信息远程传输至医生成为了亟待解决的问题。据分析，通过引入特定味觉传感器，首先将气味信息收集并分析出其化学结构是解决该问题的关键。在成功得到其化学结构之后，便可进行化学结构的建模和信息传输。而目前的传感器具有一定的高成本、难以提纯等问题，仍然需要工业界持续探索。



### 2.4.3 声音信息分类和处理

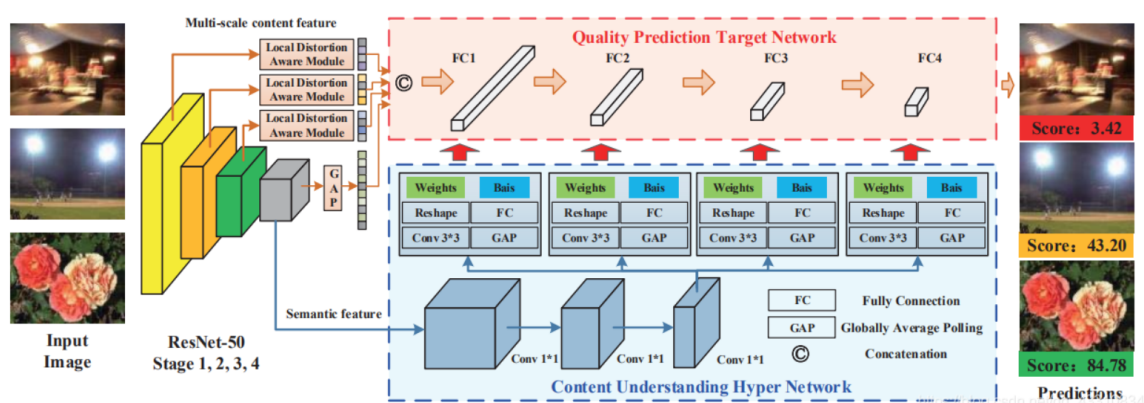
语音识别已经成为自然语言处理领域较为完善的应用之一，通过将语音信号转化成电信号，将气息诊断结果存入数据库中，从而实现气息识别的数据集以供训练，最终得到声音维度的诊断结果。

### 2.4.4 脉象信息分类和处理

脉象信息类似于语义信息，首先也要将其转化为电信号，从而利用矢量化形式进行表征。脉象融合了语义信息提取与气味信息提取的方法，收集脉象信息同样需要传感器，目前市面上已经具有较为成熟的中医把脉传感器。

### 2.4.5 询问既往病史与患者情况信息

对于此类信息，预设问诊信息，利用传统机器学习决策树的方法进行分类和判断，该部分采用智能化处理可能会引起决策树分支较多、难以给出合理标签的缺点，首先可以直接通过医生与患者直接交流的方式，更有助于医生对于患者病症的判断；其次也可通过超网络的思想，即问诊的信息作为多模态网络学习的权重，用网络来更新网络的方法，从而增强多模态学习中的鲁棒性。



图例：超网络概念的迁移

论文来源：Blindly Assess Image Quality in the Wild Guided by A Self-Adaptive Hyper Network

### 2.4.6 信息融合与知识图谱构建

综合上述信息，设计多模态感知的深度神经网络，需要解决三个问题：融合什么？应该融合哪些模态？以及如何以适当方式来标识和处理这些模态。针对以上问题，前两问无论从直观上还是数据可用性程度都便于解答：即融合中医四诊之中视觉、气味、音频、脉象信息。融合方法可借鉴融合transformer方案，通过不同模态信息穿过很多个小的bottleneck，实现模型共享最为重要的信息。

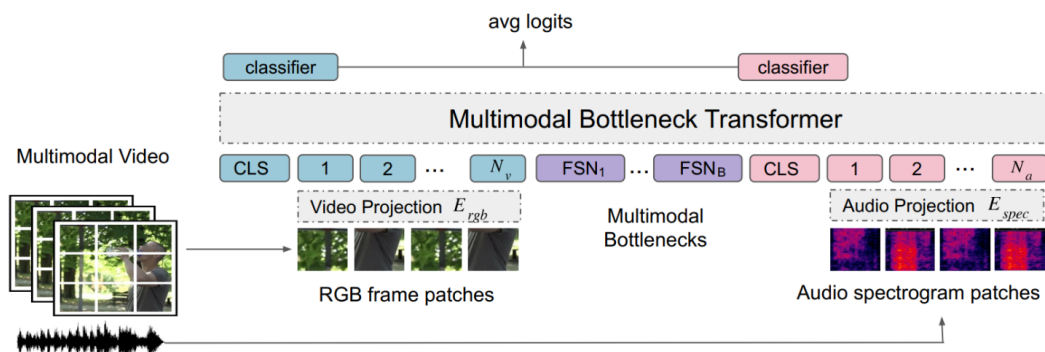


Figure 2: A Multimodal Fusion Transformer applied to audiovisual inputs. The input sequence consists of image and spectrogram patches. These are then projected into tokens and appended to special CLS (classification) and FSN (fusion bottleneck) tokens. Our transformer encoder then uses self attention to model unimodal information, and restricts cross-modal information flow via cross attention with the bottleneck tokens at multiple layers of the network.



论文来源: Attention Bottlenecks for Multimodal Fusion (NIPS 2021)

针对知识图谱构建方案, 最为重要的两点在于信息的不对等与知识图谱对齐。主要思想使用了1.3之中所提到的跨模态存储器, 并利用类别冲突损失作为训练参照。其主要逻辑为

(1)类别首先不会和自身产生冲突, 即 $m_{ii} = 0$

(2)对于本体中明确定义的不相交类别, 其 $m_{ij} = 1$

(3)对于对齐实体所对应的类别也不冲突, 假设 $e_i, e_j$ 为两对齐实体, 则他们所属类别 $m_{ij} = 0$

(4)若不属于以上三种情况, 则根据其在结构树中的距离取关联度, 即

$m_{ij} = 1 - |S(e_i) \cap S(e_j)| / |S(e_i) \cup S(e_j)|$  (类似分割任务中miou的计算方法)

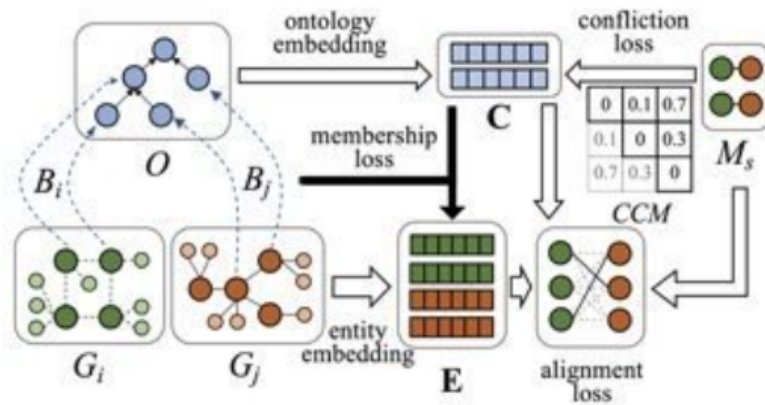


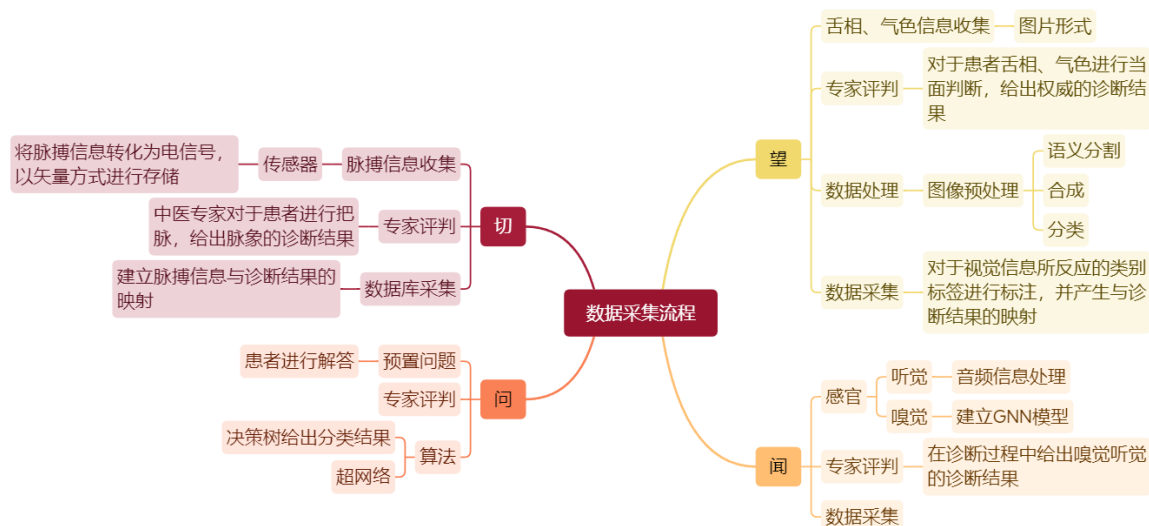
Figure 2: The framework of OntoEA.

### 第三章 功能设计

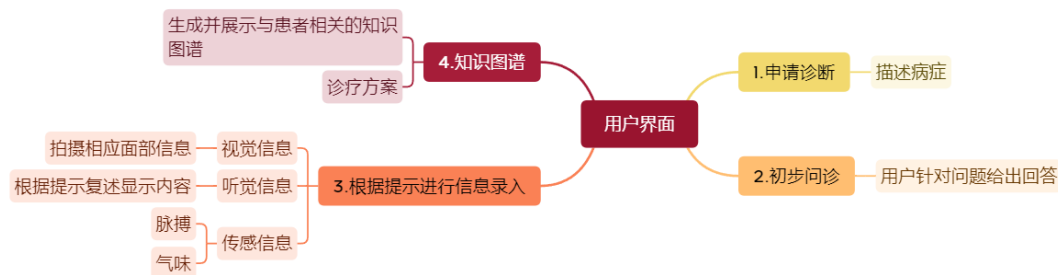
基于上述充分的知识调研和算法分析, 中医临床诊断辅助系统的功能设计包含如下环节

#### 3.1 数据采集阶段

数据采集的流程分为三个阶段: 信息收集-结合专家评判-给出分类结果。利用该流程针对中医内科学的不同场景进行诊断, 从而建立完善的数据集。以便后续多模态网络进行学习。



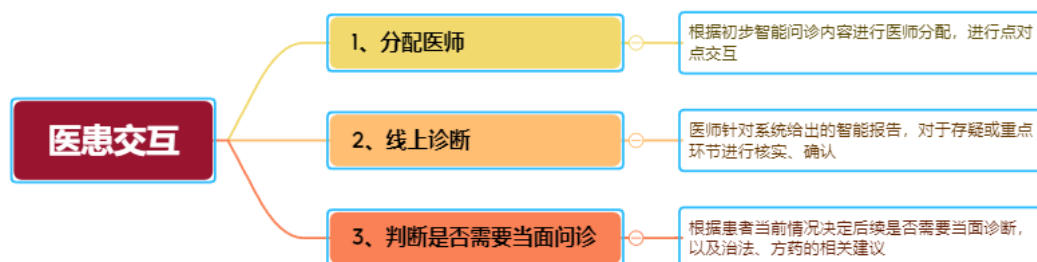
### 3.2 用户界面设计



该部分主要为线上诊断的功能设计，首先用户在系统中申请诊断，患者主动进行病症描述，初步区分病症所属范畴，从而分配该领域的医生。其次，界面显示提前内置的问诊信息，用户填写相关信息。然后根据提示进行信息录入，包括中医诊断中望闻问切的相关内容，将数据传输至数据库中，利用数据信息，给出推荐的诊疗方案，并生成与患者相关的知识图谱内容，从而给出诊断结果。

### 3.3 医患交互设计

在系统给出初步诊断后，中医充当最终的病症核实与确定环节，首先根据系统内容进行医生分配，从而对患者进行点对点诊疗。其次，针对智能诊断当中存疑或重要的问题进行再度核实。最后，根据情况判断当面问诊的必要性，并给出治法、方药的相关建议。



## 第四章 目前遇到的困难

### 4.1 技术难点

针对多模态信息处理的训练过程，由于多模态信息内容不对等、匹配性低等特点，模型具有难收敛、鲁棒性低等特点，针对多模态融合的网络构思是研究者的瓶颈之一。同时，针对知识图谱对齐的问题，中医学的系统更为繁杂，使得对齐过程中会遭到更多信息量、域适应的挑战。

### 4.2 工程难点

针对文中所提到的传感器，脉搏传感器与气味传感器的精确程度、纯度、成本等均存在一定的不足，真正落实到“每家每户”的线上问诊存在高成本的问题。线下临床诊断的辅助性尚能满足。

### 4.3 安全问题

针对用户数据隐私的保护，以及诊疗误判情况的存在，均可能导致严重的安全或伦理问题，针对该类问题，更多的是依靠算法、数据存储的完备性，以及对应法律的建成来进行有效改进。

## 第五章 总结

多模态与知识图谱技术将是近几年最为热门的技术话题，针对中医本体论、系统论的哲学思想，依靠信息融合进行综合诊断是智能诊断大势所趋。通过多模态信息处理的接入可有效减少诊断过程的主观性，通过多学科交叉模式进行定量定性分析，将数理逻辑作为中医诊断处理方法的辅助，必将成为“互联网+医疗”的重要组成模式，辅助病患减少医疗负担，提升生活质量。