

对话系统任务综述与基于 POMDP 的对话系统

李林琳，赵世奇

(注：本文的第一部分主要内容基于“Review of spoken dialogue systems”(López-Cózar et al., 2015); 第二部分翻译自英文原文“POMDP-based statistical spoken dialog systems: A review”(Yong et al., 2013))

摘要：本文第一部分对对话系统的领域研究现状做一个领域综述，第二部分我们详细介绍一下当今对话领域的一个热门课题：基于 POMDP 的统计对话管理系统。首先，我们对对话系统涉及的五种主要技术做一个综合介绍；然后，我们简要讨论一下对话系统的发展历史及其在健康等领域的应用实例。由于对话系统的一个核心组成部分是用户，我们还会简单讨论一下用户建模及其类型；接下来我们讨论一下对话系统的研究趋势，包括语音识别、多模交互、以及对话管理系统的主流研发技术。在第二部分，我们集中介绍基于统计的 POMDP 对话系统所涉及的思想及具体建模方法。

第一部分：对话系统概述

随着人机交互技术的进步，类似“环境智能（Ambient Intelligence）”这类课题更多的强调用户友好性和智能交互性。为了确保自然的智能交互，非常有必要研发安全可靠的交互技术以在人机之间建立沟通的桥梁。本着这个初衷，过去的几十年有很多研究致力于模仿人-人交互的模式来构建人机交互系统，我们称之为对话系统（Spoken Dialogue Systems, SDSs）(Lopez-Cozar & Araki, 2005; McTear, 2004; Pieraccini, 2012)。

最早的对话系统仅仅处理简单的交流任务，比如提供飞行旅行信息(Hempel, 2008)。当今的对话系统已经越来越多的出现在更复杂的任务中，比如，智能环境（Intelligent Environments, Heinroth & Minker, 2013），车载系统（Geutner et al., 2002），个人助手(Janarthanam et al., 2013)，智能家居(Krebber et al., 2004)，及人机交互(Foster et al., 2014) 等。对话系统涉及的另外一个重要课题是用户建模(Andrade et al., 2014)。

本文第一部分我们对对话系统的领域研究现状做一个任务综述，并指出该领域的研究趋势，其中各小节内容如下：1) 对话系统涉及的五种关键技术；2) 对话系统的演进历史以及相关的科研项目；3) 对话系统用户建模；4) 对话系统课题的研究趋势；更智能化、多样化，且具有可移植性以及多模功能。本文的第二部分将讨论一种具体的数据驱动的统计对话系统。

1.1 对话系统涉及的五种主要技术

对话系统通常涉及五个主要的模块：

- 语音识别（ASR）
- 口语理解（SLU）
- 对话管理（DM）
- 自然语言生成（NLG）
- 文本生成语音（TTS）

由于 ASR 模块和 TTS 模块涉及声音信号的处理，不是本文的重点，我们仅限于讨论 SLU, DM 以及 NLG 三个模块。需要说明的是，对话系统还常常有一个单独的模块来存储对话历史。

1.1.1 口语理解（SLU）

口语理解是语音识别模块的下游模块，其任务是获取输入语音信号的语义表示，该语义通常以 frame 的形式表示（Allen, 1995）。而每一个 frame 通常又包含多个槽位，例如，在飞机订票系统中，一个 SLU 的 frame 通常包含如下槽位：

- speechActType
- departureCity
- destinationCity
- destinationCity
- departureDate
- arrivalDate
- airline

一个示例 SLU 模块的输出如下：

- speechActType: 订票
- departureCity: 北京
- destinationCity: 上海
-

SLU 模块的高精度实现是很困难的，因为语言中常常带有歧义、人称代词、指示代词、以及省略成分。该模块的实现有两种主流的方式，要么借助于句法分析，要么利用统计机器学习，也可以是两者结合（Griol et al., 2014）。SLU 模块

也常常借助于对话历史模块的信息，其目的是搜索用户最近的对话历史找到当前 query 中没有显式提到的信息以更好的理解对话。

同时，口语理解系统还应该具有较好的鲁棒性来处理语音识别的错误和噪音。相关的技术有，放宽语法检查尺度，仅关注关键词；或对识别的句子进行部分分析，提取关键成分。很多统计方法在这方面有所应用，如(Lemon & Pietquin, 2012)等。

1.1.2 对话管理（DM）

口语理解的下游模块是对话管理器 DM。该模块的任务是决定系统如何回复给定的用户输入 (McTear, 2004)，例如，向用户提供信息，向用户发出询问以确认系统理解，或者请求用户换一种句型表达需求。在上述飞机订票的对话系统实例中，DM 可能会决定触发一个询问请求让用户来确认“出发城市”这个槽位的值是不是“北京”，其触发原因可能是语音识别系统对“北京”这个词的置信度值过低。

为了给用户提供信息，对话管理器通常需要查询数据库或者互联网，同时它还需要考虑对话历史模块提供的信息。例如，DM 模块可能根据对话历史上下文发出询问请求以补充缺失数据。对话管理模块模式可以有多种：rule-based, plan-based 和基于增强学习的 (Frampton & Lemon, 2009)。本文的第二部分将详细讨论数据驱动的对话系统 POMDP。

1.1.3 自然语言生成（NLG）

对话管理器的决策会输出给下游模块进行自然语言生成 (NLG)。由于对话管理器的输出是抽象表达，我们的目标是将其转换为句法和语义上合法的自然语言句子，同时考虑对话上下文的连贯性 (Lemon, 2011)。许多主流的自然语言生成系统是基于模板的，其根据句子的类型制定相应的模板 (Baptist & Seneff, 2000)。这些模版的某些成分是固定的，而另一些成分需要根据对话管理器的输出结果进行填充。例如，下述模板可以用来生成句子以表示两个城市间的航班。

我查到了 *<flightAmount>* 趟 *<departureDate>* 号 *<departureCity>* 到 *<destinationCity>* 的航班。

在该模板中，尖括号部分需要根据对话管理器的输出值进行填充，填充后该 NLG 模块输出为：我查到了 30 趟 20 号从北京到上海的航班。

考虑到对话系统的连贯性, NLG 在生成语句的时候通常需要根据对话上下文综合考虑。相关技术涉及如何省略表达, 即, 省略前文中已经提到概念, 或者使用代词指代前文已经出现的名词。该过程依赖对话历史模块的存储信息。NLG 模块还需要过滤冗余重复信息, 该过程又被称为句子聚合 (Dalianis, 1999)。除了基于模板的 NLG 技术以外, 最近也有一些基于统计的方式进行自然语言生产, 如 (Dethlefs et al., 2013; Rieser et al., 2014)。由于篇幅限制, 本文不对相关技术一一介绍。

1.2 对话系统的发展历程

对话系统的历史可以追溯到 18、19 世纪, 最早是利用机械方式制造机器人来模仿人类行为, 该学科的发展具有较长的演进史。

1.2.1 早期系统以及科研项目

20 世纪初, Stewart (1922) 发明了用电能产生声音的机器。20 世纪 40 年代出现了第一台计算机, 随后图灵提出了智能计算机的概念 (Turing, 1950)。60 年代第一次出现了处理自然语言的计算机系统, 例如, ELIZA (Weizenbaum, 1966) 基于关键词提取以及预定义模板将用户输入转换为系统回答。而随着语音识别、自然语言处理和语音合成等技术的发展, 80 年代第一次出现了对话系统, 两个代表性项目是: 美国的 DARPA 口语系统和欧洲的 Esprit SUNDIAL 系统。在这两个对话系统之后, MIT 和 CMU 相继进行了一系列对话系统的研究。90 年代, DARPA Communicator 项目由美国政府资助进行语音技术研发。到如今, SDSs 的研发已涉及更广的技术, 包括智能推理、多模交互以及跨语种交互等等 (Heinroth & Minker, 2013)。

1.2.2 应用实例

SDSs 有着广泛的应用实例, 包括自动旅游旅行信息系统 (Glass et al., 1995), 天气预报系统 (Zue et al., 2000), 银行系统 (Hardy et al., 2006; Melin et al., 2001) 和会议协作系统 (Andreani et al., 2006) 等。我们仅就其在健康领域和 embodied agent 领域的应用做一个简要介绍。

健康领域：

SDSs 在医疗应用中能协助诊疗病人，例如，Bickmore & Giorgino (2006) 研发了对话诊断系统，该系统用对话方式和病人交流，诊断疾病。该研究对于不方便现场就诊的病人具有现实意义，同时有助于处理个人隐私和敏感信息问题。一些病人无法很好的当面与医护工作者交流，第一，因为诊疗时间非常有限；第二，病人有可能忌讳一些非常隐私的问题（比如酗酒、抑郁、艾滋病等）。但是，智能对话系统就能有效的解决这些问题 (Ahmad et al., 2009)。

过去的 20 年中，SDSs 已广泛的应用在医疗领域，如医疗咨询 (Ghanem et al., 2005; Hubal & Day, 2006; Pfeifer & Bickmore, 2010)，慢性病监测 (Black et al., 2005)，辅助开药(Bickmore et al., 2010)，饮食引导 (Delichatsios et al., 2001)，帮助戒烟 (Ramelson et al., 1999)，及医疗诊断 (Maglogiannis et al., 2009)。

Embodied Agent:

SDSs 另一个常用场景是 Embodied Agent. 该领域相关的系统原型有：COLLAGEN (Rich & Sidner, 1998), AVATALK (Hubal & Day, 2006) , COMIC (Catizone et al., 2003)，以及智能对话系统 NICE (Corradini et al., 2004)。

1.3 用户建模

用户建模是 SDSs 的一个重要课题，其过程不仅仅局限于在对话系统设计阶段建设静态的 profile，也可以根据用户状态建立动态 profile。SDSs 可以根据用户的语音信息，说话状态及情景建立大量的用户 profile 数据，然后利用这些数据理解用户对话，并预测用户行为。更具体来说，用户建模涉及情感建模 (Balahur et al., 2014; Schuller & Batliner, 2013; Moors et al., 2013)，人物性格建模 (Nass & Yen, 2012)，和上下文场景建模 (Zhu & Sheng, 2011) 等主要技术。

1.4 研究趋势

语言是人类的特殊能力，SDSs 是 AI 领域的一个高难度题目，因为它涉及多个语言相关的子领域，如语音识别、语音合成、语言理解、语义表示、对话管理、语言生产、情感建模、以及多模交互。Grand View Research 公司做的一项研究调查报告显示 2012 年对话系统的市场市值大约为 3.5 亿美金，该公司预测到 2020 年会增长 31.7%。同时该项报告还指出，考虑到市场以及经济效应，人工

客服领域将是未来 SDSs 的一个大方向。下面我们讨论一下 SDSs 主要研究方向。

1.4.1 语音识别

对话系统一个重要的组件是语音识别，其输出错误是 SDSs 所面临的第一难题，未来如何降低语音识别噪音仍是一个大方向。其相关的技术有：降低背景噪音，预测用户输入，以及 ASR 系统情感识别 (Batliner et al., 2010)。不同于传统的基于 HMM 的语音识别，新趋势越来越多的依赖深度学习 (Dahl et al., 2011)。考虑到语音识别错误可能直接导致下游的对话管理器无法正确理解语义，降低语音识别噪音任重道远。

1.4.2 多模交互

口语理解是 SDSs 另一个重要组成部分，其输入不仅仅依赖于语音，还依赖于用户表情、动作等多模信号 (Bui, 2006; Lopez-Cozar, 2005)。该方面的研究很多，如欧盟的 Horizon 2020 项目的一个子领域是语言技术 (ICT-22-2014)，其致力于多模计算机交互。SDSs 多模技术还包括语义多模融合 (Russ et al., 2005)，以及借助多模信号降低语音识别模块噪音 (Longe et al., 2012)。

1.4.3 对话管理

对话管理器是协调人机交互的中心模块，它是一个相对成熟的研究课题，大致可分为如下四种类型 (Jurafsky & Martin, 2009)：

- 基于有限状态和对话语法
- 基于 Frame 的系统
- 信息状态更新 (ISU)
- 马尔可夫决策过程 (MDPs) 和部分可观测的马尔可夫过程 (POMDPs)

有限状态模型认为对话是一个状态转移序列图，图中每一个结点表示隐含的对话状态，对应于系统行为（如回答、询问、确认、等等），结点之间的状态转移控制对话流。Nuance 的自动银行系统就是根据这种方法设计的 (McTear, 2002)。有限状态模型常应用于自助语音服务系统，其优点是简单易实现，缺点是缺乏灵活性，难以处理复杂对话逻辑。

Frame-based 系统在前面简单介绍过，其基本思想是填槽位。该方法可以在当前对话轮中填一个或者多个槽位，也可以覆写或修正前面对话轮的填充内容。基于 Frame 的对话管理系统还有一些衍生系统，如 agenda (Bohus & Rudnicky, 2003)，task structure graphs，和 type hierarchies and blackboards (Rothkrantz et al., 2004)等。

ISU 方式利用“信息状态”将对话过程的所有可用信息进行建模 (Larsson & Traum, 2000)，即整合对话参与者的所有信息，而后建模对话行为。

以上三种对话管理模式都需要计算语言学专家设计并编写对话方案，该策略会增加对话系统的设计开发成本，同时也会降低系统的可维护性。为了克服这些局限性，近来出现了基于机器学习的对话管理系统，典型的代表是 MDP 和 POMDP。这些系统的基本思想是利用统计框架从大量的对话语料中自动学习对话管理模型 (Young et al., 2013)。这种方式有两个主要的优点：第一，可以将不确定性表示引入到模型中，相对基于规则的系统，其对语音和语义理解的噪音有更好的鲁棒性。第二，这种框架具有自动学习功能，可以极大的降低人工开发成本。当然，这种方法也存在缺点，首先我们需要收集大量的对话数据，然后还需要对这些数据进行标注。在本文第二部分，我们集中讨论基于 POMDPs 的统计对话管理系统。

第二部分：基于 POMDP 的统计对话管理系统

统计对话管理系统是一种数据驱动的方法，无需人工构建对话管理机制，具有较好的鲁棒性。部分可观察的马尔可夫决策过程 (POMDPs) 具有良好的数据驱动性，但问题是完整的建模和优化计算代价巨大，甚至不可实现。在实际过程中，需对 POMDP-based 的系统近似优化求解，以下内容将对 POMDP-based 的系统的领域现状做一个综合介绍。

2.1 概述

传统的 SDS 的主要组成部分如下图 1 所示。口语理解模块 (SLU) 将语言转换成抽象语义表示，即用户对话行为 u_t ，而后系统更新其内部状态 s_t ，然后系统通过决策规则 $a_t = \pi(s_t)$ 确定系统行为，最后语言生成模块 (NLG) 将系统行为 a_t 转化为自然语言字符串。其中，状态变量 s_t 包含跟踪对话过程的变量，以及表示

用户需求的属性值（又称为 *slots*）。在传统对话系统中，决策规则是通过流程图的方式实现的，图中的结点表示状态和行为，而边则表示用户输入（Oshry et al., 2009; Paek & Pieraccini, 2008）。

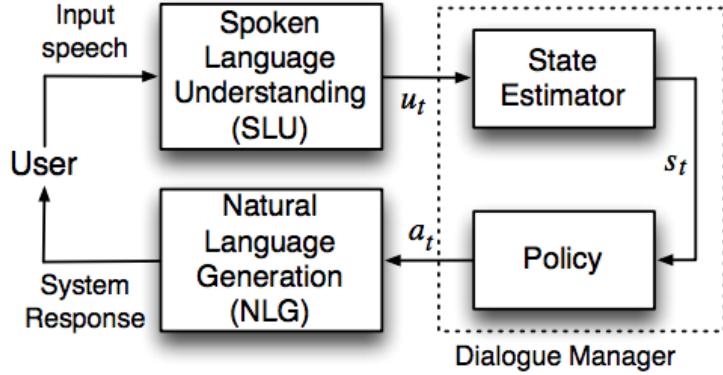


图 1. 对话系统的主要组成部分

尽管语音识别技术在过去的几十年不断进步，但是在噪音环境中（如公共场所或者汽车内）的语音识别错误率仍高达 15%-30% (Lippmann, 1997; Black et al., 2011)。对话处理机制必须要处理噪音问题，包括自动检错和恢复。正因为如此，传统的基于流程图的对话管理系统非常难以制定和维护。

新型的对话管理系统基于部分可观察的马尔可夫决策过程 (POMDPs) (Williams & Young, 2007)，该方法假定对话过程是马尔可夫决策过程，也就是说，对话初始状态是 s_0 ，每一个后续状态用转移概率来表示： $p(s_t|s_{t-1}, a_{t-1})$ 。状态变量 s_t 是无法直接观察到的，它代表了对用户需求理解的不确定程度。系统把 SLU 的输出看作是一个带噪音的基于用户输入的观察值，这个观察值的概率为 $p(o_t|s_t)$ ，这里的转移概率和生成概率用恰当的随机统计模型表示，又称为对话模型 M，而每个步骤中采取哪个行动则由另一个随机模型控制，该模型称之为对话策略 P。在对话过程中，每一步还需要一个回报函数来体现理想中的对话系统特性。对话模型 M 和对话策略 P 的优化是通过最大化回报函数的期望来实现的，该过程可以通过直接用户交互在线训练，也可以利用离线的语料库训练。详见下图 2。

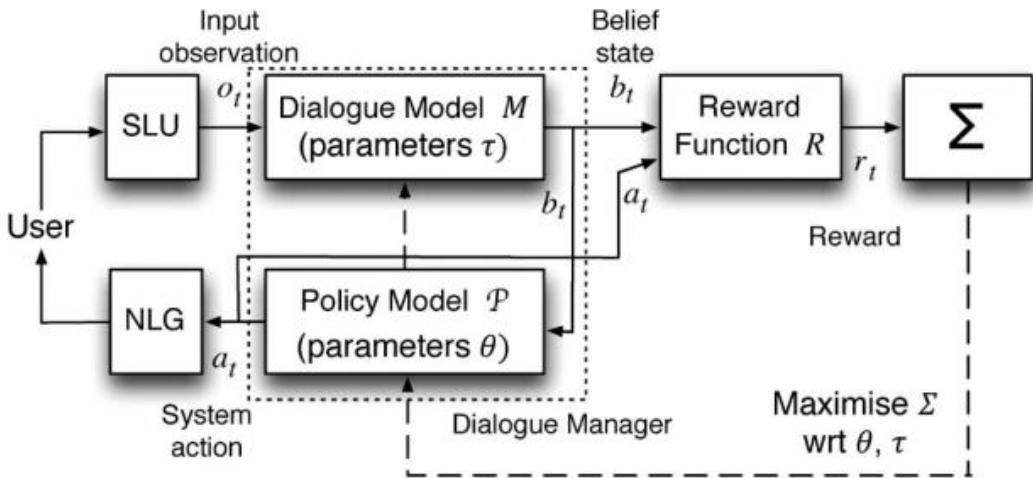


图 2. 基于 POMDP 的对话系统

基于 POMDP 的对话系统融合了两个核心观点：置信状态跟踪和增强学习。这两个方面可以放在同一个框架下学习。与传统方式相比，该方法具有如下优点：

- 1) 置信状态为语音识别噪声提供了更好的鲁棒性 (Williams & Yong, 2007)。置信状态在用户输入后的后验概率可以借助于一种称为“置信监督 (belief monitoring)”的贝叶斯推理过程更新。在设计置信状态的过程中，可以借助模型先验概率去捕捉用户行为，而借助推理过程去探索所有的识别假设空间，如模糊网络和 N-best lists。其特点是，融合多轮证据以降低单个错误的影响。与传统方法不同，用户的反复行为是得到激励的，如果用户足够多次的重复一种表达，系统对他们所说的内容的置信度也会随之增加（前提是正确的假设候选出现在 N-best lists 中）。
- 2) 通过保存各个状态的置信分布，系统可以并行的追踪各种对话路径，它不是贪婪的选择当前最优解而是综合考虑各种状态的全局解。当用户输入一个负反馈信号时，当前最可能解的概率被降低，焦点会聚集到另外一个状态。因此，不需要回溯或者修改对话机制。强大的对话策略可以简单的嵌入置信状态到对话行为的映射中。
- 3) 显式的表达状态和行为能将回报函数和状态行为对关联起来。其回报综合值组成了对话效果的客观衡量标准，因此可以用离线语料库或者在线用户互动方式，借助增强学习提升效果。该方法具有最优的决策策略，避免了人工调优的劳动，可以容纳复杂的规划机制。

然而在实践中运用 POMDP 并不容易，有许多实际问题需要解决。SDS 的状态行为空间巨大，求解这个空间需要复杂的算法和软件。实时的贝叶斯推理也非常难，完整的 POMDP 的学习策略是不可实现的，因此必须利用近似法求解。优化基于 POMDP 的 SDS 的最直接方式是通过直接用户对话。但是，通常难以找到足够数量的用户帮助训练系统，所以实践中常常通过用户模仿器的方式来对参数模型进行优化。

2.2 部分可观察的马尔可夫决策过程

部分可观察的马尔可夫决策过程用一个多元组 $(S, A, T, R, O, Z, \gamma, b_0)$ 表示，其中 S 是状态集； A 是行为集合； T 表示转移概率 $P(s_t|s_{t-1}, a_{t-1})$ ； R 是回报的期望值； O 是观测值集合； Z 代表观测概率 $P(o_t|s_t, a_{t-1})$ ； γ 是几何衰减系数，其值在 0-1 之间； b_0 是置信状态的初始值。

POMDP 的过程如下：在每一个过程中，真实世界是一个无法观察的状态 s_t 。因为 s_t 是未知的，变量置信状态 b_t 表示所有可能状态的分布， $b_t(s_t)$ 表示处在某个特定状态 s_t 的概率。系统基于 b_t 选择行为 a_t ，得到一个激励值 r_t ，然后转化到状态 s_{t+1} ，这里 s_{t+1} 仅仅依赖于 s_t 和 a_t 。然后系统得到一个观察值 o_{t+1} ，该值依赖于 s_{t+1} 和 a_t 。这个过程如图 3 所示：

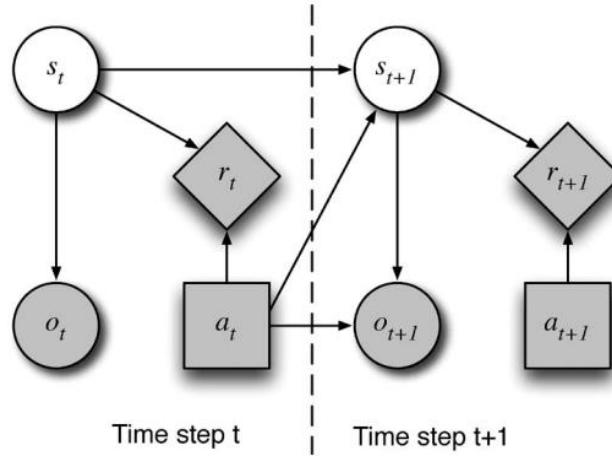


图 3. 用 influence diagram 表示 POMDP；圆圈表示隐变量，带阴影的圆圈表示观测值，方块表示系统行为，棱形表示回报值，箭头表示因果关系。

给定置信状态 b_t ，最近一次的系统行为 a_t 以及观察值 o_{t+1} ，新的置信状态 b_{t+1} 的更新可以表示如下 (Kaelbling et al., 1998)：

$$b_{t+1}(s_{t+1}) = \eta P(o_{t+1}|s_{t+1}, a_t) \sum_{s_t} P(s_{t+1}|s_t, a_t) b_t(s_t)$$

这里 $\eta = P(o_{t+1}|b_t, a_t)$ 是正规化常量， b_0 是系统没有任何行动之前的初始置信状态分布。

系统行为由策略 π 控制。最通用的策略是直接将置信状态和行为进行直接映射 $\pi(b) \in A$ ，或者通过一个概率函数进行对应 $\pi(a|b) \in [0,1]$ ，这里 $\pi(a|b)$ 是在置信状态 b 下采取行动 a 的概率，其满足 $\sum_a \pi(a|b) = 1$ 。

以置信状态 b_t 为开始的策略 π 的综合回报函数定义如下：

$$V^\pi(b_t) = r(b_t, \pi(b_t)) + \gamma \sum_{o_{t+1}} P(o_{t+1}|b_t, \pi(b_t)) V^\pi(b_{t+1})$$

该公式对应确定性策略 (deterministic policy)；

$$V^\pi(b_t) = \sum_{a_t} \pi(a_t|b_t) \left\{ r(b_t, a_t) + \gamma \sum_{o_{t+1}} P(o_{t+1}|b_t, a_t) V^\pi(b_{t+1}) \right\}$$

该公式对应随机策略 (stochastic policy)。

最佳对话策略 π^* 通过优化回报函数得到：

$$V^*(b_t) = \max_{a_t} \left[r(b_t, a_t) + \gamma \sum_{o_{t+1}} P(o_{t+1}|b_t, a_t) V^*(b_{t+1}) \right]$$

该策略也称为 Bellman 优化公式 (Bellman, 1957)。在 POMDP 参考文献中，寻找最优策略 π^* 的过程称之为“求解”和“优化”过程。Kaelbling et al. (1998) 应用了精确求解方案，(Pineau et al., 2003; Smith & Simmons, 2004) 提出了近似求解方案。但问题是通用的 POMDP 方法复杂度高，难以大规模的应用到实用对话系统。即使中小型规模，其涉及的状态、行为、和观察值很容易达到 10^{10} 量级。穷举 $P(s_{t+1}|s_t, a_t)$ 是不可实现的，因此，直接优化更新置信状态优化回报函数并不可行。通常情况下，我们需要简化模型近似求解。接下来我们将详细讨论。

2.3 置信状态表示和监测

本小节集中讨论图 2 中对话系统模型 M 。实用 SDS 中，状态必须包含三种不同类型的信息：用户的目标 g_t ，用户的真实意图 u_t ，以及对话历史 h_t (Williams &

Young, 2007)。用户目标包含需要完成任务所有信息，用户真实意图是指用户实际想表达的意图而非系统识别出的意图，对话历史跟踪之前的对话流。由此，对话中的一个状态包含三个因子：

$$s_t = (g_t, u_t, h_t).$$

引入条件独立性假设以后，该过程可以表示为图 4。将状态分解成以上三个因子可以对状态转移矩阵进行降维，同时也减少了系统的条件依赖性。

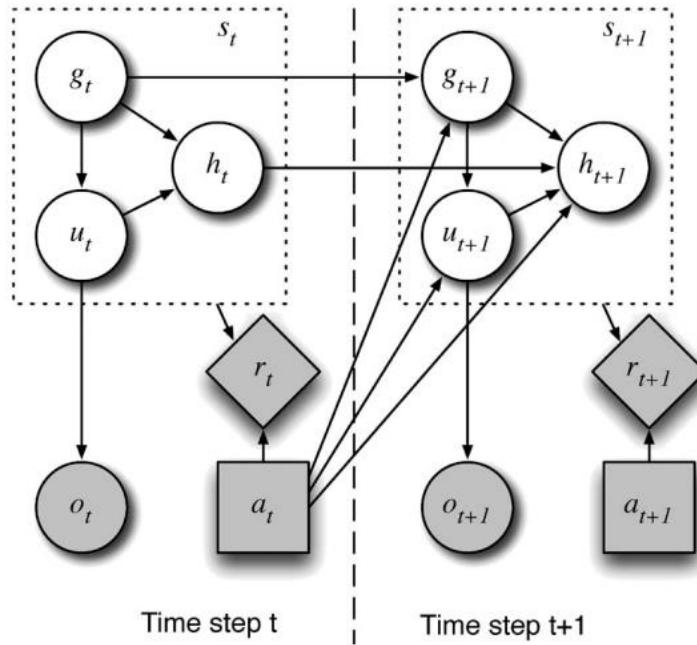


图 4 表示 SDS-POMDP 各个状态因子的 influence diagram

结合置信更新和状态因子两公式，SDS 的更新策略可以表示为：

$$\begin{aligned} b_{t+1}(g_{t+1}, u_{t+1}, h_{t+1}) &= \eta P(o_{t+1}|u_{t+1}) \\ &\quad \cdot P(u_{t+1}|g_{t+1}, a_t) \\ &\quad \cdot \sum_{g_t} P(g_{t+1}|g_t, a_t) \\ &\quad \cdot \sum_{h_t} P(h_{t+1}|g_{t+1}, u_{t+1}, h_t, a_t) \\ &\quad \cdot b_t(g_t, h_t) \end{aligned}$$

以上公式包含了对话系统的四个要素：

(a) 观察模型表示给定用户真实表达 u ，观察值 o 的概率。它包含了语音识别系

统的错误率。

- (b) 用户模型表示在给定系统前一轮输出和当前系统状态下，用户真实表达 u 的概率。它建模了用户行为。
- (c) 目标转移模型表现了用户目标转换的可能性。
- (d) 历史模型：系统记忆的对话历史。

虽然状态因子模型极大的简化了 POMDP 模型的复杂度，但是它仍旧复杂，难以在实际的系统中应用。因此还需要进一步近似化处理，通常有两种常用技术：

- 1) N-best 方法，包括剪枝和重组 (Gasic & Yong, 2011)
- 2) 贝叶斯网络法 (Thomson & Yong, 2010)

2.4 策略模型和增强学习

策略模型 P 提供了置信状态 b 和系统行为 a 的映射。我们的目标是寻找一个最优的策略最大化对话回报函数的综合期望。

POMDP 的置信空间是一个高维空间。置信空间中临近的点必须具有相同的行为值，因此，有必要用一种策略将置信空间的点进行区域划分，确保同区域的点具有相同的行为值。当然，该行为映射必须保证让每一个区域都有最佳的行为值。POMDP 系统的策略模型可以通过压缩置信空间 (Crook & Lemon, 2011) 或者动态状态赋值 (Doshi & Roy, 2008) 实现。现实应用中，我们通常需要对策略进行简化表示以降低推理复杂度。

我们可以根据实际应用加一些限定条件。首先，通常的对话系统仅仅用到相对小的置信空间；第二，可行行为的范围在给定的置信空间内通常是限定的。由此，这里有一个精简的特征空间，我们称之为 *summary space*。其中，状态和行为都被简化以方便策略表示和优化 (Williams & Yong, 2005; Williams & Yong, 2007)。*Summary space* 是全局问题空间的一个子空间，*belief tracking* 在整个问题空间中求解，而决策执行和策略优化仅在 *summary space* 中运行。问题空间和 *summary space* 中的运行方案如下：在置信更新后，问题空间中的置信状态 b 被映射为一个特征向量 \hat{b} 和候选行为集合 $\{\hat{a}\}$ 。策略从一系列的候选行为集合中选取，然后将 *summary space* 中的 \hat{a} 映射回原问题空间中一个完整的行为 a 。

Summary space 的转换需要两个组成部分：问题空间中选择候选行为的机制，以及从置信状态和候选行为中抽取特征的函数。最简单的选取候选行为的方法是将可以对应到概念以及槽位（如：地点类型、食物类型、星级评价、等等）的对

话行为（如：问候、问题、肯定、告知、等等）都包括在内，然后通过置信度对各个槽位赋值（Thomson & Young, 2010; Williams & Young, 2005）。该方法是全自动的，但它可能会包含一些错误的候选行为，如在对话的正中间进行问候，或者在没有被问到某个值的时候，对该值进行确认。另外一类选择候选行为的方法是 partial program (Andre & Russell, 2002; Williams, 2008)，或者马尔可夫逻辑网络 (Lison, 2010)。这类方法可以在对话流中任意容纳人工知识，同时显式设定业务逻辑，比如，在进行资金转账之前要求输入用户密码。同时，限定候选行为空间能更快的收敛到最优策略，因为它已经将很多的错误候选排除在外 (Williams, 2008)。但是，这些规则需要人工制定，而人工规则可能错误的将最优候选行为排除在外。作为一种折衷方案，有些系统允许每一个对话行为作为候选，但通过人工规则限定对话行为槽位 (Yong et al., 2010)。

Summary space 转换的第二个重要组成部分是从置信状态以及候选行为中抽取特征的函数。我们通常为每一个对话行为构建一个二进制特征，也可以针对每一个有效的行动 / 槽位对，例如 `confirm(food)`。这种方式通常会产生 20-30 维的行为特征，其每一维表示一个唯一行为。状态特征通常是异构的，包含实数值，二进制值，和枚举类值。典型的状态特征包括：最重要的 N 个用户目标的置信度；各个槽位的 marginal belief；最重要的用户目标属性（如，匹配到的数据库实体的数量）；表示哪些系统行为可行的隐变量；对话历史属性值（如，是否已确认最重要的用户目标）；历史用户行为；或者上述特征的组合 (Thomson & Yong, 2010; Williams & Yong, 2005; Williams, 2008; Yong et al., 2010)。通常系统具有 5 到 25 个特征，这些特征一般是手工选取的，也有些研究探索自动特征选取 (Williams & Balakrishnan, 2009)。状态特征不仅仅局限于置信状态信息，也可以在置信状态以外寻找特征，如数据库中的信息，过去的对话历史，或者上下文信息。

给定一个特定的 summary space，策略可以用表示为一个显示的映射 $\pi(\hat{b}) \rightarrow \hat{a}$ ，或者表示为一个概率分布 $\pi(\hat{b}, \hat{a}) = p(\hat{a}|\hat{b})$ ，后者行为的选取是通过对概率分布的抽样实现的。策略是 summary belief state (状态) 及 action (行为) 的函数，而不是原始问题空间的置信状态和行为。可以认为新函数是对原问题空间的近似，也可以认为是另一个马尔可夫过程，在这个过程中状态和行为是 summary state 和 summary action。

在显式的映射策略中，最常用的方法是寻找一个行为 a 以最大化 Q 函数（回报函数的数学期望），即：

$$\pi^*(\hat{b}) = \arg \max_{\hat{a}} \{Q^*(\hat{b}, \hat{a})\}.$$

这里的 Q 函数可以是带参数的，也可以是无参的。如果不带参数，置信状态会在一个离散的编码集 $\{\hat{b}_i\}$ 中，对任意一个状态 \hat{b}_i 都可以计算出其 Q 值。

有五种常用的方法来优化策略：1) planning under uncertainty; 2) value iteration; 3) Monte-Carlo 优化；4) 最小平方策略迭代 (LSPI); 5) natural actor-critic (NAC)。这五种方式常常应用在 end-to-end 的对话系统中。除此之外，还有 Q-learning (Scheffler & Young, 2002) 和 SARSA (Henderson et al., 2008) 等方法。由于篇幅限制，我们不对这些策略做一一介绍，感兴趣的读者可以阅读相关参考文献。

总的来说，planning under uncertainty 策略将置信状态作为概率分布，而 value iteration 和 Monte-Carlo 优化需要首先将置信空间量化，LSPI 和 NAC 基于置信状态特征的线性模型进行函数近似。

2.5 用户模拟器

直接从语料库中学习对话策略存在很多问题，比如收集数据中的状态空间可能与策略优化数据不同。另外，这种方式无法在线互动学习。因此我们可以构建一个用户模拟器，让这个模拟器与对话系统直接进行互动。用户模拟器不仅仅可以用来学习对话系统，也可以用它来评估对话系统 (Schatzmann et al., 2006)。用户模拟器通常运行在抽象对话行为层。给定一系列的用户行为和系统回复，其目标是对用户回复的概率分布进行有效的建模：

$$p(u_t | a_t, u_{t-1}, a_{t-1}, u_{t-2}, \dots)$$

正如前文提到的，在实际系统中对话管理系统的观测值被噪音变量影响，因此用户回复同时被用户模型和噪音模型控制。模型 $p(u_t | \dots)$ 匹配在语料中的用户回复，另外需要 error model 建模语音识别及理解错误 (Hastie, 2012; Pietquin & Hastie, 2012; Schatzmann et al., 2005; Williams, 2008)。

2.6 系统和应用

前面的几个小节讲述了统计对话系统的几个主要模块，其相关的技术在随着时间进步完善。尽管在商业上推广这些技术有一定的难度，但是在具体的场景中

已有一定范围应用。这里我们简单的提几个基于 POMDP 框架的对话系统。

这些系统大多都是非正式的 inquiry 系统，包括语音呼叫 (Janarthanam et al., 2011)，旅游信息 (Thomson & Yong, 2010)，日程安排 (Kim & Lee, 2007) 和汽车导航 (Kim et al., 2008) 等。POMDP 也可应用于基于命令控制的系统，如通过多模接口控制家电 (Williams, 2007)。

POMDP 曾在 CMU 举办的“Let’s Go”竞赛任务中被应用，其为 Pittsburgh 区域的居民播报非忙时段的公交车信息 (Thomson et al., 2010)。在该应用中，用户可能从多种不同的手机装置来电，且通话环境通常有噪音，结果显示，基于 POMDP 的系统明显优于传统系统 (Black et al., 2011)。

2.7 小结

统计对话系统提出数据驱动的框架，该方式可以有效的降低人工编写复杂对话管理规则的开销，同时对在噪音环境中的语音识别错误具有良好的鲁棒性。通过一个显式的对不确定性建模的贝叶斯模型和一种回报驱动的策略优化机制，POMDP 建立了一个良好的对话系统框架。

但是，基于 POMDP 的对话系统非常复杂，通常需近似求解。这里有许多实际的研究问题，例如，如何在保持模型的复杂度的同时，让置信状态序列可解？如何减少策略学习的迭代次数，从而能在真实用户上训练模型而不是借助用户模拟器？同时，还需要将对话系统打包以及平台化以使得非专家也能使用这些技术。

除此之外，POMDP 框架严重依赖回报函数。原则上来说，这是对话系统的优点，因为它可以提供一个客观的目标机制定义对话系统的设计标准。但是，实际应用中，我们很难直接从用户那里获得可信的回报信号，即使是最简单的成功 / 失败也很难获得，比如，在被问到“系统是否回答了你所问的问题？”时，许多用户出于礼貌会直接说“是”，或者由于对系统不切实际的过高期望而直接回答“否”。回报函数可以基于用户满意度来获取，而该满意度值可以通过客观的可以衡量的特征进行回归计算获取，该方法在 PARADISE 系统 (Levin et al., 1997) 上有所应用，也有其他的研究 (Singh et al., 1999) 在这个思路框架下进行。尽管如此，很多经验告诉我们，根据真实用户反馈的在线学习策略必须要结合成熟的生物识别技术，用客观标准测量用户情感满意度。

总结

本文对对话系统的领域研究现状做了一个整体介绍。对话系统的核心问题是处理多轮交互，让机之间的互动高效、自然、智能。在本文中我们涉及了对话系统的主要任务模块，并做了简单的概述，同时指出了对话系统所面临的问题和挑战。我们还介绍了对话系统的演进历史及其应用实例，并从口语交流、多模交互和对话管理方面介绍了该领域的研究趋势。接下来，本文详细介绍了 POMDP 统计对话管理器的相关技术及领域现状，同时也指出了其中的问题及面临的挑战。

参考文献

- López-Cózar, & R., Araki, M. (2005). Spoken, multilingual and multimodal dialogue systems: Development and assessment. John Wiley
- López-Cózar, R., Callejas, Z., Griol, D., & Quesada, J. F. (2015). Review of spoken dialogue systems. Loquens, 1(2), e012.
- McTear, M. F. (2002). Spoken dialogue technology: Enabling the conversational user interface. ACM Computing Surveys, 34(1), 90–169. <http://dx.doi.org/10.1145/505282.505285>
- McTear, M. F. (2004). Spoken dialogue technology. Toward the conversational user interface. Springer. <http://dx.doi.org/10.1007/978-0-85729-414-2>
- Pieraccini, R. (2012). The voice in the machine: Building computers that understand speech. Cambridge, MA: MIT Press.
- Hempel, T. (2008). Usability of speech dialogue systems: Listening to the target audience. Springer.
- Heinroth, T., & Minker, W. (2013). Introducing spoken dialogue systems into Intelligent Environments. New York: Springer. <http://dx.doi.org/10.1007/978-1-4614-5383-3>
- Geutner, P., Steffens, F., & Manstetten, D. (2002). Design of the VICO spoken dialogue system: Evaluation of user expectations by Wizard-of-Oz experiments. Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC '02), Canary Islands.
- Janarthanam, S., Lemon, O., Liu, X., Bartie, P., Mackaness, W., & Dalmas, T. (2013). A multithreaded conversational interface for pedestrian navigation and question answering. Proceedings of the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), 151–153.
- Krebber, J., Möller, S., Pegam, R., Jekosch, U., Melichar, M., & Rajman, M. (2004). Wizard-of-Oz tests for a dialog system in smart homes. Paper presented at the Joint Congress CFA/DAGA '04, Strasbourg.
- Foster, M. E., Giuliani, M., & Isard, A. (2014). Task-based evaluation of context-sensitive referring expressions in human-robot dialogue. *Language, Cognition and Neuroscience*, 29(8), 1018–1034. <http://dx.doi.org/10.1080/01690965.2013.855802>
- Andrade, A. O., Pereira, A. A., Walter, S., Almeida, R., Loureiro, R., Compagna, D., & Kyberd, P. J. (2014). Bridging the gap between robotic technology and health care. *Biomedical Signal Processing and Control*, 10, 65–78. <http://dx.doi.org/10.1016/j.bspc.2013.12.009>
- Allen, J. (1995). Natural language understanding. Redwood City, CA: The Benjamin Cummings.
- Callejas, Z., Griol, D., Engelbrecht, K.-P., & López-Cózar, R. (2014). A clustering approach to assess real user profiles in spoken dialogue systems. In J. Mariani, S. Rosset, M. Garnier-Rizet & L. Devillers (Eds.), *Natural interaction with robots, knowbots and smartphones* (pp. 327–334). New York: Springer. http://dx.doi.org/10.1007/978-1-4614-8280-2_29
- Griol, D., Callejas, Z., López-Cózar, R., & Riccardi, G. (2014). A domain-independent statistical methodology for dialog management in spoken dialog systems. *Computer Speech and Language*, 28(3), 743–768. <http://dx.doi.org/10.1016/j.csl.2013.09.002>
- Lemon, O. (2011). Learning what to say and how to say it: Joint optimisation of spoken dialogue management and natural language generation. *Computer Speech and Language*, 25(2), 210–221. <http://dx.doi.org/10.1016/j.csl.2010.04.005>
- Lemon, O., & Pietquin, O. (Eds.) (2012). Data-driven methods for adaptive spoken dialogue systems: Computational learning for conversational interfaces. Springer. <http://dx.doi.org/10.1007/978-1-4614-4803-7>
- Frampton, M., & Lemon, O. (2009). Recent research advances in reinforcement learning in spoken dialogue systems. *Knowledge Engineering Review*, 24(4), 375–408. <http://dx.doi.org/10.1017/S0269888909990166>
- Baptist, L., & Seneff, S. (2000). GENESIS-II: A versatile system for language generation in conversational system applications. *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP '00)*, 3,

271–274.

- Dalianis, H. (1999). Aggregation in natural language generation. *Computational Intelligence*, 15(4), 384–414. <http://dx.doi.org/10.1111/0824-7935.00099>
- Dethlefs, N., Hastie, H., Cuayáhuitl, H., & Lemon, O. (2013). Conditional random fields for responsive surface realisation using global features. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, 1254–1263.
- Rieser, V., Lemon, O., & Keizer, S. (2014). Natural language generation as incremental planning under uncertainty: Adaptive information presentation for statistical dialogue systems. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 22(5), 979–994. <http://dx.doi.org/10.1109/TASL.2014.2315271>
- Stewart, J. Q. (1922). An electrical analogue of the vocal organs. *Nature*, 110, 311–312. <http://dx.doi.org/10.1038/110311a0>
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 236, 433–460. <http://dx.doi.org/10.1093/mind/LIX.236.433>
- Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <http://dx.doi.org/10.1145/365153.365168>
- Glass, J., Flammia, G., Goodine, D., Phillips, M., Polifroni, J., Sakai, S., ... & Zue, V. (1995). Multilingual spoken-language understanding in the MIT Voyager system. *Speech Communication*, 17(1–2), 1–18. [http://dx.doi.org/10.1016/0167-6393\(95\)00008-C](http://dx.doi.org/10.1016/0167-6393(95)00008-C)
- Zue, V., Seneff, S., Glass, J. R., Polifroni, J., Pao, C., Hazen, T. J., & Hetherington, L. (2000). JUPITER: A telephone-based conversational interface for weather information. *IEEE Transactions on Speech and Audio Processing*, 8, 85–96. <http://dx.doi.org/10.1109/89.817460>
- Hardy, H., Biermann, A., Bryce Inouye, R., McKenzie, A., Strzalkowski, T., Ursu, C., ... & Wu, M. (2006). The AMITIÉS system: Data-driven techniques for automated dialogue. *Speech Communication*, 48(3–4), 354–373. <http://dx.doi.org/10.1016/j.specom.2005.07.006>
- Melin, H., Sandell, A., & Ihse, M. (2001). CTT-bank: A speech controlled telephone banking system—An initial evaluation. *TMH-QPSR* 42(1), 1–27.
- Andreani, G., Di Fabbrizio, D., Gilbert, M., Gillick, D., Hakkani-Tur, D., & Lemon, O. (2006). Let's DISCOH: Collecting an annotated open corpus with dialogue acts and reward signals for natural language helpdesks. *IEEE 2006 Workshop on Spoken Language Technology*, 218–221. <http://dx.doi.org/10.1109/SLT.2006.326794>
- Bickmore, T., & Giorgino, T. (2006). Health dialog systems for patients and consumers. *Journal of Biomedical Informatics*, 39(5), 556–571. <http://dx.doi.org/10.1016/j.jbi.2005.12.004>
- Ahmad, F., Hogg-Johnson, S., Stewart, D. E., Skinner, H. A., Glazier, R. H., & Levinson, W. (2009). Computer-assisted screening for intimate partner violence and control: A randomized trial. *Annals of Internal Medicine*, 151(2), 93–102. <http://dx.doi.org/10.7326/0003-4819-151-2-200907210-00124>
- Ghanem, K. G., Hutton, H. E., Zenilman, J. M., Zimba, R., & Erbelding, E. J. (2005). Audio computer assisted self interview and face to face interview modes in assessing response bias among STD clinic patients. *Sexually Transmitted Infections*, 81(5), 421–425. <http://dx.doi.org/10.1136/sti.2004.013193>
- Hubal, R., & Day, R. S. (2006). Informed consent procedures: An experimental test using a virtual character in a dialog systems training application. *Journal of Biomedical Informatics*, 39(5), 532–540. <http://dx.doi.org/10.1016/j.jbi.2005.12.006>
- Pfeifer, L. M., & Bickmore, T. (2010). Designing embodied conversational agents to conduct longitudinal health interviews. *Proceedings of Intelligent Virtual Agents*, 4698–4703.
- Black, L. A., McTear, M. F., Black, N. D., Harper, R., & Lemon, M. (2005). Appraisal of a conversational artefact and its utility in remote patient monitoring. *Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems*, 506–508. <http://dx.doi.org/10.1109/CBMS.2005.33>
- Bickmore, T. W., Puskar, K., Schlenk, E. A., Pfeifer, L. M., & Sereika, S. M. (2010). Maintaining reality: Relational agents for antipsychotic medication adherence. *Interacting with Computers*, 22(4), 276–288. <http://dx.doi.org/10.1016/j.intcom.2010.02.001>
- Delichatsios, H., Friedman, R. H., Glanz, K., Tennstedt, S., Smigelski, C., Pinto, B., ... & Gillman, M. W. (2001). Randomized trial of a “talking computer” to improve adults’ eating habits. *American Journal of Health Promotion*, 15(4), 215–224. <http://dx.doi.org/10.4278/0890-1171-15.4.215>
- Ramelson, H. Z., Friedman, R. H., & Ockene, J. K. (1999). An automated telephone-based smoking cessation education and counseling system. *Patient Education and Counseling*, 36(2), 131–144. [http://dx.doi.org/10.1016/S0738-3991\(98\)00130-X](http://dx.doi.org/10.1016/S0738-3991(98)00130-X)
- Maglogiannis, I., Zafiroopoulos, E., & Anagnostopoulos, I. (2009). An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. *Applied Intelligence*, 30(1), 24–36. <http://dx.doi.org/10.1007/s10489-007-0073-z>
- Rich, C., & Sidner, C. L. (1998). COLLAGEN: A collaboration manager for software interface agents. *User Modeling and User-Adapted Interaction*, 8(3–4), 315–350. <http://dx.doi.org/10.1023/A:1008204020038>
- Catizone, R., Setzer, A., & Wilks, Y. (2003). Multimodal dialogue management in the COMIC project. *Proceedings of the EACL-03 Workshop on ‘Dialogue Systems: Interaction, Adaptation and Styles of Management’*. European Chapter of the Association for Computational Linguistics, 25–34.
- Corradini, A., Fredriksson, M., Mehta, M., Königsmann, J., Bernsen, N. O., & Johannesson, L. (2004). Towards

- believable behavior generation for embodied conversational agents. Proceedings of the Workshop on Interactive Visualisation and Interaction Technologies (IV&IT), 946–953.
- Balahur, A., Mihalcea, R., & Montoyo, A. (2014). Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications. *Computer Speech and Language*, 28(1), 1–6. <http://dx.doi.org/10.1016/j.csl.2013.09.003>
- Schuller, B. W., & Batliner, A. (2013). Computational paralinguistics: Emotion, affect and personality in speech and language processing. John Wiley & Sons. <http://dx.doi.org/10.1002/9781118706664>
- Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, 5(2), 119–124. <http://dx.doi.org/10.1177/1754073912468165>
- Nass, C., & Yen, C. (2012). The man who lied to his laptop: What we can learn about ourselves from our machines. *Current Trade*.
- Zhu, C., Sheng, W. (2011). Motion- and location-based online human daily activity recognition. *Pervasive and Mobile Computing*, 7(2), 256–269. <http://dx.doi.org/10.1016/j.pmcj.2010.11.004>
- Batliner, A., Seppi, D. Steidl, S., & Schuller, B. (2010). Segmenting into adequate units for automatic recognition of emotion-related episodes: A speech-based approach. *Advances in Human Computer Interaction*, 2010. <http://dx.doi.org/10.1155/2010/78202>
- Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2012). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio, Speech and Language Processing*, 20(1), 30–42. <http://dx.doi.org/10.1109/TASL.2011.2134090>
- Bui, T. H. (2006). Multimodal dialogue management- State of the art. Human Media Interaction Department, University of Twente (Vol. 2).
- Russ, G., Sallans, B., & Hareter, H. (2005). Semantic based information fusion in a multimodal interface. *Proceedings of the International Conference on Human–Computer Interaction, HCI '05*, Las Vegas. Lawrence Erlbaum.
- Longé, M., Eyraud, R., & Hullfish, K.C. (2012). Multimodal disambiguation of speech recognition. U.S. Patent No. 8095364 B2. Retrieved from <http://www.google.com/patents/US8095364>
- Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, speech recognition, and computational linguistics* (2nd ed.). Prentice Hall.
- Bohus, D., & Rudnicky, A. I. (2003). RavenClaw: Dialog management using hierarchical task decomposition and an expectation agenda. *Proceedings of the 8th European Conference on Speech Communication and Technology, EUROSPEECH 2003–INTERSPEECH 2003*, 597–600.
- Rothkrantz, L. J. M., Wiggers, P., Flippo, F., Woei-A-Jin, D., & van Vark, R. J. (2004). Multimodal dialogue management. *Lecture Notes in Computer Science*, 3206, 621–628. http://dx.doi.org/10.1007/978-3-540-30120-2_78
- Larsson, S. & Traum, D. R. (2000). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering*, 6(4), 323–340. <http://dx.doi.org/10.1017/S1351324900002539>
- Young, S., Gasic, M., Thomson, B., & Williams, J. D. (2013). POMDP-based statistical spoken dialog systems: A review. *Proceedings of the IEEE*, 101(5), 1160–1179. <http://dx.doi.org/10.1109/JPROC.2012.2225812>
- M. Oshry, P. Baggio, K. Rehor, M. Young, R. Akolkar, X. Yang, J. Barnett, R. Hosn, R. Auburn, J. Carter, S. McGlashan, M. Bodell, and D. C. Burnett, Voice extensible markup language (VoiceXML) 3.0, W3C, W3C Working Draft, Dec. 2009. [Online]. Available: <http://www.w3.org/TR/2009/WD-voicexml30-20091203/>.
- T. Paek and R. Pieraccini, “Automating spoken dialogue management design using machine learning: An industry perspective,” *Speech Commun.*, vol. 50, no. 8–9, pp. 716–729, 2008.
- R. P. Lippmann, “Speech recognition by machines and humans,” *Speech Commun.*, vol. 22, no. 1, pp. 1–15, 1997.
- A. Black, S. Burger, A. Conkie, H. Hastie, S. Keizer, O. Lemon, N. Merigaud, G. Parent, G. Schubiner, B. Thomson, J. Williams, K. Yu, S. Young, and M. Eskenaizi, “Spoken dialog challenge 2010: Comparison of live and control test results,” in *Proc. Annu. Meeting Special Interest Group Discourse Dialogue*, Portland, OR, 2011, pp. 2–7.
- J. Williams and S. Young, “Partially observable Markov decision processes for spoken dialog systems,” *Comput. Speech Lang.*, vol. 21, no. 2, pp. 393–422, 2007.
- L. Kaelbling, M. Littman, and A. Cassandra, “Planning and acting in partially observable stochastic domains,” *Artif. Intell.*, vol. 101, pp. 99–134, 1998.
- R. E. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1957.
- J. Pineau, G. Gordon, and S. Thrun, “Point-based value iteration: An anytime algorithm for POMDPs,” in *Proc. Int. Joint Conf. Artif. Intell.*, Acapulco, Mexico, 2003, pp. 1025–1032.
- T. Smith and R. G. Simmons, “Heuristic search value iteration for POMDPs,” in *Proc. Conf. Uncertainty Artif. Intell.*, Banff, AB, Canada, 2004, pp. 520–527.
- M. Gasic and S. Young, “Effective handling of dialogue state in the hidden information state POMDP dialogue manager,” *ACM Trans. Speech Lang. Process.*, vol. 7, no. 3, 2011, Article 4.
- B. Thomson and S. Young, “Bayesian update of dialogue state: A POMDP framework for spoken dialogue systems,” *Comput. Speech Lang.*, vol. 24, no. 4, pp. 562–588, 2010.
- P. A. Crook and O. Lemon, “Lossless value directed compression of complex user goal states for statistical spoken dialogue systems,” in *Proc. Annu. Conf. Int. Speech Commun. Assoc.*, Florence, Italy, 2011, pp. 1029–1032.

- F. Doshi and N. Roy, “The permutable POMDP: Fast solutions to POMDPs for preference elicitation,” in Proc. Int. Joint Conf. Autonom. Agents Multiagent Syst., 2008, pp. 493–500.
- J. Williams and S. Young, “Scaling up POMDPs for dialog management: The ‘summary POMDP’ method,” in Proc. IEEE Workshop Autom. Speech Recognit. Understand., Cancun, Mexico, 2005, pp. 177–182.
- J. Williams and S. Young, “Scaling POMDPs for spoken dialog management,” IEEE Trans. Audio Speech Lang. Process., vol. 15, no. 7, pp. 2116–2129, Jul. 2007.
- D. Andre and S. Russell, “State abstraction for programmable reinforcement learning agents,” in Proc. 18th Nat. Conf. Artif. Intell., Edmonton, AB, Canada, 2002, pp. 119–125.
- J. D. Williams, “The best of both worlds: Unifying conventional dialog systems and POMDPs,” in Proc. Annu. Conf. Int. Speech Commun. Assoc., Brisbane, Australia, 2008, pp. 1173–1176.
- P. Lison, “Towards relational POMDPs for adaptive dialogue management,” in Proc. Annu. Meeting Assoc. Comput. Linguist., Uppsala, Sweden, 2010, pp. 7–12.
- S. Young, M. Gas’ic, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu, “The hidden information state model: A practical framework for POMDP-based spoken dialogue management,” Comput. Speech Lang., vol. 24, no. 2, pp. 150–174, 2010.
- L. Li, J. D. Williams, and S. Balakrishnan, “Reinforcement learning for dialog management using least-squares policy iteration and fast feature selection,” in Proc. Annu. Conf. Int. Speech Commun. Assoc., Brighton, U.K., 2009, pp. 2475–2478.
- O. Pietquin, M. Geist, S. Chandramohan, and H. Frezza-Buet, “Sample-efficient batch reinforcement learning for dialogue management optimization,” ACM Trans. Speech Lang. Process., vol. 7, no. 3, pp. 1–21, 2011.
- K. Scheffler and S. Young, “Automatic learning of dialogue strategy using dialogue simulation and reinforcement learning,” in Proc. 2nd Int. Conf. Human Lang. Technol. Res., San Diego, CA, 2002, pp. 12–19.
- J. Henderson, O. Lemon, and K. Georgila, “Hybrid reinforcement/supervised learning of dialogue policies from fixed datasets,” Comput. Linguist., vol. 34, no. 4, pp. 487–511, 2008.
- J. Schatzmann, K. Weilhammer, M. Stuttle, and S. Young, “A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies,” Knowl. Eng. Rev., vol. 21, no. 2, pp. 97–126, Jun. 2006.
- H. Hastie, “Metrics and evaluation of spoken dialogue systems,” in Data-Driven Methods for Adaptive Spoken Dialogue Systems: Computational Learning for Conversational Interfaces. New York: Springer-Verlag, 2012.
- O. Pietquin and H. Hastie, “A survey on metrics for the evaluation of user simulations,” Knowl. Eng. Rev., 2012, DOI:<http://dx.doi.org/10.1017/S0269888912000343>.
- J. Schatzmann, K. Georgila, and S. Young, “quantitative evaluation of user simulation techniques for spoken dialogue systems,” in Proc. Annu. Meeting Special Interest Group Discourse Dialogue, Lisbon, Portugal, 2005, pp. 45–54.
- J. Williams, “Evaluating user simulations with the Cramervon Mises divergence,” Speech Commun., vol. 50, pp. 829–846, 2008.
- S. Janarthanam, H. Hastie, O. Lemon, and X. Liu, ““The day after the day after tomorrow?” A machine learning approach to adaptive temporal expression generation: Training and evaluation with real users,” in Proc. Annu. Meeting Special Interest Group Discourse Dialogue, Portland, OR, 2011, pp. 142–151.
- K. Kim and G. G. Lee, “Multimodal dialog system using hidden information state dialog manager,” in Proc. Int. Conf. Multimodal Interfaces Demonstration Session, Nagoya, Japan, 2007.
- K. Kim, C. Lee, S. Jung, and G. Lee, “A frame-based probabilistic framework for spoken dialog management using dialog examples,” in Proc. Special Interest Group Discourse Dialogue (SIGDIAL) Workshop Discourse Dialogue, Columbus, OH, 2008, pp. 120–127.
- J. Williams, “Applying POMDPs to Dialog Systems in the Troubleshooting Domain,” in Proc. HLT/NAACL Workshop Bridging the Gap: Acad. Ind. Res. Dialog Technol., Rochester, NY, 2007, pp. 1–8.
- B. Thomson, K. Yu, S. Keizer, M. Gasic, F. Jurcicek, F. Mairesse, and S. Young, “Bayesian dialogue system for the let’s go spoken dialogue challenge,” in Proc. IEEE Spoken Lang. Technol. Workshop, Berkeley, CA, 2010, pp. 460–465.
- E. Levin, R. Pieraccini, and W. Eckert, “Learning dialogue strategies within the Markov decision process framework,” in Proc. IEEE Workshop Autom. Speech Recognit. Understand., Santa Barbara, CA, 1997, pp. 72–79.
- S. Singh, M. Kearns, D. Litman, and M. Walker, “Reinforcement learning for spoken dialogue systems,” in Proc. Neural Inf. Process. Syst., 1999.