

# 机器学习在青少年性教育中的应用研究

陈龙 于一豪 薛皓天 宋博凡 梁彦泽 许兵兵

(武汉科技大学城市学院信息工程学部 武汉 430083)

**摘 要** 为了让学生们能够在线上轻松地学习到科学全面系统的性知识,受线上“会话式学习”的新模式启发,本文整合有关性教育的教案资源、整理知识数据库,设计会话式的趣味学习模块,基于微信小程序,让用户在线上轻松获取全面系统的知识,从而全面高效地传播正确的性知识与性观念。为了给学生们提供一些相对专业的心理疏导,结合人工智能,给青少年一个倾诉和寻求帮助的对象,并为知识的授予方设计了一个“小莓人”的AI tutor形象。

**关键词** 性教育;会话式学习;人工智能;小莓人;微信小程序

中图分类号 TP391.9 DOI:10.16707/j.cnki.fjpc.2020.08.040

## Research on the Application of Machine Learning in Sex Education for Teenagers

CHEN Long, YU Yihao, XUE Haotian, SONG Bofan, Lian Yanze, XU Bingbing

(School of Informatics Division, City College, Wuhan University of Science and Technology, Wuhan, China, 430083)

### 1 引言

中国现阶段教育系统和自古以来对性知识的忌讳造成性教育偏失,导致青少年性侵犯高发、婚前怀孕率高、性病罹患率高涨等社会问题。本文以人工智能在青少年与儿童性教育方面的应用为研究领域,研究人工智能辅助性教育的可行性,并尝试建立基于移动互联网、实体机器人的公益性人工智能性教育应用场景<sup>[1]</sup>。

### 2 技术路线

#### 2.1 会话式学习

小程序最重要的功能之一就是利用会话式学习模式帮助学生学习到科学系统的性知识,让学习过程变得轻松有趣,让教育环节轻度化、碎片化。本文采取的方案是利用 button 及其指向路径的不断

变换来完成会话式交互。并且将每一个知识点划分为几个小的模块,每一个小的模块包含数量不定的经验值。模块的最后一个 button 被点击完成后,用户将会得到相对应的经验值加成,以此达到用户的经验和等级的提升。

#### 2.2 智能聊天基础功能的实现

AIUI 是科大讯飞提供的一套人机智能交互解决方案,免费提供了内置的开放技能、开发者自定义技能、内置的开放问答库和开发者自定义问答库。AIUI 服务器所连接的云函数服务器位于语义理解引擎之后,运行环境为 NodeJS,可以直接与万维网通信。AIUI 在解析用户的语义后,将用户的意图传递给云函数,开发者可以自主调用万维网上的接口 API,查询到结果后,将播报结果回传 AIUI。也就是说,学生在小程序端口发送信息,信息在显示的同时会发送到 AIUI 的 API 接口中,获取一个返回的消息并且显示在界面中。接入 AIUI 的 API 接口后将发出和收到的数据都储存在一个数组中,

本文得到湖北省高等学校大学生创新创业训练计划项目(No.S201913235008)资助。陈龙,男,1997年生,主要研究领域为数字图像处理、计算机视觉。E-mail: 303549722@qq.com。于一豪,男,1998年生,主要研究领域为计算机视觉。E-mail: 152874992@qq.com。薛皓天,男,1999年生,主要研究领域为人工智能。E-mail: 1900332593@qq.com。宋博凡,男,1999年生,主要研究领域为数字图像处理。E-mail: 1139882150@qq.com。梁彦泽,男,2000年生,主要研究领域为数据挖掘处理。E-mail: 2971746737@qq.com。许兵兵,男,1998年生,主要研究领域为软件工程。E-mail: 1693579185@qq.com。

然后通过链表的方式在界面中输出。

### 2.3 语音识别技术

语音识别技术，就是能够让机器识别人的语音，识别语音中包含的文字信息，相当于给机器安装上“耳朵”<sup>[2]</sup>，使其具备“听”的功能。由于微信小程序利用 `wx.startRecord()` 和 `wx.stopRecord()` 录音并上传至服务器的录音文件都是 silk 格式的，需要通过 `ffmpeg` 将 silk 录音文件转成 wav 录音文件，然后利用 `Node.js` 来访问科大讯飞的语音识别的 API，从而知道语音识别出来的结果，实现对学生发送的语音信息进行识别和智能回复<sup>[3]</sup>。

## 3 核心算法

### 3.1 伦理引擎过滤器

互联网上的海量信息在很大程度上方便了机器的训练及自主学习，但是由于网络监控在法律制度和技術上的不完善，互联网中存在着大量的不良信息，尤其在性知识方面。网络上常常伴随一些色情暴力的信息，倘若没有一个过滤系统，不仅不利于青少年心理成长，甚至会起到反作用。为此，需要为小莓人设计一套伦理引擎，在信息上传至服务器和服务器获取回复之前，对传输的信息进行敏感词识别与过滤，如图 1 所示。这一功能将会利用 DFA 算法来实现<sup>[4]</sup>。

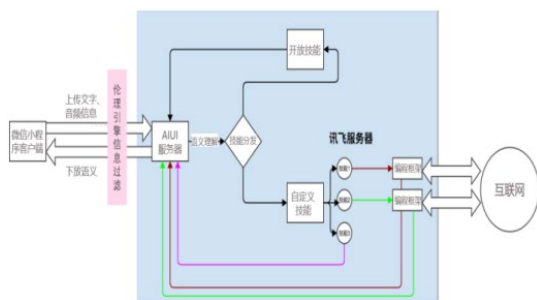


图 1 DFA 算法流程图

所有的违禁词均由 ASCII 码组成的，而待过滤文本的本质也是 ASCII 码的集合。将这些 ASCII 码集合创建为敏感词库，利用 DFA 算法将敏感词库构建成类似一颗一颗的树。在识别敏感词时首先查询“色”字是否存在于词库中，如果不存在则证明以“色”开头的敏感词不存在。如果“色”字开头的敏感词存在，则依次识别“情”、“图”等字直到最后一个字，以此判断该词汇是否为敏感词。如果出现敏感词中填充无意义字符来混肴视听的情况

(“例如色&&情”)，可以在对敏感词搜索时，做一个无意义词的过滤。当循环到这类无意义的字符时跳过，避免干扰，这样判断一个词是否为敏感词时就大大减少了检索的匹配范围。其流程如图 2 所示。

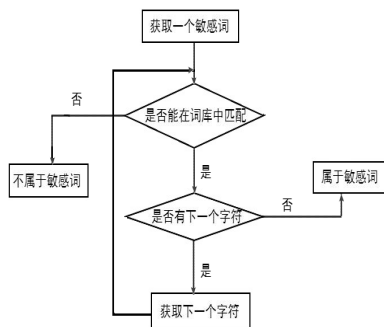


图 2 敏感词识别流程图

### 3.2 类人聊天算法

为了让小莓人与用户的聊天交互更为自然，让 AI 能够智能记忆用户的习惯用语，甚至是以往的对话内容，从而在聊天过程中做出更智能更人性化的反应，根据用户个性化迭代，提高用户体验。本项目选择利用 BP 算法训练人工智能的自主学习，从而实现类人聊天。

## 4 总结

基于机器学习的人工智能技术在青少年性教育中的应用研究，是当前教育领域的盲点与性教育领域的难点，具有较大的研究价值和实践价值。本项目围绕研究选题，前期已经完成了主要的原型设计与功能的开发，验证了方案的合理性与可行性。本论文研究 DFA 与 BP 算法<sup>[6]</sup>，目前已实现了部分代码，增加特有的过滤引擎、训练 AI 自主学习相关语义，进一步地优化完善核心功能。

### 参考文献

- [1] 李祥,陈凤.国际人工智能教育研究的进展及其对专业人才培养的启示.黑龙江高教研究,2020(04):256-260
- [2] 张会云.语音情感识别研究综述.信息通信,2019(11):78-80
- [3] 凌利,陈龙,祝扬,宋博凡,杨华勇.TensorFlow智能医疗服务平台设计.福建电脑,2019,35(04):95-96
- [4] 肖红.大数据下的机器学习算法探讨.通讯世界,2017(06):41-44
- [5] 张石清,李乐民,赵知劲.人机交互中的语音情感识别研究进展.电路与系统学报,2013(02):48-50
- [6] 彭志勇,龙虎.人工智能背景下量子机器学习算法的概论.信息与电脑(理论版),2020(04):69-72