

6-4-2020

Research and Implementation of Question Answering System Oriented Wargaming Based on AIML

Zejian Sun

The Department of Information Operation& Command Training, NDU, Beijing 100091, China;

Guangya Si

The Department of Information Operation& Command Training, NDU, Beijing 100091, China;

Liu Yang

The Department of Information Operation& Command Training, NDU, Beijing 100091, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Research and Implementation of Question Answering System Oriented Wargaming Based on AIML

Abstract

Abstract: Wargaming exercise is an important means of combat simulation, but the high query and operation threshold of wargaming system limit its use and extension. Through investigating the actual operation process of a wargaming exercise, collecting and analyzing the trainees use requirements, *the intelligent question answering system centering on access to the database by making knowledge documents, index tables and access modules was designed based on AIML(Artificial Intelligence Markup Language) and ALICE program, adopted heuristic rules.* The implementation of question answering system made commanders participating in wargaming query data by natural language, provided ideas for the further improvement of wargaming system as well as an interface for the collection of question answering data in wargaming.

Keywords

AIML, question answering system, wargaming, SQL query

Recommended Citation

Sun Zejian, Si Guangya, Liu Yang. Research and Implementation of Question Answering System Oriented Wargaming Based on AIML[J]. Journal of System Simulation, 2017, 29(10): 2489-2496.

基于 AIML 的兵棋演习智能问答系统研究与实现

孙泽健, 司光亚, 刘洋

(国防大学信息作战与指挥训练教研部, 北京 100091)

摘要: 兵棋演习是重要的作战模拟手段, 但兵棋系统较高的查询操作门槛限制了其使用和扩展。通过调查了解某一兵棋演习的实际运行过程, 搜集并分析参训人员的使用需求, 设计了以访问兵棋数据库为主的智能问答系统。问答系统以 AIML 语言(Artificial Intelligence Markup Language, 人工智能标记语言)和 ALICE 程序为基础, 采用启发式规则, 制作了特定的 AIML 文档、数据库索引表和数据库访问模块, 实现了对某一兵棋数据库的访问。系统的完成使参演人员可以以自然语言查询兵棋演习交战数据, 实时快速地解答参演人员的各种疑问, 并实现演习问答数据的全时采集及交战数据的快速统计分析。

关键词: AIML; 问答系统; 兵棋演习; SQL 语句

中图分类号: TP391.1

文献标识码: A

文章编号: 1004-731X (2017) 10-2489-08

DOI: 10.16182/j.issn1004731x.joss.201710033

Research and Implementation of Question Answering System Oriented Wargaming Based on AIML

Sun Zejian, Si Guangya, Liu Yang

(The Department of Information Operation& Command Training, NDU, Beijing 100091, China)

Abstract: Wargaming exercise is an important means of combat simulation, but the high query and operation threshold of wargaming system limit its use and extension. Through investigating the actual operation process of a wargaming exercise, collecting and analyzing the trainees use requirements, *the intelligent question answering system centering on access to the database by making knowledge documents, index tables and access modules was designed based on AIML(Artificial Intelligence Markup Language) and ALICE program, adopted heuristic rules.* The implementation of question answering system made commanders participating in wargaming query data by natural language, provided ideas for the further improvement of wargaming system as well as an interface for the collection of question answering data in wargaming.

Keywords: AIML; question answering system; wargaming; SQL query

引言

智能问答系统(Question Answering System,



收稿日期: 2017-04-30 修回日期: 2017-10-12;
基金项目: 国家自然科学基金(U1435218, 61403401);
作者简介: 孙泽健(1988-), 男, 青岛, 硕士生, 研究方向为计算机战争模拟; 司光亚(1967-), 男, 博导, 教授, 研究方向为计算机战争模拟。

QA 系统)是一种新的信息检索系统, 也是一种新的人机交互方式, 用户可以用自然语言向 QA 系统提问, QA 系统能自动查询知识库并返回答案。近年来, 问答系统逐渐由学术研究转为技术应用, 如 Apple 公司的移动个人助理 Siri 可以通过语音或文字与手机用户交互, 实现个人事务的查询及代办功能; IBM 公司的 Watson 则在智力竞答节目 Jeopardy

<http://www.china-simulation.com>

• 2489 •

中战胜了两名人类高手而赢得冠军,展现了知识图谱和自然语言处理技术相结合带来的诸多可能性,种种科技前沿进展都为当前兵棋系统进一步的改进提供了思路。

当前存在的多种兵棋系统多是典型的“人在回路”系统,人是其中的主导和关键因素,但目前受数据查询和系统操作的复杂性困扰,经常出现指令下达出现错误,数据反馈不及时等问题,改变这一现状的重要方法是将自然语言处理技术引入兵棋系统的使用过程中,使参演人员在兵棋演习过程中,不再只面对“冷冰冰”的指令和数据,也可以遇到“有温度”的“人话”,构建兵棋推演智能问答系统(以下简称兵棋问答系统)就是一个初步的尝试,希望在演习过程中解决使用人员产生的各种疑问,快速准确地反馈演习数据,减轻保障人员工作量,提高兵棋演习的质量。

1 需求和技术思路

1.1 兵棋演习对智能问答的需求

通过参加某兵棋演习,并调查其采用的兵棋系统发现,演习的训练保障工作仍有许多亟待提高的地方,一是由于演习前的培训时间短,参演人员对于课程辅导内容消化吸收不够,导致后续操作不熟

练,产生大量疑问需要导调人员解答;二是传统的查询方式难以迅速跟踪演习进程,大型兵棋系统能够描述陆、海、空、天、电、网等多维战场空间各种作战实体在不同作战行动支配下的各种复杂交互,仿真实体规模巨大,数量众多,极易给使用人员造成战争炫目^[1-2];三是上面两个问题造成了人力的浪费,演习过程中需要大量保障人员全时值守,往返于各个不同作战席位,重复回答很多问题且工作量巨大。

将人工智能和兵棋服务保障工作相结合,则至少可部分地解决上述问题。兵棋问答系统不但可以实现数据的自然语言访问,还可以统一回答不同席位参训人员的问题,减少使用界面间的切换。依据现有的商业开发经验^[3-4],问答系统可以提供结合图片、动画的富媒体答案,甚至可以收集使用领域产生的大量文本问题数据,为下一步的大数据分析和系统的更高智能化打下基础。

1.2 技术思路

通过参加兵棋演习,和参演人员进行交流,收集其在演习中产生的疑问,组成一个小样本的兵棋问题集来对具体需求进行分析。其所涉及的问题种类多样,大致可以分为 5 类,表 1 给出了分类的主要依据和典型例句。

表 1 兵棋演习过程中的常见问题示例

Tab. 1 Examples of common problems in the process of wargaming exercises

类别	问题范围	所需知识库	例句
第一类	兵棋系统操作和运行规则	FAQ 库	地图中的六角格有什么作用? 飞机返航过程中可以调整返航机场吗?
第二类	查询想定和实时信息	兵棋数据库	XXX 部队是否出发? Y 导弹的射程是多少?
第三类	查询战况统计信息	兵棋数据库	我方还有多少艘潜艇? 我方击落了多少架敌机?
第四类	询问军事概念和战法策略	军事百科/自由文本	什么是 X 型装备? 潜艇游猎活动会不会攻击己方潜艇?
第五类	其他		联指的作战决心有没有下发?

根据数据统计和兵棋系统现有条件进行分析,可知用户输入的问题主要有以下特征:

(1) 大多数问题是带疑问词的简单句,其中又

以前 3 类问题数量最多。

(2) 询问简单句中大量简称、代字甚至不规则文法,对其进行分词和句法分析的效果一般。

(3) 许多问题需要通过生成数据库查询语言, 访问数据库来查找答案。

经过多方面比较研究^[5-7], 拟采用启发式模板匹配的方法, 以 AIML 语言^[8]和 ALICE 程序^[9-10]为基础, 增加相应的处理标签和程序处理模块, 创建兵棋演习领域的知识文档, 以 PyCharm 开发环境, 通过访问兵棋数据库来回答第二、三类问题以及部分第一类问题。

2 问答系统的设计

基于 AIML 的兵棋问答系统其框架结构如图 1 所示。主要包括 AIML 文档、数据库访问索引表、数据库访问参数项和处理模块、问句归一化模板和文本答案。首先, 对用户输入的问题进行规范化处

理和初步的分类, 将 FAQ 类问题和数据库访问类问题分别转入相应的处理模块。其次, 对于访问数据库的问题, 通过中间处理生成完整的 SQL 查询语句。最后, 访问数据库返回数据, 并和 AIML 的文本模板生成文本答案返回给用户。

兵棋数据查询需要访问兵棋系统的关系型数据库, 从表和索引中查找相应答案, 故此子系统实现的关键在于找到输入问句提供的有限信息和查询目标的表列信息间的映射关系, 从而让系统自动生成完整的 SQL 语句。据此必须分别对 AIML 语言和 ALICE 程序进行扩充和改进, 增加相应的查询标签和生成 SQL 语句的程序模块, 同时编写合理的数据库索引表, 再据此编写相应的 AIML 条目。

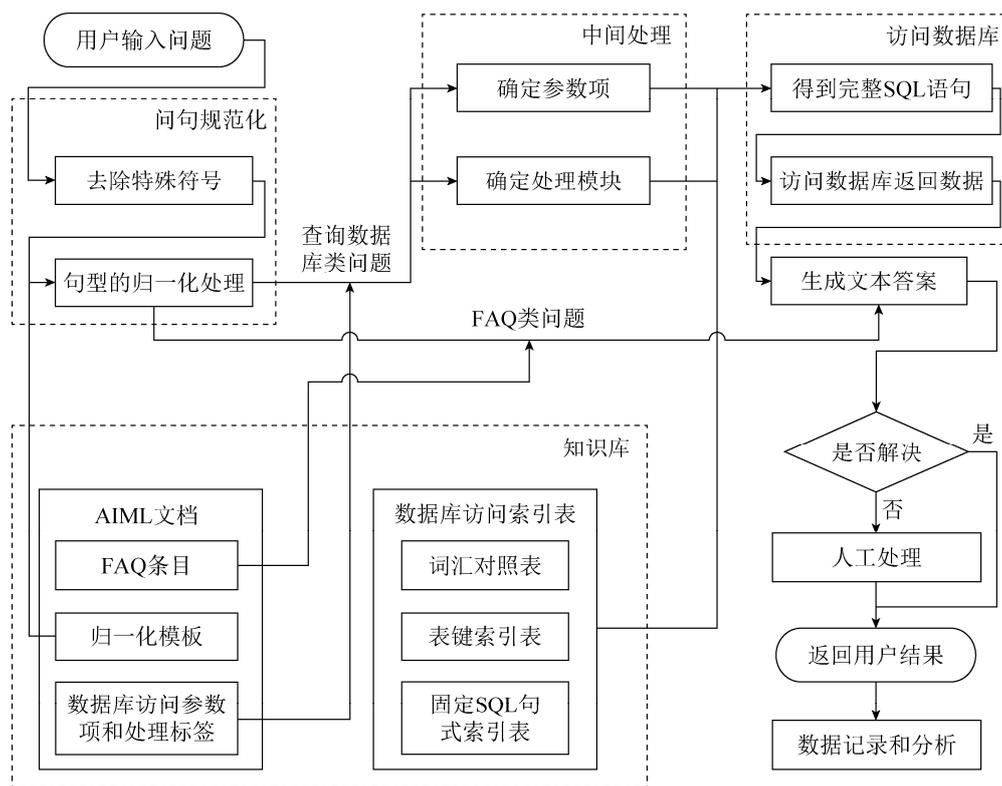


图 1 问答系统框架

Fig. 1 QA system frame

2.1 数据库访问的参数项

用户输入的访问兵棋数据库的问题虽然是一整个问句, 但其提供的信息不超过几个关键词, 比如“某型导弹的射程有多远?”问句和“某型导弹

射程”两个关键词提供的信息量大致相同, 而“蓝方现在还有几艘潜艇?”问句和“蓝方 潜艇 数量”提供的信息量大致相同, 因为兵棋演习过程中问得一般是当前情景, 一般查询的是实时信息, 故“现在”

等时间词汇可以省略。经过分析,我们可以在编辑归一化模板化简输入问句的同时,编辑合适的参数项记录输入问句提供的信息。本文设计了多个参数,主要包括

Name='search', Value='LOSS'/'RANGE'/'REMNANT'/'.....', 记录查询的类别。

Name='side', Value='WE'/'FOE'/'NEU'/'.....', 记录查询的集团名称。

Name='domain', Value='AIR'/'NAVAL'/'SSM'/'.....', 记录查询目标的可能范畴。

Name='wanted', Value='SUBMARINE'/'ROTORY'/'CARRIER'/'JAMMAR'/'.....', 记录查询超类名称的查询目标。

Name='table', Value='TABLE_X'/'TABLE_Y'/'TABLE_Z'/'.....', 记录某一数据库表或索引。

Name='column', Value='Column_X'/'Column_Y'/'Column_Z'/'.....', 记录某一数据库字段。

AIML 系统可以通过 3 种方式传递参数,为系统的扩展和补充提供了便利,分别为:

(1) 利用自定义标签的默认参数项传递参数。

AIML 系统为每一个标签预留了特定的参数项位置,只要输入参数项的名称和系统修改的名称相同,参数就可以有效传递进相关处理模块,如编辑 `<sdr cloumn='FUEL' domain='NAVAL'/>`,相关参数就可以传递给 sdr 处理模块。

(2) 利用 `<star>` 标签传递参数。AIML 系统不但可以通过 `*` 通配符实现了知识的有效组织,还通过 `<star>` 标签将所替换的字段记录下来。此功能可以用来记录具体特定的查询目标或者代字,比如“Z 舰船”,“XXYY001”。只要编辑合适的索引表,再利用模糊查询(LIKE 语句)句式,就可以实现有效的数据库查询。

(3) 利用 `<think>` 标签传递参数。`<think>` 标签和 `<set>` 以及 `<get>` 标签相配合,可以记录系统未设置的参数项,如通过编辑代码段 `<think><set name="wargame_name">"XX2017"</set>`,可以创建 "wargame_name" 参数项,并传递参数值 "XX2017"

给它。利用 `<think>` 标签可以保留的系统信息,如兵棋数据库的名称和配置情况、兵棋演习的编号或者席位信息。

2.2 数据库访问处理模块

经过比较和分析,用户输入的需要调用兵棋数据库的相似问题一般有相似的 SQL 查询语句,如“X 飞机的航程是多少?”和“W 舰船的最大装载量是多少?”,查询目标在兵棋数据库的同一张原型表中,只是调用的条件和查询的字段不同,所以只要找到其中最常见的几种固定模式就可以使系统覆盖相当比例的问题。故本文除了利用 AIML 原有的 `<condition>` 标签和处理模块之外,还在 AIML 源文件基础上扩展构建了几种数据库访问处理模块,每一模块绑定一个或几个固定的 SQL 查询句式(如 "select distinct __from__where__")。

2.2.1 提取参数项并转换为 SQL 数据项

输入问题通过 AIML 的标签被提取生成相关的参数项后(见 2.1),根据数据库访问索引表(见 2.3)将参数项映射成 SQL 语句的目标字段、目标表以及查询条件等数据项。如通过参数项 Name='domain', Value='AIR' 可得,数据要查询 'TABLE_AIR_UNIT' 实体表和 'TABLE_AIR_PROTO' 原型表,从索引表中可以找到 'TABLE_AIR_UNIT.Column_X=TABLE_AIR_PROTO.Column_Y'。

2.2.2 合成完整 SQL 语句

将所有的参数项全部转换为相关的数据项之后,可以根据每一模块绑定的一个或几个固定的 SQL 查询句式,合成完整的 SQL 语句。对于某些嵌套而成的 SQL 语句,可以先生成部分条件子句,将其作为一个数据项,和其他数据项一起生成最终完整的 SQL 语句。

2.2.3 访问数据库查询数据

得到完整的 SQL 语句后,调整数据库配置信息,通过原有接口访问兵棋数据库,得到数据值,并将其本模块的结果返回给 AIML 条目,和其他文

本数据一起生成文本答案(见 2.4)。

2.3 数据库访问索引表

数据库索引表是参数项到完整 SQL 查询语句数据项的“桥梁”,其编写必须着眼于保证系统最终得到的完整 SQL 语句无语法错误,实际可用,同时尽量减少参数项的设置,易于编写 AIML 的条目。

2.3.1 词汇对照表

将自然语言的词汇转换成兵棋数据库可识的字段值或数据值,也可以包括部分的系统缩略词和代字。如“歼十/歼十飞机/歼 10 飞机/歼 10B/J10/J-10/J10 飞机”在数据库中都表示同一原型或同一实体,其数据值一般为“J10”或“J-10”,词汇对照表就是要建立自然语言词汇和数据库数据值的映射关系,配合部分模糊查询的 SQL 语句(如 SQL 数据项“%J[-]10%”),就可以大大提高问答系统的召回率。同时结合 AIML 自带的<star>标签,就可以极大地简化 AIML 知识库的编写。

2.3.2 数据库表键索引表

记录了数据库每张实体表或者原型表和其主键或其他字段的对应关系,以及同一领域实体表和原型表部分字段的对应关系。由于输入问题提供的信息量有限,但通过表示查询目标的单个词汇,就可以确定查询的原型表或实体表和其主键,如查询“X 型航母”基本可以确定查询舰船的原型表,通过表键索引表就可以知道需要查询的字段。战况统计查询中多是潜艇、三代机、雷达等超类的查询目标,从中还可以确定超类的字段名。除此之外,部分 SQL 查询语句包括数据库里的多张数据表,将数据表之间的固定关系(其字段之间的对应关系)编成索引表,就可以由所得的参数项生成 SQL 查询语句的条件(where 子句)。

2.3.3 固定 SQL 句式索引表

记录了常见的 SQL 查询语句的模板。通过总

结归纳常见问题的 SQL 查询语句,从中提取出通用的 SQL 句式和部分子句句式,进行编号,构建成索引表。在数据库访问模块中,通过将数据项填入绑定的 SQL 句式,最终生成了完整的 SQL 查询语句(见 2.2)。

2.4 知识库

2.4.1 归一化问句模板

面向兵棋演习的同一问题有多种问法,如“Y 飞机能飞多远?”“Y 飞机的航程是多少?”“Y 最远能飞多少公里?”,等等不一而足,但查询的都是原型名为 Y 的飞机的航程。AIML 语言提供了可以有效归纳输入问题的<srai>标签,比如利用<srai>标签,可以在 AIML 知识库中将询问 Y 飞机航程的问题全部转换到名为“Y 航程”的条目,甚至配合<star>标签直接将所有类似问题转换到“* 航程”的条目,再通过识别“*”来查询 Y 飞机航程。

针对上述情况,编辑覆盖面广泛的归一化模板来规范化输入的问题变得尤为重要,成为提高系统召回率的关键,毕竟兵棋系统的所有数据都存在数据库中,只要确定其 SQL 查询语句,相关数据都可以查到。但更多的模板有赖于在系统使用过程中对输入数据的收集和分析。

2.4.2 文本答案

从数据库中返回的数据只是单个数值,没有数值单位,没有时间戳,也没有说明来源,由此就会产生歧义,甚至误导询问系统的参训人员,如系统中多采用公里、公里每时,但实际某些情况多采用海里等单位,容易产生歧义,需要在文本生成,返回答案的过程中予以消除。AIML 提供的<template>标签默认提供文本编辑方式,可以将从数据库返回的数据项嵌入到文本答案中,在知识库的编辑过程添加相应单位、时间戳以及数据来源,可以有效解决问题。

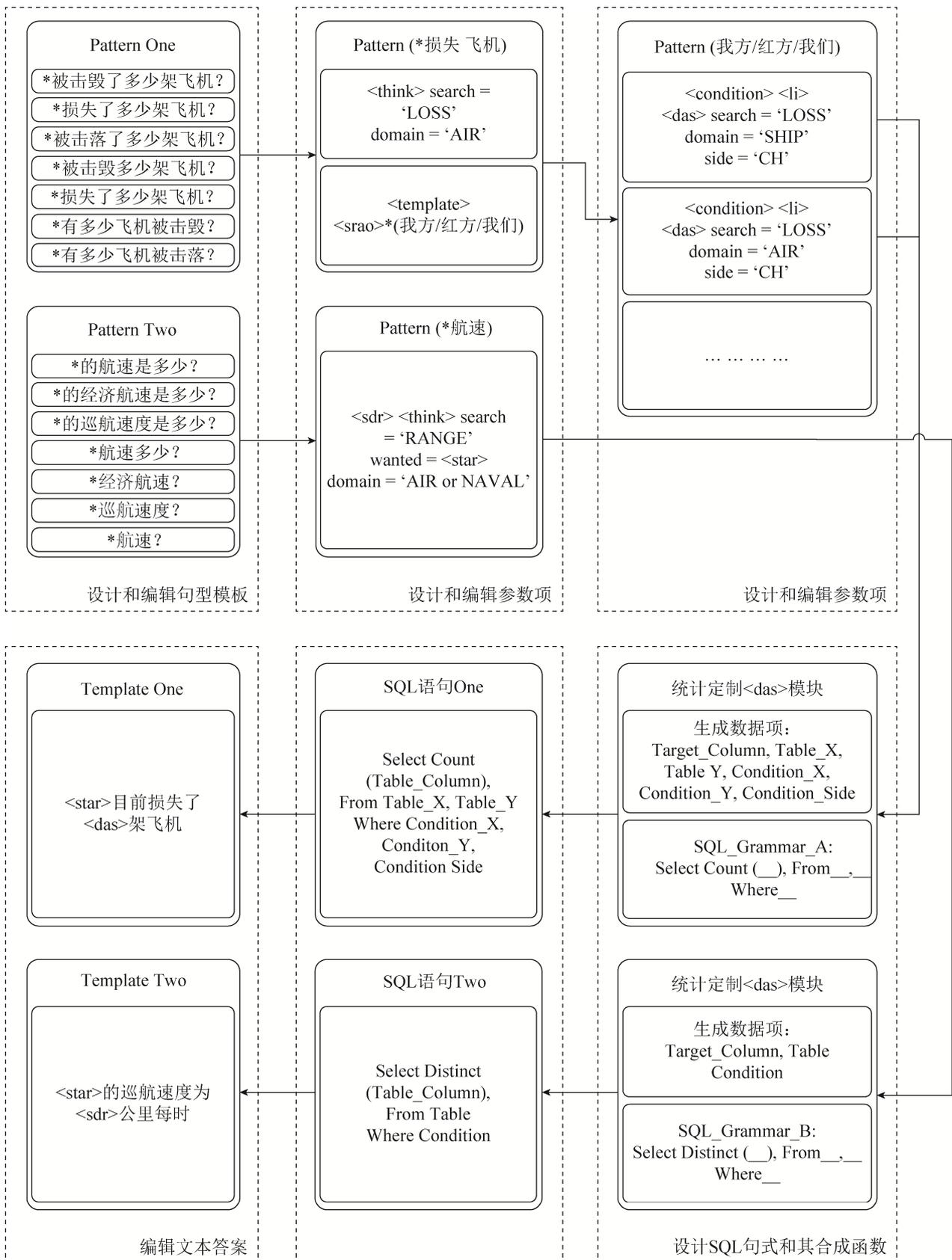


图 2 问答系统示例
Fig. 2 QA system example

3 问答系统的实现

利用 PyCharm 开发环境、PyAIML 工具包以及部分开源的 AIML 文档, 初步实现了面向兵棋演习的智能问答系统。支持 GUI 界面的文字输入, 以问题“红方损失了多少架飞机?”为例, 说明问答系统的主要工作流程(见图 2)。

- (1) 对问句进行归一化处理, 通过<srai>标签转到了问题“*损失 飞机”;
- (2) 通过<srai>标签转到了问题“我方/红方/我们”;
- (3) 提取参数项’search’和’domain’传递给系统;
- (4) 通过“我方/红方/我们”条目里的<condition>标签和标签找到访问数据库的处理模块<das>;
- (5) 增加参数项’side’;
- (6) 通过数据库访问索引表生成数据项 Target_Column, Table_X 等;
- (7) 生成完整 SQL 语句;
- (8) 生成文本答案红方目前损失了<das>架飞机;
- (9) 返回用户结果。(见图 3)

```

chat chat chat chat chat chat
C:\Anaconda2\python.exe C:/Users/Administrator/Pycha
Loading std-startup.xml... done (0.02 seconds)
Loading wargame-faq.aiml... done (0.00 seconds)
Loading wargame-das.aiml... done (0.00 seconds)
Enter your message >> 红方现有多少艘潜艇
红方现有 M 艘潜艇。
Enter your message >> X飞机的航程是多少
X飞机的航程为 Y 公里。
Enter your message >> |
  
```

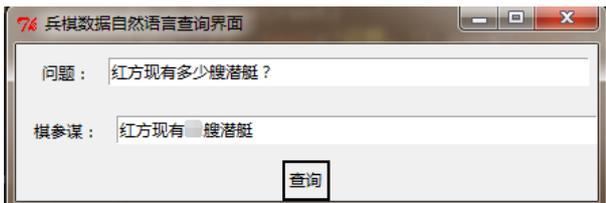


图 3 问答系统演示

Fig. 3 QA system demonstration

4 结论

智能问答系统为兵棋系统的进一步改进提供了一个新的发展方向, 即使不用语义分析、情感分析等高端的自然语言处理技术, 只需要部分自然语言规则匹配技术, 也可以让问答系统减轻演习调理人员的工作量, 使数据查询变得简单, 加快兵棋推演的进程^[11-12]。2016 年的双十一期间, 很多淘宝商家的“小二”就是由聊天机器人充当的, 很多顾客还茫然不知, 这当然由于机器人越来越“拟人化”, 但更多的是由于通过建立简单的诸如“身高、体重”到“尺码、型号”的规则匹配数据映射关系, 大量的咨询问题得到自动回复和解决, 以至于不少网购顾客觉得“小二们”如此智能, 甚至无所察觉。

但同时我们也应看到, 兵棋问答系统的智能化之路仍然任重道远, 主要表现以下方面:

- (1) 收集的问答数据有限, 没有编写领域词库, 无法进行同义词的扩展, 影响系统的召回率。
- (2) 对于涉及兵棋系统操作和运行规则的问题很难批量导入到 AIML 知识库, 目前也无法支持富媒体答案(图片、动画)的编写。
- (3) AIML 语言的多轮对话和学习功能没有开发, 系统也没有采用中文分词和句法分析的步骤。

下一步, 系统改进的重点要放在组合多种问答系统技术, 改进问答系统对兵棋系统操作和运行规则类问题的回答效能。

参考文献:

- [1] 胡晓峰, 罗批. 战争系统复杂性与战争模拟 [J]. 国防科技, 2007, 28(2): 6-11. (Hu X F, Luo P. War System Complexity and War Simulation [J]. National Defense Science & Technology, 2007, 28(2): 6-11.)
- [2] 胡晓峰, 司光亚, 吴琳, 等. 战争模拟原理与系统 [M]. 北京: 国防大学出版社, 2009. (Hu X F, Si G Y, Wu L, et al. War Simulation Principles and Systems [M]. Beijing, China: NDU Press, 2009.)
- [3] 何世柱, 张元哲, 刘康, 等. 智能问答技术 [J]. 中国人工智能学会通讯, 2016, 6(1): 6-11. (He S Z, Zhang Y Z, Liu K, et al. Intelligent Question Answering Technology [J]. CAAI Communication, 2016, 6(1): 6-11.)

(下转第 2506 页)