

DOI: 10.20189/j.cnki.CN/61-1527/E.202602011

作战仿真想定行动动态调整系统设计与应用

董志明¹, 胡忠奇^{1,2*}, 李冠辰³, 高建成¹

(1. 陆军兵种大学, 北京 100072;

2. 陆军某部队, 洮南 137100;

3. 陆军某部队, 北京 100000)

摘要: 针对陆军战术级作战仿真想定中作战行动参数调整效率较低的问题, 设计了基于大语言模型的作战行动动态调整系统。该系统由语音指令解析、调整指令生成、行动参数调整3个核心模块组成, 语音指令解析模块引入基于词性标注的关键信息抽取规则, 从仿真技术人员语音指令中准确抽取作战单位、时间、地域、行动方式等关键要素; 调整指令生成模块采用融合语义理解与术语匹配的混合检索策略, 在调整指令模板库中检索最优模板, 生成调整指令文本并自动转换为JSON格式代码; 行动参数调整模块通过基于时间关联的检索算法, 实现调整指令代码与对应的作战行动实例代码的映射替换。实验结果表明: 在使用特定测试指令集的情况下, 系统的关键信息抽取准确率达到94.5%。调整指令模板检索准确率达到93.5%。相较于传统人工调整方式, 作战行动参数调整效率提升显著, 平均耗时减少35.6%。该系统可以为作战仿真想定的作战行动参数快速调整提供有效的技术支撑。

关键词: 作战仿真想定; 作战行动参数调整; 大语言模型; 提示工程; 检索增强生成

中图分类号: TP391.9

文献标志码: A

开放科学标识码(OSID):

文章编号: 2097-2741(2026)02-0114-15



听语音
与作者
聊科研

作战仿真想定以作战想定文书为依据, 由仿真技术人员对文书内容进行细化, 并借助想定编辑软件实施数字化处理, 将自然语言描述的作战文书转换为XML、JSON等结构化计算机语言, 形成支撑作战仿真系统运行的初始化数据集。其中, 作战行动作为作战仿真想定的核心组成, 涉及作战单位、时间、地域、目标、行动方式以及关键点位坐标等多类参数的协同配置, 其调整的灵活性与准确性直接影响仿真推演效果。在当前主流想定编辑软件中, 作战行动参数调整需要通过逐层遍历菜单、手动定位与输入的方式来完成, 操作流程较为繁琐, 有时还会打断仿真技术人员战术思维的连续性^[1-2]。

目前, 相关研究多专注于作战仿真想定行动

从无到有的生成问题, 庞宁等^[3]针对单一大语言模型在作战行动生成任务中的能力局限, 提出一种基于混合智能体的生成方法; 曹占广等^[4]围绕作战行动序列编排复杂与执行控制困难的问题, 提出一种基于复合约束条件的作战行动序列生成与执行控制方法; 钟赞等^[5]聚焦空中作战行动过程设计研究, 结合动态影响网理论和改进遗传算法, 提出一种优选方法。然而, 针对作战仿真想定行动参数的动态调整, 尚未有研究人员开展系统性探索。

近年来, 随着大语言模型技术^[6]的迅速发展, 以及工具接口调用^[7]、提示工程^[8]、检索增强生成^[9]等关键技术的创新突破, 大语言模型的语义理解、任务推理与智能生成能力大幅提升, 为解决上述问题提供了技术突破口。通过构建基于大语

收稿日期: 2025-10-22

修回日期: 2026-01-04

录用日期: 2026-01-15

基金项目: 军事类研究生重点资助课题(XT2024B2086)

第一作者: 董志明(1977—), 男, 教授, 主要从事装备论证与仿真研究。E-mail: dongzhiming10@163.com

***通信作者:** 胡忠奇(1992—), 男, 硕士研究生, 主要从事作战仿真研究。E-mail: 625452456@qq.com

引用格式: 董志明, 胡忠奇, 李冠辰, 等. 作战仿真想定行动动态调整系统设计与应用[J]. 火箭军工程大学学报, 2026, 40(2): 114-128. DONG Zhiming, HU Zhongqi, LI Guanchen, et al. Design and application of a dynamic adjustment system for operational simulation scenario actions[J]. Journal of Rocket Force University of Engineering, 2026, 40(2): 114-128.

言模型的作战仿真想定行动动态调整系统, 仿真技术人员能够以自然语言交互的方式直接对作战行动参数进行精准调整, 将传统“查找界面、输入参数”的复杂操作流程, 简化为“发出指令、确认结果”的语音交互方式。

1 系统设计

1.1 总体架构

系统由语音指令解析模块、调整指令生成模块和行动参数调整模块3个核心模块构成, 其总体

架构如图1所示, 展示从语音指令输入到行动参数调整的端到端数据流以及各模块与外部系统的数据接口。

1.2 交互关系

1) 仿真技术人员与系统的交互关系

系统接收仿真技术人员的语音指令, 生成初步调整指令文本并返回给仿真技术人员, 由仿真技术人员进行参数填充与结果确认, 并返回系统, 由系统形成最终的调整指令文本。交互关系详情如表1所示。

2) 系统内部的交互关系

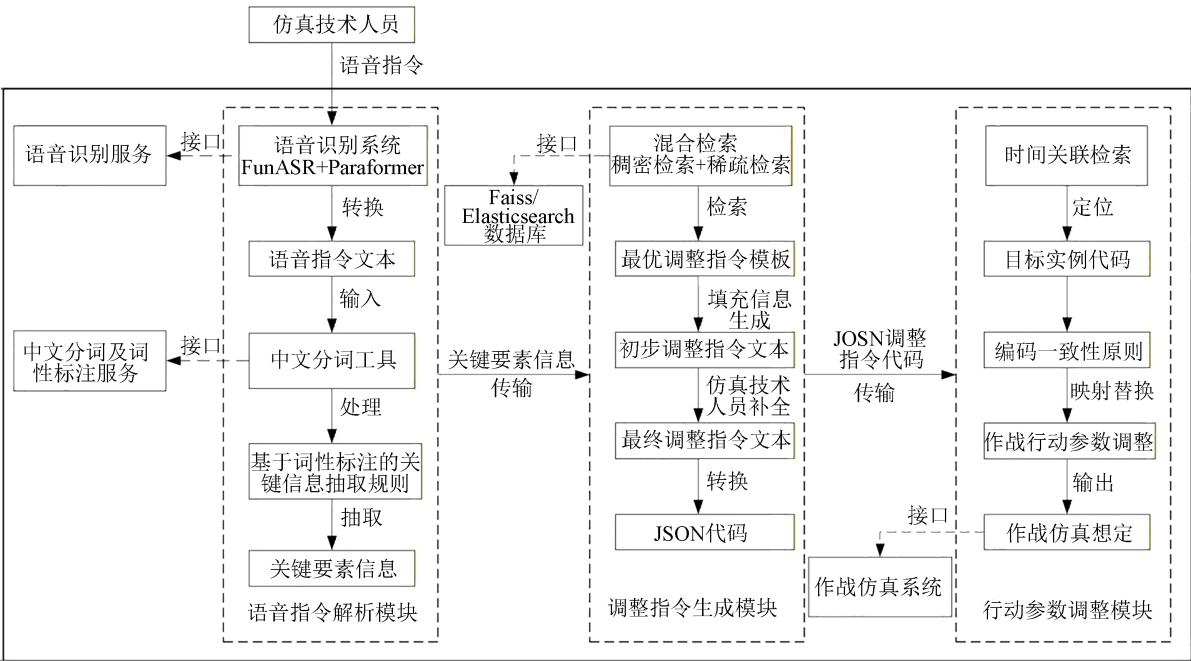


图1 系统总体框架

Fig. 1 Overall system framework

表1 仿真技术人员与系统的交互

Table 1 Simulation technician-system interaction		
输入方	接收方	信息内容
仿真技术人员	系统	语音指令
系统	仿真技术人员	初步调整指令文本
仿真技术人员	系统	最终调整指令文本

系统内部的信息交互主要由2个部分构成: ①语音指令解析、调整指令生成与行动参数调整3个核心模块之间的有序协作, 通过API接口实现数据交换, 依次完成从语音解析、指令生成到参数调整的完整处理流程, 最后经由API接口驱动作战仿真想定系统完成更新; ②各模块与外部支撑系统之间的协同调用, 以大语言模型作为系统的智能调度中枢, 通过API接口调用语音识别、分

词标注、模板检索等外部服务, 实现对原始语音指令解析、关键信息抽取与最佳调整指令模板检索, 为核心业务流程提供能力支撑。具体交互关系如表2所示。

1.3 核心流程

1) 语音指令解析。仿真技术人员通过语音下达指令后, 语音指令解析模块调用语音识别系统将音频转换为文本, 随后对其进行中文分词、词性标注与基于规则的关键信息抽取等处理, 从而获取关键要素信息, 如作战单位、时间、行动方式、目标等。

2) 调整指令生成。调整指令生成模块以抽取的关键要素信息为查询条件, 采用混合检索策略从模板库中检索最优模板, 并将关键信息填充至模板中, 生成初步的调整指令文本。经仿真技术

人员确认与补全后，自动转换为标准化的 JSON 格式代码。

表 2 系统内部的信息交互
Table 2 Information interaction within the system

输入方	接收方	信息内容
语音指令解析模块	调整指令生成模块	关键要素信息
调整指令生成模块	行动参数调整模块	JSON 格式调整指令代码
行动参数调整模块	作战仿真想定系统	JSON 格式仿真想定代码
大语言模型	语音识别系统	语音识别请求
语音识别系统	大语言模型	语音识别结果
大语言模型	中文分词工具	分词与词性标注请求
中文分词工具	大语言模型	带词性标签的分词结果
大语言模型	Faiss/Elasticsearch 数据库	检索请求
Faiss/Elasticsearch 数据库	大语言模型	检索结果

3) 行动参数调整。行动参数调整模块解析 JSON 代码，通过基于时间关联的检索算法在作战行动仿真想定片段中精确定位待调整的行动实例代码，并依据编码字段一致性原则，将调整指令代码映射并替换至目标实例代码，最终完成想定作战行动参数的调整。

2 系统模块功能实现路径

2.1 语音指令解析功能

由于大语言模型不具备语音信号处理能力，因此，需要先通过 API 接口调用语音识别系统，将仿真技术人员的语音指令转换为可处理的指令文本；然后，再通过 API 接口调用中文分词工具，对指令文本进行分词和词性标注；最后，基于词性标注的关键信息规则，从指令文本中抽取关键信息。

2.1.1 语音识别及分词处理

为提升语音识别与转换的准确性，基于阿里巴巴达摩院开源的 FunASR 框架^[10]构建一套集成多模块的语音识别系统，在其核心的 Paraformer-large 语音识别模型^[11]基础上，融合框架内置的 fs-mn_vad 语音端点检测模型^[12]、punc_ct 标点恢复模型^[13]与 speech_ngram 热词增强模型^[14]，可以将语音指令准确转换为带标点符号的文本，在增强语

义连贯性与可读性的同时，也为后续指令解析与处理提供可靠基础。

得到语音指令文本后，大语言模型通过 API 接口调用 jieba 中文分词工具^[15]进行分词处理，为词性标注奠定基础。同时，基于军语、条令条例与战术教材等专业资料，构建一个涵盖 26 类共 8 587 个术语的词典作为外部词表提供给分词工具，提升分词工具对专业军事词汇的识别能力。

2.1.2 关键信息抽取

大语言模型本质上是基于大规模规范文本语料训练得到的大规模参数模型，在处理结构清晰、语义明确的文本语言方面具有明显优势^[16]。然而，仿真技术人员所使用的口语化指令往往包含大量省略内容和模糊表述，会给大语言模型的理解能力带来挑战，导致其难以从中有效抽取关键信息。因此，需要设计一套基于词性标注的关键信息抽取规则，引导大语言模型能够从口语化的指令文本中，按照预设的语义类别，系统地定位并抽取其中的关键信息。首先，在分词的基础上，为每个具备独立语义的军事词汇或短语赋予相应的词性标签，作战仿真领域常见词性标记如表 3 所示。然后，结合作战行动中常见的要素信息，如作战单位、作战时间、作战地域、作战方式等，共设计 21 条关键信息抽取规则。部分抽取规则示例如表 4 所示。

以语音指令文本“夺控分队沿 5 号高地、6 号高地向敌装甲步兵 2 连防御阵地发起攻击”为例，大语言模型依据基于词性标注的关键信息抽取规则从语音指令文本抽取关键信息如下：作战单位为夺控分队（nt）；作战方向为沿 5 号高地、6 号高地（ns）；作战目标为敌装甲步兵 2 连防御阵地（nt）；行动方式为攻击（v）。

2.2 调整指令生成功能

调整指令生成的准确性高度依赖于检索方法的设计。在调整指令模板检索过程中，单一索引键可能无法充分表达模板的细微差别。例如，索引键“攻击”和“攻占”在语义上较为接近，但其对应的调整指令模板内容却存在显著差异。若只采用基于表示的稠密检索方法^[17]，可能因这两个词在向量空间中位置接近，而导致返回错误的模板。相比之下，基于关键词的稀疏检索^[18]依赖于词项的精确匹配，即使词语语义相近，只要词项不同仍可有效区分。因此，在基于表示的稠密检索的基础上，引入基于关键词的稀疏检索，构建混合检索方法，以兼顾语义理解与术语匹配的需求。

表 3 作战仿真领域常见词性标记

Table 3 Common part-of-speech tags in operational simulation

标签	名称	解释	标签	名称	解释
n	名词	表示人、装备、阵地、天气、障碍工事, 如: 指挥员、坦克、防御阵地、风、一列桩铁丝网	ns	地名	表示地理名称或高地编号, 如: 张村、郭庄、123 高程点、6 号高地
nt	机构	表示作战单位或作战群, 如: 主攻分队、装甲步兵 1 连	nz	专有名词	表示专有名词, 如: 利剑行动、铸剑行动
v	动词	表示动作或行为, 如: 攻击、摧毁、占领	a	形容词	表示描述性质或状态, 如: 坚固的、敌方的、主要的
d	副词	表示修饰动词或形容词, 如: 迅速、全力	p	介词	表示方位或关系, 如: 沿、向、从、在
c	连词	表示连接词语或句子, 如: 和、或、但	m	数词	表示数量或顺序, 如: 11、12、第 1、两
q	量词	表示单位, 如: 个、枚、发	r	代词	表示代替名词, 如: 我方、我部、敌方、其
w	标点符号	表示符号, 如: , . ! ;	Time	时间	表示时间表达, 如: 今日、3 时、2025-01-01 03:00:00

表 4 关键信息抽取规则部分示例

Table 4 Partial examples of key information extraction rules

抽取类别	抽取规则	抽取示例
作战单位	①核心标签: 机构名词 nt(如部队、编队名称等); ②辅助标识: 敌我属性修饰词(如我方、敌、蓝军等); ③位置规则: 句首主语或动词前的执行主体; “由…担任”“…部队”等句式中的名词 nt; 作战单位描述专有句式(如“作战/攻击/保障/支援/侦察/夺控/警戒/伏击单位为…”); ④规则描述: 抽取完整军事单位名称 nt, 保留敌我属性前缀, 若单位名称含数字/字母编号需完整保留。	夺控分队; 敌坦克 2 连
作战时间	①核心标签: 时间词 TIME(如明日、8 时、12:00、立即等)或动词性时间描述 v(如集结后、占领后等); ②位置规则: 句首时间状语(如“明日 6 时, 1 连发起攻击”); 介词 p(如于、在等)引导的时间结构(如“于 8 时 30 分结束”); 作战时间描述专有句式(如“作战/攻击/保障/支援/侦察/夺控/警戒/伏击时间为…”); ③规则描述: 抽取显式时间词 TIME 及动词性时间描述 v, 优先保留成对时间点(开始+结束)。	8 时 30 分; 集结后立即
作战地域	①核心标签: 地名 ns(如杨树林地区、11 号高地、132 高地西侧等)+坐标 m(如(11、12)、(N 1° 1' 1"、E 1° 1' 1")等), 有时也存在只有地名 ns 的情况; ②位置规则: 动词后的地域描述(如“攻击地域为…”“机动至…”等); 作战地域描述专有句式(如“作战/攻击/保障/支援/侦察/夺控/警戒/伏击地域为…”); ③规则描述: 抽取动词 v 后的连续地名以及坐标序列。	攻击地域为杨树林地区; 机动至 1 号高地(1、1)西侧
作战方式	①核心标签: 名词 n(如并肩、钳形等)+动词 v(如攻击、突袭、包围等); ②位置规则: 方式状语“以”后的描述(如“以…方式”); 关注动词连用结构(如“实施…”); 作战方式描述专有句式(如“作战/攻击/保障/支援/侦察/夺控/警戒/伏击方式为…”); ③规则描述: 抽取“以”字后的作战方式名词 n, 或连续动词 v 构成的行动组合。	钳形攻击; 中路突破、两翼卷击
作战行动	①核心标签: 连续动词 v(如突袭、包围、穿插等); ②辅助标签: 方式状语 d(如密集火力、快速机动等); ③位置规则: 动词连用结构(如“发起攻击并夺取”); 识别“以…方式”后的 v 动词(如“以纵深突贯方式攻击”); 作战行动描述专有句式(如“…作战/攻击/保障/支援/侦察/夺控/警戒/伏击行动”“作战/攻击/保障/支援/侦察/夺控/警戒/伏击行动为…”); ④规则描述: 抽取连续动词 v 构成的核心行动序列, 若含方式状语需与动词 v 组合提取(如“以…方式”+v); 避免提取单字动词 v(如“攻”需结合上下文判断是否为完整行动)。	攻击; 纵深穿插
作战目标	①核心标签: 地名 ns(如 1 号高地、张庄、立交桥等)或机构名 nt(如敌指挥所、蓝军炮兵阵地等)或名词 n(如通信枢纽、补给线等); ②位置规则: 动词 v 后的直接宾语(如摧毁敌指挥所、夺取 1 号高地、占领张庄等); 介词 p 引导的结构(如针对蓝军炮兵阵地、以通信枢纽为重点等); 目标描述专有句式(如“攻击/保障/支援/侦察/夺控/警戒/伏击目标为…”“首要攻击/保障/支援/侦察/夺控/警戒/伏击…”); ③规则描述: 抽取动词 v 或介词 p 后的显式目标, 优先保留军事术语, 目标可能伴随修饰词(如敌、蓝军等), 需与目标名称合并抽取。	敌坦克 2 连防御阵地; 敌通信枢纽

在确定检索方法后, 针对性构建调整指令模板库, 具体包括调整指令模板向量库与文本库 2 个子库。针对向量库, 构建以“索引键-目标值”为形式的关联索引; 针对文本库, 则构建基于指令模板文本的倒排索引, 以支持高效混合检索。基于上述设计, 调整指令生成功能实现路径为: 大语言模型基于语音指令解析模块抽取的关键信息, 运用混合检索方法在调整指令模板库中匹配最合

适的模板, 并将关键信息填充至模板中的对应占位符, 从而生成初步的调整指令文本。该文本在系统界面中呈现, 由仿真技术人员进一步补充与完善, 最终形成符合规范的 JSON 格式调整指令。

2.2.1 混合检索

采用基于表示的稠密检索 (L2) 和基于关键词的稀疏检索 (BM25) 线性组合的混合检索方法, 具体流程如图 2 所示。

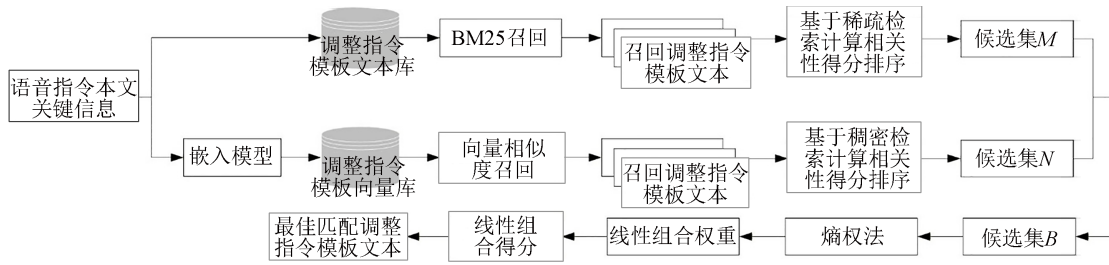


图 2 混合检索算法流程

Fig. 2 Flow of hybrid retrieval algorithm

混合检索算法核心流程如下:

① 计算基于表示的稠密检索相关性得分 S_d , 根据得分的高低顺序选取 Top- k_1 个模板, 构建候选集 $N = \{N_1, N_2, \dots, N_{k_1}\}$;

② 计算基于关键词的稀疏检索相关性得分 S_s , 根据得分的高低顺序选取 Top- k_2 个模板, 构建候选集 $M = \{M_1, M_2, \dots, M_{k_2}\}$;

③ 合并两个候选集, 并去除重复模板, 得到最终调整指令模板候选集 $B = \{B_1, B_2, \dots, B_{k_3}\}$, $k_3 \leq k_1 + k_2$ (在本实验中, $k_1 = k_2 = 5$), 基于熵权法自适应确定线性组合权重 w_d 与 w_s , 计算候选集 B 中各个模板的线性组合得分 S_h , 最后选取 S_h 最高的调整指令模板。

1) 基于表示的稠密检索相关性得分

利用 bce-embedding-base_v1 嵌入模型将从文本指令中抽取的关键信息编码为 768 维查询向量, 计算与索引键向量的欧氏距离, S_d 分值越小表示越相似。

$$S_d = \sqrt{\sum_i (q_i - d_i)^2} \quad (1)$$

式中: q_i 、 d_i 分别为查询向量与索引向量的第 i 维分量。

2) 基于关键词的稀疏检索相关性得分

采用 ATIRE BM25 算法, 使用 IDF 作为权重, 并防止 IDF 出现负值, 计算从语音文本指令中抽取的关键信息和调整指令模板倒排索引构中的关键词相关性得分, S_s 分值越高则表示越相似。

$$S_s(T, Q) = \sum_{i=1}^n I(t_i) \cdot \frac{f(t_i, T) \cdot (k+1)}{f(t_i, T) + k \cdot \left(1 - b + b \cdot \frac{l_T}{l_{avg}}\right)} \quad (2)$$

其中

$$I(t_i) = \log \frac{N - n(t_i) + 0.5}{n(t_i) + 0.5} \quad (3)$$

式中: Q 表示从语言文本指令中抽取的关键信息集合; T 表示被检索的调整指令模板; t_i 表示查询中的第 i 个词项; $I(t_i)$ 表示词项 t_i 的逆文档频率; $f(t_i, T)$ 表示词项 t_i 在模板 T 中出现的次数; l_T 表示模板 T 的长度; l_{avg} 表示所有模板的平均长度; k 、 b 为模型参数, 一般 $k \in [1.2, 2.0]$, $b \in [0, 1]$, 在本实验中, $k = 1.2$, $b = 0.75$ 。

3) 基于混合检索策略计算召回得分

首先, 对 S_d 和 S_s 进行最大最小归一化处理, 将其映射到 $[0, 1]$ 区间, 表达式为

$$S'_d = \frac{S_d - S_{d,min}}{S_{d,max} - S_{d,min}} \quad (4)$$

$$S'_s = \frac{S_s - S_{s,min}}{S_{s,max} - S_{s,min}} \quad (5)$$

式中: $S_{d,min}$ 为最小稠密检索分数; $S_{d,max}$ 为最大稠密检索分数; $S_{s,min}$ 为最小稀疏检索分数; $S_{s,max}$ 为最大稀疏检索分数。

然后, 将归一化后的 S'_d 和 S'_s 分别乘以相应的权重并相加, 计算线性组合得分 S_h 为

$$S_h = w_d S'_d + (1 - w_d) S'_s \quad (6)$$

式中: w_d 表示稠密检索分数归一化后的权重; $1 -$

w_d 表示稀疏检索分数归一化后的权重。

4) 基于熵权法确定线性组合权重

在基于语音指令关键信息的检索过程中, 不同查询对检索方式的依赖程度存在差异。语义复杂的指令更需稠密检索的深层理解, 术语精确的指令则更依赖稀疏检索的关键词匹配。这种查询条件的动态特性要求权重确定方法能够依据每次检索结果自适应地评估2种方法的重要性, 而非采用固定权重。

相较于依赖主观判断的层次分析法, 或旨在降维的主成分分析法, 熵权法^[19]可以基于指标数据的离散程度进行客观赋权。当查询中术语关键时, 稀疏检索得分在不同模板间差异显著, 熵权法自动赋予其更高权重, 确保精确匹配模板优先返回; 当查询语义模糊或存在词汇不匹配时, 稠密检索得分的离散度更高, 从而获得主导权重, 发挥其语义理解优势。因此, 熵权法是确定线性组合权重的优选方案。

假设验证集包含 k_3 个优化指令模板文本, 归一化后得分如表5所示。

表 5 熵权法验证集

Table 5 Entropy weight method validation set

文本序号	S'_d	S'_s
1	X_1	Y_1
2	X_2	Y_2
3	X_3	Y_3
$\vdots$	$\vdots$	$\vdots$
k_3	X_{k_3}	Y_{k_3}

在使用熵权法确定线性组合参数 w 时, 设定指标集合为 $S'_d = \{X_1, X_2, \dots, X_{k_3}\}$ 以及 $S'_s = \{Y_1, Y_2, \dots, Y_{k_3}\}$, 依据每个指标分别计算各优化指令模板在2个指标集合中的得分占比 p_{ij} 为

$$p_{ij} = \frac{S_{ij}}{\sum_{i=1}^n S_{ij}}$$

(7)

式中: S_{ij} 为指标集合; n 为调整指令模板文本的个数。

分别计算稠密检索信息熵 E_1 和稀疏检索信息熵 E_2 , 表达式为

$$E_j = -\ln^{-1}(n) \sum_{i=1}^n p_{ij} \ln(p_{ij})$$

(8)

式中: p_{ij} 为各调整指令模板在2个指标集合中的得分占比; n 为调整指令模板文本的个数 k_3 。

然后, 分别计算检索差异系数 g_j , 表达式为

$$g_j = 1 - E_j$$

(9)

最后, 确定稠密检索的权重 w_d 和稀疏检索的权重 w_s 分别为

$$w_d = \frac{g_d}{g_d + g_s}$$

(10)

$$w_s = 1 - w_d$$

(11)

进而计算线性组合得分 S_h , 选择 S_h 分数最高的调整指令模板。线性组合权重 w_d 与 w_s 通过上述熵权法自适应确定, 无需人工预设。

2.2.2 调整指令生成

检索得到调整指令模板文本后, 构建提示词引导大语言模型基于词性语义相同原则, 对齐模板中的占位符与抽取的关键词, 将关键词填充到模板中, 生成初步的调整指令文本, 提示词如表6所示。大语言模型将生成的初步调整指令返回给仿真技术人员, 后者依据实际需求对指令中的占位符内容进行补全, 形成最终的调整指令文本, 并自动转换为JSON格式, 以便于后续处理与执行。

依据仿真技术人员语音指令的完整性, 系统具备2种参数填充模式。①全自动填充模式, 当仿真技术人员语音指令中包含所有必需且能无歧义映射到调整指令模板的参数时, 系统可自动完成全部调整, 例如, 语音指令“你部(装甲步兵1连)的攻击方式从钳形攻击改为并肩攻击”, 系统自动抽取“装甲步兵1连”“攻击方式”“钳形攻击”“并肩攻击”等关键信息, 从调整指令模板库中检索得到攻击方式调整模板, 将关键信息自动填充至模板占位符中, 生成初步的调整指令文本并将其返回给仿真技术人员, 仿真技术人员确认无误后, 系统直接生成最终的调整指令文本, 全过程无需任何人工干预。②协同式填充模式, 当仿真技术人员语音指令描述模糊或参数数据不全时, 系统负责完成其所能处理的部分, 并引导仿真技术人员完成剩余部分, 例如, 语音指令“你部(侦察分队)加快行军速度”, 系统自动完成作战单位【侦察分队】和参数类型【行军速度】的识别与填充, 并在数值输入框位置高亮显示, 并提示“请输入具体速度值”, 需由仿真技术人员手动输入“45公里/小时”。

针对不同类型的参数信息, 系统提供差异化的填充机制: ①类似于【开始时间】【结束时间】【行军速度】【行军间隔】这种数值类的信息, 可以在实际界面对话框中手动输入填充; ②类似于【攻击目标点位1~4个】【左分界线点位2~7个】【右分界线点位2~7个】【突破点位1~7个】

表 6 生成初步指令的提示词

Table 6 Prompt words for generating preliminary commands

【角色设定】 1. 身份:想定作战行动参数调整指令初步生成助手 2. 核心能力: -指令模板解析能力:识别占位符,并做好词性标注 -地理坐标转换能力:识别地名,并在地理坐标库中检索到相应的经纬坐标 -关键信息填充能力:基于填充规则,将关键信息填充至调整指令模板
【任务描述】 基于词性相同原则,填充关键词到调整指令模板中,生成初步的调整指令,并返回给仿真技术人员
【任务流程】 1. 输入:基于语音指令文本抽取的关键信息、基于混合检索策略得到的调整指令模板 2. 处理阶段: ■ 指令模板解析:识别占位符,做好词性标注 ■ 地理坐标转换:识别地名,并在地理坐标库中检索到相应的经纬坐标 ■ 关键信息填充:将关键信息填充至指令模板 ■ 填充规则:①词性匹配:过滤出词性标签与占位符要求完全一致的关键词;②语义相似度计算:对词性不匹配但语义相似的关键词,计算与占位符的语义相似度,选择语义角色兼容且相似度>0.85的关键词 3. 输出:初步的调整指令文本
【输入示例】 1. 输入基于语音指令文本抽取的关键信息: 作战单位:夺控分队(nt) 作战方向:沿5号高地、6号高地(ns) 作战目标:敌装甲步兵2连防御阵地(nt) 行动方式:攻击(v) 2. 输入基于混合检索得到的调整指令模板: 【作战单位】担任攻击任务,于【开始时间】以【攻击方式】方式攻击【攻击对象】,攻击地域为【攻击目标点位1~4个】,攻击方向为【攻击方向点位2~7个】,左分界线为【左分界线点位2~7个】,右分界线为【右分界线点位2~7个点】,突破口位置为【突破点位1~7个】,当前任务线为【当前任务线点位2~7个】,后续任务线为【后续任务线点位2~7个】,于【结束时间】结束攻击任务
【输出示例】 【夺控分队】担任攻击任务,于【开始时间】以【攻击方式】方式攻击敌装甲步兵2连防御阵地,攻击地域为【攻击目标点位1~4个】,攻击方向为【5号高地坐标数值】,【6号高地坐标数值】,左分界线为【左分界线点位2~7个】,右分界线为【右分界线点位2~7个】,突破口位置为【突破点位1~7个】,当前任务线为【当前任务线点位2~7个】,后续任务线为【后续任务线点位2~7个】,于【结束时间】结束攻击任务

【当前任务线点位2~7个】【后续任务线点位2~7个】这种点位类的信息,可以通过在电子地图上选取坐标点位的方式填充;③类似于【攻击方式】【伏击方向】【警戒方向】【行军队形】这种选择性特别广泛的信息,则依赖已预设的信息由仿真技术人员自行选择,如【攻击方式】预设了“并肩攻击”“钳形攻击”“相向攻击”“向心攻击”等攻击方式,【伏击方向】预设了“0”“ $\pi/2$ ”“ π ”“ $3\pi/2$ ”“ 2π ”等方向。

值得注意的是,如果仿真技术人员给出的【攻击方式】为“中路突破、两翼卷击”之类的不在预设攻击方式内的信息,系统还会以【攻击方式】返回仿真技术人员,同时提示仿真技术人员依据预设选项重新选择,以确保参数的规范性。最终的调整指令代码示例如表7所示。

2.3 行动参数调整功能

在行动参数调整模块中,大语言模型识别调整指令代码中的“作战单位”“开始时间”和“结

束时间”等字段,在作战行动仿真想定片段中对应找到该作战单位在特定时间段的作战行动实例代码,并基于编码字段一致性原则,将调整指令代码映射至作战行动实例代码,从而完成对作战行动参数的动态调整。

2.3.1 时间关联检索

根据每个作战单位执行的作战行动的时间域都独一无二的特点,可以计算调整指令代码时间域与各个作战行动实例代码时间域之间的耦合度,通过耦合度的数值来定位调整指令代码对应的作战行动实例代码。因此,以Jaccard相似度系数为基础,提出一种基于时间关联的检索算法。在本系统中,作战行动实例代码时间域定义为作战行动实例代码中记录的计划执行时间区间,即从m_nBeginTime(开始时间)到m_nFinishTime(结束时间)的闭区间,该区间仅包含行动的核心执行阶段,不包括准备、收尾等附加时间,以此确保调整指令能精准匹配到行动的核心执行时段,

表 7 最终的调整指令代码示例
Table 7 Example of final adjustment command codes

调整指令文本
【夺控分队】担任攻击任务,于【2025-06-20 04:50:40】以【钳形攻击】方式攻击【敌装甲步兵2连防御阵地】,攻击地域为【[坐标1数值],[坐标2数值],[坐标3数值],[坐标4数值]】,攻击方向为【[坐标5数值],[坐标6数值]】,左分界线为【[坐标7数值],[坐标8数值],[坐标9数值]】,右分界线为【[坐标10数值],[坐标11数值],[坐标12数值]】,突破口位置为【[坐标13数值],[坐标14数值],[坐标15数值]】,当前任务线为【[坐标16数值],[坐标17数值],[坐标18数值]】,后续任务线为【[坐标19数值],[坐标20数值],[坐标21数值]】,于【2025-06-20 05:50:40】结束攻击任务”
调整指令代码
<pre>{ "m_nBeginTime": "2025-06-20 04:50:40", "m_nExecutorID": "45634351", "m_nFinishTime": "2025-06-20 05:50:40", "m_nOrderType": "60", "m_nReceiveID": "第1夺控队", "m_nSendID": "", "m_nSendTime": "2025-05-23 10:41:40", "iTargetArray1": 2, // 攻击方式, "m_nTargetID1": "32345217", "m_sPoint1": "[// 攻击方向,[坐标5数值],[坐标6数值]]", "m_sPoint2": "[// 左分界线,[坐标7数值],[坐标8数值],[坐标9数值]]", "m_sPoint3": "[// 右分界线,[坐标10数值],[坐标11数值],[坐标12数值]]", "m_sPoint4": "[// 突破口位置,[坐标13数值],[坐标14数值],[坐标15数值]]", "m_sPoint5": "[// 当前任务线,[坐标16数值],[坐标17数值],[坐标18数值]]", "m_sPoint6": "[// 后续任务线,[坐标19数值],[坐标20数值],[坐标21数值]]", "m_sArea1": "[// 攻击地域,[坐标1数值],[坐标2数值],[坐标3数值],[坐标4数值]]", "receiveSide": "" }</pre>

避免因时间区间定义过宽而导致误匹配。

Jaccard 相似度系数是衡量 2 个集合相似度的经典指标^[20], 定义为 2 个集合的交集大小与并集大小之比。将其应用于时间区间, 可将时间域 U_m 和 U_n 视为时间轴上的 2 个集合, 其耦合度 C 即为时间的 Jaccard 相似度, 可表示为

$$C = \frac{|U_m \cap U_n|}{|U_m \cup U_n|} \quad (12)$$

式中: 耦合度 C 值域为 $[0, 1]$, 值为 1 表示时间域完全重合, 值为 0 则表示无重叠。

将 Jaccard 相似度系数应用于作战仿真领域, 可将时间域视为时间轴上的集合, 其交集与并集的比例能够客观反映调整指令代码时间域与各个作战行动实例代码时间域的双向匹配程度, 相较于仅考虑指令时长或行动时长的单向度量方式, Jaccard 系数能够客观反映 2 个时间域之间的双向匹配程度, 有效识别时间线上整体重合度最高的行动实例。

因此, 在时间关联检索算法中, 可假设调整指令代码时间域为

$$T_{instr} = (t_s^{instr}, t_e^{instr}) \quad (13)$$

式中: t_s^{instr} 为调整指令代码中的开始时间; t_e^{instr} 为调整指令代码中的结束时间。

作战行动实例代码时间域为

$$T_{act} = (t_s^{act}, t_e^{act}) \quad (14)$$

式中: t_s^{act} 为作战行动实例代码中的开始时间; t_e^{act} 为作战行动实例代码中的结束时间。

二者时空耦合度的量化表达式为

$$K = \frac{|T_{instr} \cap T_{act}|}{|T_{instr} \cup T_{act}|} \quad (15)$$

式中: $|T_{instr} \cap T_{act}|$ 表示时间域交集长度; $|T_{instr} \cup T_{act}|$ 表示时间域并集长度。

$$|T_{instr} \cap T_{act}| = \max(0, \min(t_e^{instr}, t_e^{act}) - \max(t_s^{instr}, t_s^{act})) \quad (16)$$

$$|T_{instr} \cup T_{act}| = (t_e^{instr} - t_s^{instr}) + (t_e^{act} - t_s^{act}) - |T_{instr} \cap T_{act}| \quad (17)$$

基于时间关联的检索算法具体实现流程为：①大语言模型解析调整指令代码，识别其中的“编组 ID”“开始时间”和“结束时间”3 个字段名称；②利用“编组 ID”在作战行动仿真想定片段中检索 ID 一致的全部作战行动实例代码；③依次计算调整指令代码“开始时间”至“结束时间”和各个作战行动实例代码“开始时间”至“结束时间”的时间范围；④采用时间区间交集算法，动态计算调整指令代码时间域与各个作战行动实例代码时间域的时间耦合度；⑤基于耦合度的数值排序结果，选取“编组 ID”一致且数值最高的作战行动实例代码。若最高耦合度为 0，则判定为无匹配实例，系统将向仿真技术人员返回“未找到对应时间段内的作战行动”提示，需检查指令时间参数是否正确。此外，本系统未设置最低耦合度阈值，只要存在重叠即认为可能匹配，再依据“编组 ID 一致”与“耦合度最高”原则唯一确定目标实例。

2.3.2 行动参数调整

大语言模型在完成作战行动实例代码定位后，依托提示词基于编码字段一致性原则，实现调整指令代码与目标代码的映射替换，并提示仿真技术人员已完成作战行动参数调整工作。仿真技术人员可以在作战行动参数动态设置界面查看参数的更新情况，检查无误后保存调整后的作战仿真想定。

基于编码字段一致性原则是指，系统在参数调整过程中严格遵循预设的代码结构和字段命名规范，确保调整指令代码与作战行动实例代码之间建立准确的对应关系。该原则要求两类代码在数据结构、字段命名和数值格式等方面保持统一，从而保证调整指令能够精准定位到想定中的目标参数。通过基于编码字段一致性原则，系统实现从调整指令代码到作战行动实例代码的自动化映射替换，既可以保障参数调整的准确性，又能够维护仿真想定数据结构的完整性和系统运行的可靠性。

3 实验验证

3.1 系统环境

系统的大模型服务基于 vLLM 推理框架，实现 DeepSeek-R1-70B-AWQ 模型的本地化部署，后端采用 FastAPI 构建标准化 RESTful 接口，提供统一服务接入。硬件平台配置 4 张 NVIDIA

RTX 4090 显卡，3 张专用于承载 DeepSeek-R1-70B-AWQ 模型推理任务，另外 1 张则用于承载语音识别与向量检索服务。同时，为了支持高效的混合检索，采用 Faiss 向量数据库^[21]与 Elasticsearch 文本数据库^[22]分别存储与管理调整指令模板的向量化索引与倒排索引，实现语义匹配与关键词检索的高效协同。

选用 DeepSeek-R1-70B-AWQ 作为系统底座模型，主要考虑其在复杂推理任务中展现出的较强能力。该模型采用的思维链技术架构有助于准确解析仿真技术人员的语音指令，70 B 级别的参数规模为理解军事领域的专业术语和复杂语义关系提供了必要基础，通过 AWQ 量化技术处理，模型在保持核心性能基本完整的同时，显存占用量降至约 40 GB 水平。在模型选型方面，相较于 GPT-4 等闭源模型，其更符合内网部署的安全需求；相较于参数量相近的 Llama 3 等通用模型，其在复杂逻辑推理方面表现更佳；相较于参数规模较小的模型，其性能上限更能满足作战行动参数动态调整这一复杂任务的要求。此外，该模型在预训练阶段充分涵盖中英文语料，具备良好的双语理解能力，能够准确处理不同语言的语音指令。

值得注意的是，系统通过提示工程技术和检索增强生成技术，已经为 DeepSeek-R1-70B-AWQ 模型构建了清晰的任务流程框架和相关规则约束。该底座模型凭借其庞大的参数量，已具备足够的能力遵循此类精细指引，可以较好地完成作战行动参数动态调整任务。此外，由于微调训练需要庞大的 GPU 内存和计算时间，并且需要构建大规模、高质量的作战仿真领域指令数据集，消耗大量的人工成本；若仅基于小规模数据进行微调，则容易导致大语言模型过拟合，反而损害其任务泛化能力。基于上述原因，本实验未对 DeepSeek-R1-70B-AWQ 模型进行微调训练。

3.2 实验设计与结果分析

为评估系统的综合性能，围绕语音指令解析、调整指令生成与行动参数调整等核心环节，设计 4 项针对性评估指标、1 套测试指令集以及系列消融和对比实验，旨在定量验证系统各功能模块的性能表现及其相较于传统方法的综合优势。

3.2.1 评估指标

为全面评估系统的综合性能，选取关键信息抽取准确率、调整指令模板检索准确率与参数调整效率 3 项性能评估指标，分别用于评估系统对语音指令关键信息的抽取能力、对调整指令模板的

检索能力以及对作战行动参数的调整效率。同时, 引入统计显著性检验作为关键的推断性指标, 旨在验证不同对比组之间性能差异的统计学意义, 确保实验结论的科学性与可靠性。

1) 关键信息抽取准确率

采用关键信息抽取准确率评价基于词性标注的抽取规则有效性, F_1 分数能更均衡地评估抽取的准确性和完整性, 其表达式为

$$F_1 = 2 \times \frac{P \times R}{P + R} \quad (18)$$

式中: $P = \frac{\text{正确抽取的实体数量}}{\text{方法抽取的实体总数量}}$;
 $R = \frac{\text{正确抽取的实体数量}}{\text{标准答案中的实体总数量}}$ 。

2) 调整指令模板检索准确率

通过调整指令模板检索准确率评估混合检索方法性能, 系统一次性找到最匹配模板的能力是检索性能的核心指标, 其表达式为

$$A = \frac{N_c}{N_t} \quad (19)$$

式中: A 为调整指令模板检索准确率; N_c 为返回的 TOP-1 模板为正确答案的任务数; N_t 为总任务数。

MRR 是一个在信息检索领域广泛使用的评估指标, 用于衡量检索系统的排序质量^[23]。其不仅关注“是否找到了正确答案”, 更关注“正确答案被排在第几位”。正确答案排名越靠前, 则 MRR 得分越高, 其表达式为

$$\text{MRR} = \frac{1}{H} \sum_{i=1}^H \frac{1}{r_i} \quad (20)$$

式中: MRR 表示正确的调整指令模板在检索列表中的排名分数; H 表示查询的总任务数; r_i 表示对于第 i 个查询, 正确的调整指令模板在检索结果列表中的位置。

3) 参数调整效率

以单次参数调整平均耗时为核心度量, 计算作战仿真想定行动动态调整系统相对于传统人工调整方式的效率提升幅度, 可表示为

$$\eta = \frac{|T_t - T_s|}{T_t} \quad (21)$$

式中: η 表示效率提升幅度; T_t 表示传统人工方法的单次参数调整平均耗时; T_s 表示系统的单次参数调整平均耗时。

4) 统计显著性检验

为避免实验结论仅源于数据对比而缺乏统计推断支撑, 特引入统计显著性 p 值作为核心评估指标之一。本研究采用配对样本 t 检验^[24], 在显著性水平 $\alpha = 0.05$ 的标准下, 对上述定量指标的组间差

异进行显著性检验。配对 t 检验并通过计算 t 统计量来得到 p 值, 其计算过程如下。

设 $x_{A,i}$ 和 $x_{B,i}$ 分别代表第 i 个样本在 A 组和 B 组方法下的观测值, 共有 n 个配对样本。

首先, 计算每对观测值的差值为

$$o_i = x_{A,i} - x_{B,i} \quad (22)$$

而后, 计算差值的均值和标准差分别为

$$\bar{o} = \frac{1}{n} \sum_{i=1}^n o_i \quad (23)$$

$$S_o = \sqrt{\frac{\sum_{i=1}^n (o_i - \bar{o})^2}{n-1}} \quad (24)$$

随后, 计算 t 统计量为

$$t = \frac{\bar{o}}{S_o / \sqrt{n}} \quad (25)$$

该 t 统计量服从自由度为 $v = n - 1$ 的 t 分布, p 值即为在此 t 分布下, 观测到当前 t 统计量或更极端的概率, 通过查询 t 分布表得到。

因此, 显著判断标准为:

当 $0.01 \leq p < 0.05$ 时, 认为组间差异显著;

当 $p < 0.01$ 时, 认为组间差异极为显著;

当 $p \geq 0.05$ 时, 认为组间差异不显著。

3.2.2 测试指令集

在作战仿真想定中, 作战行动具体涵盖 3 种参数类型。①数值型参数: 如作战时间、行军速度、行动时间、兵力数量等数值参数; ②行动型参数: 如攻击方式、作战方向、行军队形等需要从预定义战术选项选择的配置参数; ③点位型参数: 如目标点位、分界线、航线节点等空间位置参数。测试指令集既要包含基础的单参数调整指令, 也要涵盖需要多参数调整的复杂指令。为此, 依托 1 份陆军战术级作战仿真想定, 通过 3 种类型参数的有效组合, 形成 8 类具有代表性且难度均衡的作战行动参数调整任务, 每类任务下设 20 条语义相近但具体内容各异的测试指令, 总计 160 条任务实例, 并以此构成测试指令集, 其部分示例如表 8 所示。

3.2.3 消融实验

1) 关键信息抽取消融实验

主要用于验证基于词性标注的关键信息抽取规则的有效性。为此, 设置 4 个实验组, 分别是基于词性标注的关键信息抽取规则的 A 组、基于少样本提示的 B 组、基于基础提示词的 C 组和基于关键信息抽取模型的 D 组, 其中, A、B、C 组均使用相同的底座模型 DeepSeek-R1-70B-AWQ,

B组的提示词包含3~5个提示示例，C组仅使用基础提示词，没有任何具体规则与提示示例，D组采用基于Qwen2.5-Coder-7B通过全参数微调训练得到的关键信息抽取模型。实验过程中，将测试指令集的语音指令文本分别输入至各实验组，记录各组抽取的关键信息内容，将输出与人工标注的标准答案进行比对，并计算各组的 F_1 分数，实验数据如表9所示。

表 8 测试指令集部分示例
Table 8 Examples of test command set

任务类型	测试指令	需调整参数类型
基础数值参数调整	你部(装甲步兵1连)的行军速度从每小时30公里提升到45公里	数值型参数
基础行动参数调整	你部(电子对抗连)的攻击方式从干扰压制改为欺骗干扰	行动型参数
基础点位坐标调整	你部(装甲步兵1连)的攻击目标点位从[坐标1]变更为[坐标2]	点位型参数
模糊参数调整	你部(装甲步兵1连)加快行军速度	数值型参数
多参数关联调整	你部(火力连)火力打击时间推迟到T+15分,持续时间缩短为20分	数值型参数
复杂参数调整(行动参数+点位参数)	你部(装甲步兵1连)伏击方向从 π 方向调整为 $3\pi/2$ 方向,并重新设定左战斗分界线点位	行动型参数 点位型参数
复杂参数调整(数值参数+行动参数)	你部(装甲步兵1连)渡河后,攻击方式由向心攻击改为钳形攻击	数值型参数 行动型参数
复杂参数调整(多单位多参数)	你部(装甲步兵1连)与坦克2连,在T+30分同时向1号高地发起并肩攻击	数值型参数 行动型参数 点位型参数

表 9 关键信息抽取对比实验数据
Table 9 Comparative experimental data of key information extraction

实验分组	精确度/%	召回率/%	F_1 /%	与A组显著性检验
A	95.2	93.8	94.5	基准
B	89.5	87.2	88.3	$p<0.01$
C	85.3	80.1	82.6	$p<0.01$
D	92.1	94.5	93.3	$p>0.05$

由表9可以看出，A组的 F_1 分数显著高于B组和C组， p 值均小于0.01，表明基于词性标注的关键信息抽取规则在性能上明显优于少样本提示方

法及基础大语言模型直接抽取方法。这一结果从统计学上证明了将作战仿真领域知识固化为结构化规则，并引导大语言模型执行，相较于依赖模型从少量样本中自行领悟或仅凭其自身能力，能更为有效地完成关键信息抽取任务。同时，A组的 F_1 分数与D组相当，而D组的 p 值大于0.05，未达到显著性水平，说明无需投入大量计算资源与数据标注成本训练专用模型，通过精心设计的提示工程与规则引导，直接使用大型底座模型即可达到相同的效果，为系统模块功能的实现提供了更具可操作性与可扩展性的技术路径。

2) 调整指令模板检索消融实验

主要用于评估混合检索方法在调整指令模板检索任务中的有效性，验证混合检索方法能否克服稠密检索或稀疏检索等单一检索方法的局限性，在检索准确率与排序质量上实现显著提升。

实验设置3个实验组，分别为基于稠密检索的A组、基于稀疏检索的B组和基于混合检索的C组。其中，A组单独使用Faiss向量数据库，基于查询与模板的语义向量相似度进行检索；B组单独使用Elasticsearch及ATIRE BM25算法，基于查询与模板的关键词匹配程度进行检索；C组采用线性加权方法，综合稠密检索与稀疏检索的得分，以实现语义匹配与术语精确性的高效协同。在实验过程中，将测试指令集中的关键信息作为查询条件，将同一组查询分别输入至3个实验组，记录各组返回的TOP-1模板结果，计算检索准确率和平均倒数排名的数值，结果如表10所示。

表 10 调整指令模板检索对比实验数据
Table 10 Comparative experimental data of adjustment command template retrieval

实验分组	检索方法	准确率/%	MRR	与C组显著性检验
A	稠密检索	85.5	0.801	$p<0.01$
B	稀疏检索	88.0	0.832	$p<0.05$
C	混合检索	93.5	0.901	基准

由表10可以看出，C组的检索准确率达到93.5%，较A组和B组分别提升8.0%和5.5%， p 值均小于0.05；C组平均倒数排名的数值为0.901，同样明显高于A组和B组。说明混合检索策略的性能优势具有统计学意义，其通过融合语义理解与术语精确匹配，能够实现检索精度与排序稳定性的显著提升。

通过案例分析进一步揭示混合检索的融合优

势。面对“攻占1号高地”这类查询，稀疏检索凭借精确术语匹配成功返回结果，而稠密检索因“攻占”与“攻击”的语义相似性产生误判；混合检索通过加权融合有效利用稀疏检索的精确性，规避语义混淆。在应对查询条件为“从左翼包抄”，而对应模板为“实施左翼迂回”等词汇不匹配场景时，混合检索则通过稠密检索的语义理解能力弥补稀疏检索的局限。说明混合检索在不同类型的查询场景下均能发挥协同优势，为系统提供更加精准有效的模板检索能力。

3.2.4 对比实验

为定量评估系统在作战行动参数调整任务中的作业效能，通过与传统人工调整方式进行对比，验证系统在任务效率方面的提升效果。实验数据如表11所示。

表 11 参数调整效率对比实验数据					
Table 11 Comparative experimental data of parameter adjustment efficiency					
任务类型	A组时间/s	B组时间/s	平均缩短时间/s	效率提升/%	<i>p</i> 值
基础数值参数调整	8.9	12.1	3.2	26.5	$p<0.05$
基础行动参数调整	7.5	14.6	7.1	48.6	$p<0.01$
基础点位坐标调整	10.2	15.8	5.6	35.4	$p<0.01$
模糊参数调整	14.8	16.3	1.5	9.6	$p>0.05$
多参数关联调整	13.1	20.5	7.4	36.1	$p<0.01$
复杂参数调整 (行动样式+坐标点位)	15.6	25.7	10.1	39.3	$p<0.01$
复杂参数调整 (作战时间+行动样式)	12.4	19.2	6.8	35.4	$p<0.01$
复杂参数调整 (多单位多参数)	17.1	30.5	13.4	43.9	$p<0.01$
所有任务平均时间	12.45	19.34	6.89	35.6	$p<0.01$
标准差	3.37	6.13	—	—	—

实验由10名具备作战仿真想定编辑经验且来自作战仿真想定开发一线团队的技术人员作为参与者，其中4人具备5年以上想定编辑经验，4人具有3~5年经验，2人为3年以下经验，这种经验

分布符合实际作战仿真团队的人员构成特点，也能够全面反映不同熟练度操作人员的使用效果。每位参与者需分别使用2种方法完成相同的测试任务集。A组使用作战想定行动动态调整系统，B组使用基于想定编辑软件的传统人工方式。实验过程中，通过屏幕录制与系统日志相结合的方式，精确记录每个任务的起始与结束时间。任务计时从仿真技术人员发出语音指令起，至参数确认调整完成止，确保时间记录的准确性与一致性。由表11可以看出，A组的8类作战行动参数调整任务平均时间为12.45 s，相较于B组缩短了6.89 s，*p*值小于0.01，整体效率提升35.6%。其中，在基础参数调整任务中，A组平均耗时较B组缩短约30%，尤其在涉及多参数联动的复杂任务中，A组效率提升更为明显，相较于B组，均达到35%以上，这是因为系统通过单次指令即可解析并应用所有参数，而人工方式需要在多个单位、多个参数栏间频繁切换。同时，A组在不同类型任务中的耗时标准差为3.37，远小于B组的6.13，表明系统的性能表现更稳定，受任务复杂度和操作人员个体差异的影响较小。

4 结论

针对传统人工调整作战行动参数操作繁琐、效率低下、易打断战术思维的问题，围绕陆军战术级作战仿真想定，综合运用工具接口调用、提示工程以及检索增强生成技术，构建基于大语言模型的作战仿真想定行动动态调整系统，通过语音指令解析、调整指令生成、行动参数调整等功能模块，实现作战行动参数的动态调整。通过实例测试和对比实验，证明该系统相较于传统人工方式，可以显著提高作战行动参数调整任务效率。

在实验中发现该系统仍存在2点不足。一是需要仿真技术人员在安静环境中与系统实施语音交互，由于语音识别模型很容易受到外界噪声的干扰，限制了系统的适用场景；二是系统只能智能化地执行作战行动参数动态调整任务，达不到完全智能效果。系统很依赖于仿真技术人员语音指令的具体性和完整性，语音指令中的作战行动要素如作战时间、作战点位、作战方式等越具体，系统运行效果越好，如果仿真人员下达的指令比较模糊，系统运行效果则较差，只比与传统人工方法略强。

因此，后续需要从3个方向对系统进行优化升

级。一是尝试开发或寻找性能更出色的语音识别模型,提高在噪声下对仿真技术人员语音指令的识别率;二是尝试用战斗条令条例训练 1 个大语言模型智能体,让智能体自动填补仿真技术人员指令中空缺的参数信息,使系统能够更好的理解并执行多步骤、带条件的复杂想定参数调整指令;三是尝试加入作战编组、战场环境和初始态势等参数调整功能,将系统从单一的作战行动参数调整功能扩展为想定全要素参数调整功能。

参考文献

- [1] 冯亚军,尹康银,王前,等.面向雷达兵指挥推演的想定生成系统设计与实现[J].空军预警学院学报,2021,35(3):212-215.
Feng Yajun, Yin Kangyin, Wang Qian, et al. Design and implementation of scenario generation system for radar troops command deduction[J]. Journal of Air Force Early Warning Academy, 2021, 35(3): 212-215.
- [2] 龚建兴,王子沐,杨奇龙.基于粒子群算法的训练仿真想定优化生成方法[J].系统仿真学报,2023,35(9):1860-1870.
Gong Jianxing, Wang Zimu, Yang Qilong. Training simulation scenario generation based on particle swarm optimization[J]. Journal of System Simulation, 2023, 35(9): 1860-1870.
- [3] 庞宁,肖卫东,赵翔,等.基于混合智能体的作战行动方案生成方法[J].指挥与控制学报,2025,11(2):181-190.
Pang Ning, Xiao Weidong, Zhao Xiang, et al. Course of action based on mixture-of-agents[J]. Journal of Command and Control, 2025, 11(2): 181-190.
- [4] 曹占广,于屏岗,尹宗润,等.基于复合约束的作战行动序列生成与执行控制方法[J].火力与指挥控制,2021,46(9):105-112.
Cao Zhanguang, Yu Pinggang, Yin Zongrun, et al. Generation and execution control method for course of action based on composite constraints[J]. Fire Control & Command Control, 2021, 46(9): 105-112.
- [5] 钟赞,万路军,张杰勇,等.区间不确定性下的空中作战行动过程优选方法[J].航空学报,2021,42(2):324282.
Zhong Yun, Wan Lujun, Zhang Jieyong, et al. Optimized selection method for air combat course of action with interval uncertainty[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(2): 324282.
- [6] 许鹏,刘梦雅,范新峰,等.人工智能算法在核应急中的应用综述[J].火箭军工程大学学报,2024,38(3):104-110.
Xu Peng, Liu Mengya, Fan Xinfeng, et al. Overview of application of artificial intelligence algorithms in nuclear emergency[J]. Journal of Rocket Force University of Engineering, 2024, 38(3): 104-110.
- [7] Schick T, Dwivedi-Yu J, Dessì R, et al. Toolformer: language models can teach themselves to use tools[EB/OL].(2023-02-09)[2025-10-19].<https://arxiv.org/abs/2302.04761>.
- [8] Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York, USA: ACM, 2020: 9459-9474.
- [9] Liu Pengfei, Yuan Weizhe, Fu Jinlan, et al. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing [EB/OL].(2022-07-28)[2025-10-19].<https://arxiv.org/abs/2107.13586>.
- [10] Gao Zhifu, Li Zerui, Wang Jiaming, et al. FunASR: a fundamental end-to-end speech recognition toolkit[EB/OL].(2023-05-18)[2025-10-19].<https://arxiv.org/abs/2305.11013>.
- [11] Gao Zhifu, Zhang Shiliang, Yan Zhijie, et al. Paraformer: fast and accurate parallel transformer for non-autoregressive end-to-end speech recognition[EB/OL].(2022-06-16)[2025-10-19].<https://arxiv.org/abs/2206.08317>.
- [12] Zhang Shiliang, Liu Cong, Jiang Hui, et al. Feed-forward sequential memory networks: a new structure to learn long-term dependency[EB/OL].(2015-11-28)[2025-10-19].<https://arxiv.org/pdf/1512.08301v1>.
- [13] Ling Tongtao, Lai Yutao, Chen Lei, et al. A small and fast BERT for Chinese medical punctuation restoration[EB/OL].(2025-06-28)[2025-10-19].<https://arxiv.org/html/2308.12568>.
- [14] Huang Kaixun, Zhang Ao, Yang Zhanheng, et al. Contextualized end-to-end speech recognition with contextual phrase prediction network[EB/OL].(2023-07-12)[2025-10-19].<https://arxiv.org/abs/2305.12493>.
- [15] 邢玲,程兵.基于结巴分词的领域自适应分词方法研究[J].计算机仿真,2023,40(4):310-316.
Xing Ling, Cheng Bing. Research on domain adaptive word segmentation method based on Jieba word segmentation[J]. Computer Simulation, 2023, 40(4): 310-316.
- [16] 王昱婷,陈波,闫强,等.基于问题导向提示学习和多路推理的检索增强生成问答[J].计算机工程与应

- 用, 2025, 61(12): 120-128.
- Wang Yuting, Chen Bo, Yan Qiang, et al. Retrieval-augmented question-answering generation based on question-oriented prompt learning and multi-channel reasoning[J]. *Computer Engineering and Applications*, 2025, 61(12): 120-128.
- [17] Karpukhin V, Oguz B, Min S, et al. Dense passage retrieval for open-domain question answering [C]// *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online. Association for Computational Linguistics, 2020: 6769-6781.
- [18] Robertson S, Zaragoza H. The probabilistic relevance framework: BM25 and beyond[J]. *Foundations and Trends in Information Retrieval*, 2009, 3(4): 333-389.
- [19] 邹龙, 杨东, 王以松, 等. 基于结构熵的交通流量实时预测模型[J]. *计算机工程与应用*, 2026, 62(7): 363-373.
- Zou Long, Yang Dong, Wang Yisong, et al. Real-time traffic flow prediction model based on structural entropy[J]. *Computer Engineering and Applications*, 2026, 62(7): 363-373.
- [20] 李顺东, 杜佳欣, 吴川宇, 等. 单云环境下强隐私保护的三维多重集相似度阈值精确查询方案[J]. *计算机学报*, 2025, 48(10): 2430-2449.
- Li Shundong, Du Jixin, Wu Chuanyu, et al. Strong privacy-preserving scheme for exact multi-dimensional multiset similarity threshold queries in single-cloud environments[J]. *Chinese Journal of Computers*, 2025, 48(10): 2430-2449.
- [21] Johnson J, Douze M, Jégou H. Billion-scale similarity search with GPUs[EB/OL]. (2017-02-28) [2025-10-19]. <https://arxiv.org/abs/1702.08734>.
- [22] Gormley C, Tong Z. *Elasticsearch: The Definitive Guide*[M]. California: O'Reilly Media Inc., 2015: 724.
- [23] White J, Fu Quchen, Hays S, et al. A prompt pattern catalog to enhance prompt engineering with ChatGPT[EB/OL]. (2021-11-21) [2025-10-19]. <https://arxiv.org/abs/2302.11382>.
- [24] 王晓东, 徐永杰. 金属封装陶瓷复合结构抗射流性能仿真[J]. *火箭军工程大学学报*, 2024, 38(2): 17-24.
- Wang Xiaodong, Xu Yongjie. Simulation research on jet penetration resistance of metal-packaged ceramic structures[J]. *Journal of Rocket Force University of Engineering*, 2024, 38(2): 17-24.

Design and Application of a Dynamic Adjustment System for Operational Simulation Scenario Actions

DONG Zhiming¹, HU Zhongqi^{1,2*}, LI Guanchen³, GAO Jiancheng¹

(1. Army Arms University of PLA, Beijing 100072, China;

2. A Troop of PLA Army, Taonan 137100, China;

3. A Troop of PLA Army, Beijing 100000, China)

ABSTRACT: To address the low efficiency of adjusting operational action parameters in army tactical-level operational simulation scenarios, a dynamic operational action adjustment system based on a large-language model was proposed. The system comprised of three core modules: voice command parsing, adjustment command generation, and action parameter adjustment. The voice command parsing module introduced key information extraction rules based on part-of-speech tagging to accurately extract critical elements such as operational units, time, location, and action methods from the voice commands of the simulation technicians. The adjustment command generation module employed a hybrid retrieval strategy that integrated semantic understanding and term matching to retrieve the optimal template from the adjustment command template library. It generated adjustment instruction text and automatically converted it into JSON format code. The action parameter adjustment module utilized a temporal association-based retrieval algorithm to map and replace the adjustment command code with the corresponding operational action in-

stance code. Under specific test command sets, the system achieves key information extraction accuracy of 94.5% and adjustment command template retrieval accuracy of 93.5%. Compared with traditional manual adjustment methods, the system significantly improves the efficiency of the operational action parameter adjustment, reducing the average time consumption by 35.6%. Experimental results demonstrate that the system provides effective technical support for rapid adjustment of operational action parameters in operational simulation scenarios.

KEYWORDS: operational simulation scenario; adjustment of operational action parameters; large-language model; prompt engineering; retrieval-augmented generation

(编辑: 于福兴)

专题征文——先进半导体与电驱电控新技术

为促进学术交流,推动先进半导体与电驱电控领域创新研究与应用发展,本刊拟开设“先进半导体与电驱电控新技术”专题,现面向社会各界开展论文征集工作,热忱欢迎本领域和相关交叉学科领域的专家、学者积极撰写相关文章,踊跃投稿。

征文范围包括但不限于以下方向:(1) 先进半导体材料与器件制备技术;(2) 仿高功率密度高效能电机设计;(3) 先进半导体基电驱系统拓扑与集成设计;(4) 低空装备先进电驱电控核心技术;(5) 仿智能化控制策略与技术;(6) 热管理与可靠性保障技术;(7) 仿新能源融合与工程化应用;(8) 测试与表征技术。

让我们以思想碰撞火花,用文字勾勒未来,共同推进先进半导体与电驱电控领域的进步与繁荣!

《火箭军工程大学学报》编辑部

二〇二六年四月