Dynamic Multi-Source Knowledge Augmented Intelligent Content System]

Authors: [Group 19：Zhu Ziyang, Tan Yifan, Sun Liuyi]

Student IDs: [224040092, 120090416,224040098]

Email: [224040092@link.cuhk.edu.cn, 120090416@link.cuhk.edu.cn, 224040098@link.cuhk.edu.cn]

Introduction

GameCopy Genius: AI Agent for Game Content Creation

In this work, we present GameCopy Genius, a specialized AI agent for gaming content generation that synergizes data-driven automation with creator expertise. Traditional LLM-based approaches face two critical limitations in gaming contexts:

I Generic Output Problem Off-the-shelf LLMs produce mechanically fluent but genre-agnostic content. And content Lacks awareness of evolving game metas, patch notes, or community slang.

II Engagement Gap Human-written scripts leverage tacit knowledge of "what makes content viral". Pure scraping-based systems miss nuanced humor/reference frameworks

Our Solution:

Machine-curated knowledge: Real-time data from 15+ official games blogger

Creator wisdom: Fine-tuning on 50K high-engagement scripts annotated by top Weibo creators

Motivation

In the current landscape of digital marketing for mobile games, especially titles like Honor of Kings (王者荣耀), crafting engaging and timely promotional copy has become increasingly essential. However, traditional methods of content creation are often manual, time-consuming, and disconnected from real-time user interests and trends. Our motivation stems from the need to streamline this process by combining intelligent topic detection, user intent understanding, and data-driven content generation. By integrating web scraping, knowledge enhancement, and large language models (LLMs), we aim to generate high-quality, context-aware promotional text that aligns with current hot topics and user preferences. Ultimately, our system serves to empower marketing teams with a scalable, automated, and intelligent solution for content creation in fast-evolving gaming communities.

Methodology

云-端协同架构
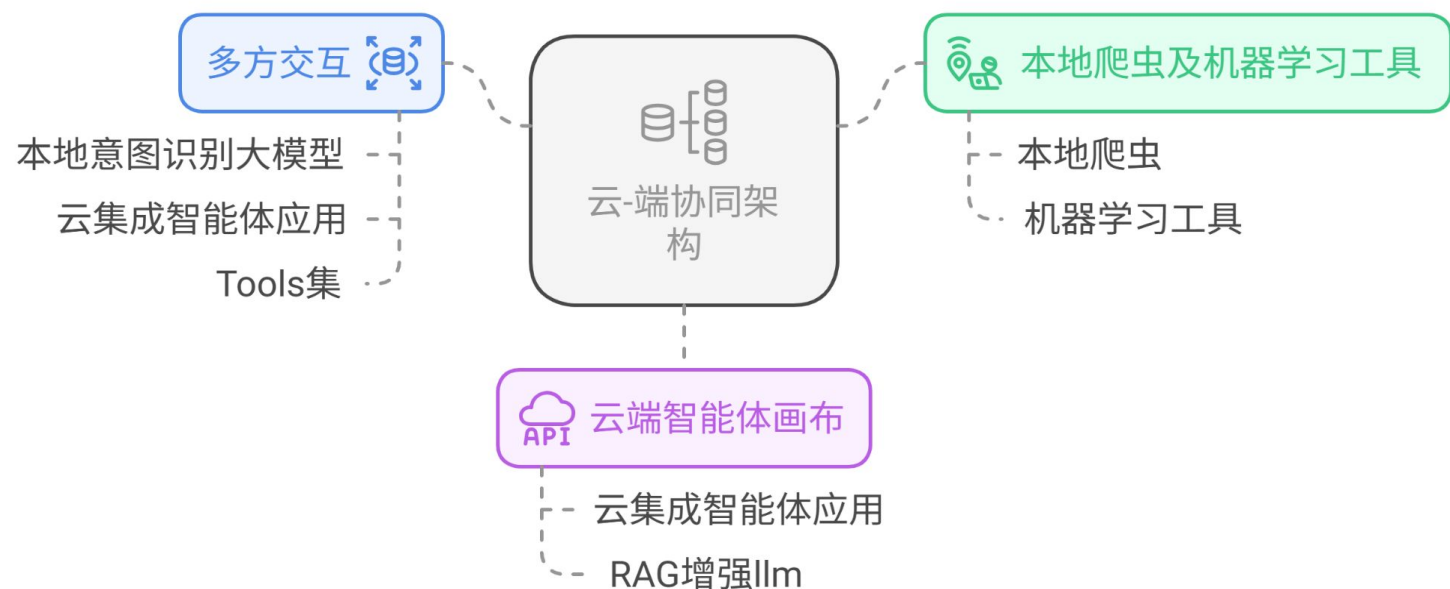
I Framework

- **Cloud-Edge Collaborative Architecture:** A layered design combining local crawlers and machine learning tools with a cloud-based agent canvas.
- **Multi-party Interaction:** Includes a local intent recognition large model, cloud-integrated agent applications, a fine-tuned large model for copywriting classification, and a suite of tools.

II. Core Methodology

Multi-source Dynamic Knowledge: Integrates a two-year historical knowledge base with real-time crawled popular copywriting

Fine-tuned Content Classification Model: Use labeled data by the llm to train a Bert model for classification

Machine Learning Methods: TF-IDF with word frequency analysis, Decision Tree like random forest and so on for classification, and LDA topic modeling—for keyword trend analysis and prediction.
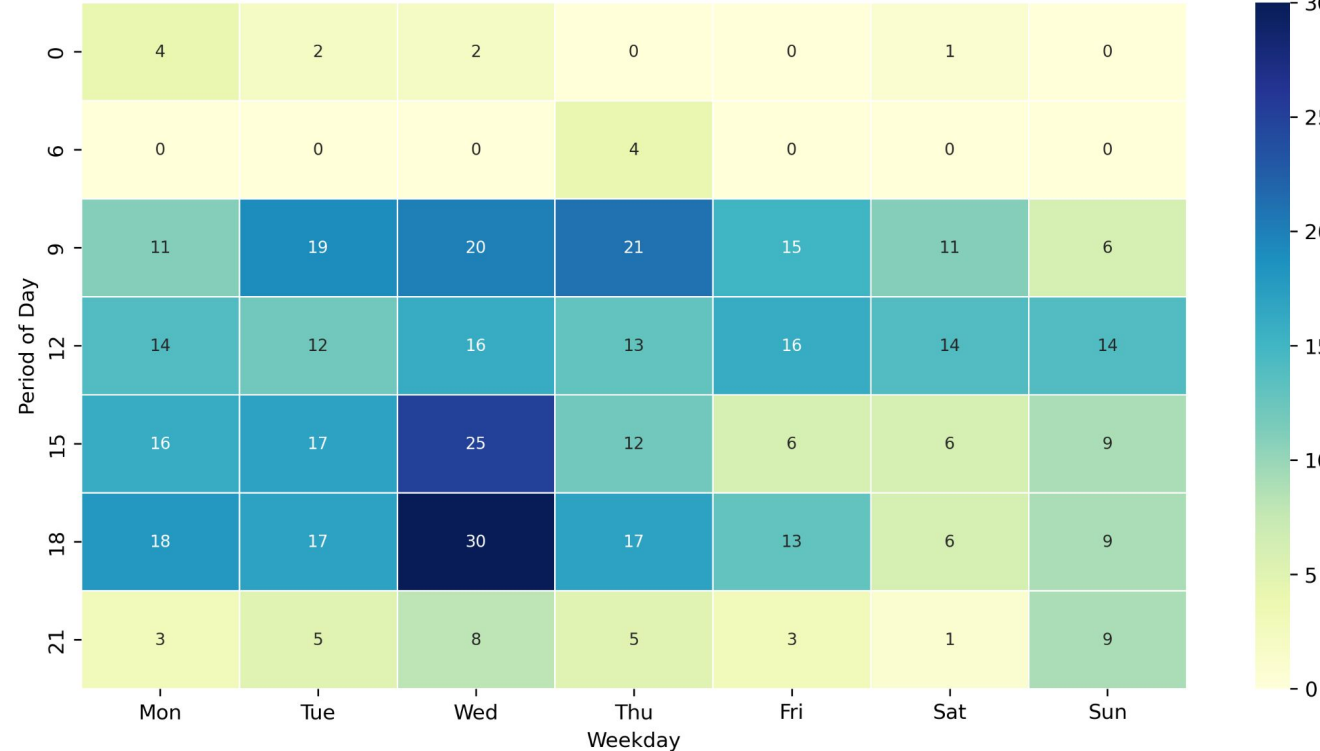
Online Posting Time Analysis: Real-time crawling of the latest copywriting updates the database, enabling active cluster feature analysis and providing users with a heatmap of optimal engagement time slots.

Results
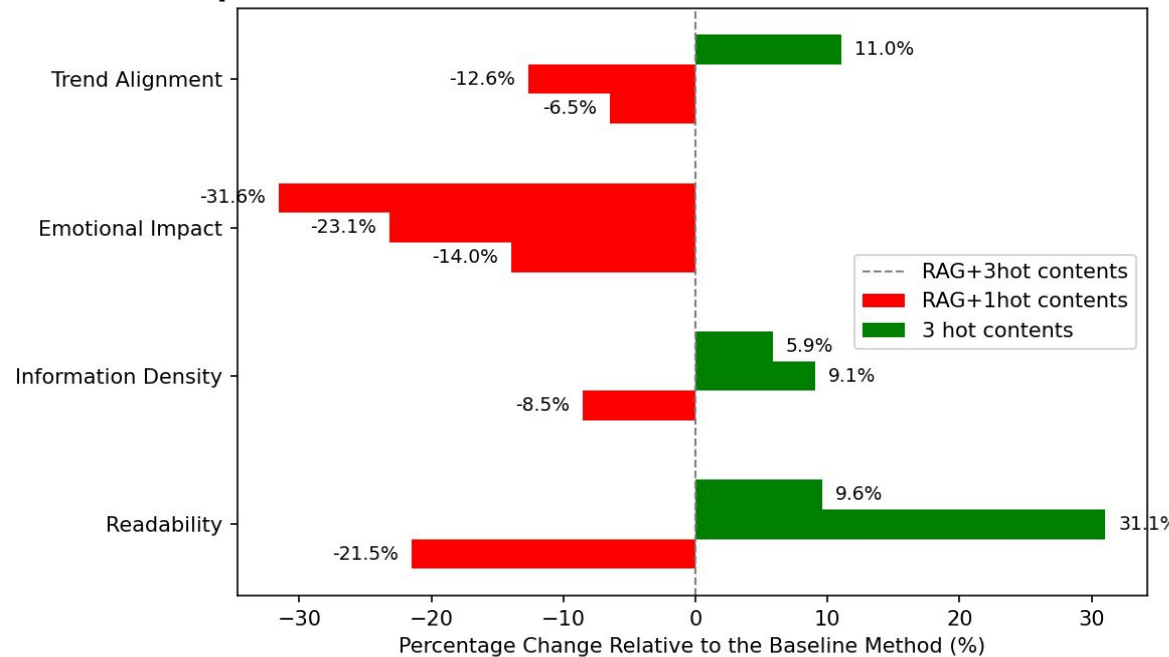
I. Evaluation Matrix:

Readability: considering average number of words per clause and ratio of adverbs and conjunctions per sentence. **Information Density:** verbs and nouns ratio. **Emotional Impact:** Based on the arousal dimension of Chinese EmoBank, a manually annotated Chinese dimensional emotion lexicon that measures the level of calmness or excitement expressed in a text. **Trend Alignment:** calculate coverage based on keywords from previously popular interactive contents.

II. Analysis:

- **RAG's Enhancement Effect:**
  RAG leads to notable improvements in Trend Alignment, Readability and Information Density, suggesting that the retrieval mechanism effectively captures the **structural patterns** and **stylistic consistency** of **historically high-performing copy**.
- **Marginal Effect of Popular Content:**
  The **diversity of examples** helps the model better capture **latent trend patterns**, but an excess of content may **compromise the semantic coherence** of the generated text.
- **Post Time Recommendation**:
  Based on the top 30% most high-interactive copy from the past six months, it is recommended to post between 12:00–18:00 on weekdays and 09:00–12:00 on weekends.