

北京大学学报(自然科学版)
Acta Scientiarum Naturalium Universitatis Pekinensis
doi: 10.13209/j.0479-8023.2015.045

古典诗词动画自动生成研究

阎思瑶 郑旭玲[†] 史晓东 郑发魁

厦门大学智能科学与技术系, 厦门 361005; [†] 通信作者, E-mail: xlzheng@xmu.edu.cn

摘要 以中国古典诗词为研究对象, 首次尝试将自然语言处理、计算诗学和计算机动画等相关研究成果相结合, 用于实现古典诗词动画的自动生成。首先对一首诗词进行自动句法语义分析, 获取其中的各种关键特征, 然后使用基于 SVM 和协同学习的分类器对其风格、题材以及时间进行自动判定, 据此自动为其选取恰当的场景脚本和各类素材, 进行场景规划, 并结合 Flash 动画技术实现自动脚本生成和动画生成, 自动为其合成一段声情并茂、视听交融、动静交错的多媒体动画短片。实验表明, 所提出的一系列方法初步解决了古典诗词的动画自动生成问题, 为后续研究提供了理论依据和实验基础。

关键词 古典诗词; 动画生成; SVM 分类; 共现分析

中图分类号 TP391

Study of Intelligent Multimedia Display System for Classic Chinese Poetry

YAN Siyao, ZHENG Xuling[†], SHI Xiaodong, ZHENG Fakui

Departement of Cognitive Science, Xiamen University, Xiamen 361005; [†] Corresponding author, E-mail: xlzheng@xmu.edu.cn

Abstract As the study of Chinese classical poetry, it is the first attempt to natural language processing, computing-related research poetics and computer animation combined to solve the classical poetry animation automatically generated. And it proposes a SVM-based method to collaboratively determine the ancient's style, theme and time. In addition, a method according to the results of time-based ancient scene classification is proposed and also basing on keywords analyzes the relationship between the use of co-occurrence of a subsequent generation of animation animated element supplement. According to the experiments, this paper show that the proposed research methods, initially solved the classical poetry animation automatically generated problems, but also provides a theoretical basis and experimental foundation for subsequent research.

Key words ancient poetry; animation generation; svm classification; co-occurrence analysis

中国古典诗词源远流长, 博大精深, 是中华民族文化艺术宝库中的瑰宝。在信息化高度发展的今天, 如何运用新的技术形式来展现中国古典诗词, 对于传统文化的传承与发展至关重要。古典诗词的动画生成研究, 即将诗词的意境用动画展示出来, 属于计算诗学^[1]的范畴, 有助于使用者更好地理解古典诗词的含义, 感受其中的意境美, 从而激发使用者对古典诗词的热爱。目前, 基于古典诗词的计算机动画生成研究还没有相关的成果, 本研究希望能为后续研究起到抛砖引玉的作用。

迄今为止, 计算诗学和文本到动画的生成(也称为文景转换, text-to-scene)两方面的研究是分离的。就中国古典诗词而言, 自 20 世纪 90 年代后, 北京大学计算语言学研究所、台湾元智大学中国语文学系、重庆大学计算机学院以及厦门大学艺术认知与计算实验室等不少学术机构和学者陆续在诗歌语料库及知识库的建设、韵律分析、风格分析、情感分析和诗歌自动生成等领域开展了汉语计算诗学的广泛研究^[2]。就动画生成而言, Sproat 小组^[3]于 2001 年开发的 WordsEye 系统利用自然语言处理技

术对故事的场景进行分析,用于确定动画角色的特征、方位和动作等,从而指导三维场景的构建。2002年陆汝钊等^[4]提出了故事的全过程计算机辅助动画生成技术。Akerberg^[5]在2003年提出的CarSim系统可用于交通事故场景的重现,系统通过对交通事故场景的文字描述进行自动分析,提取实体、方位、时间等信息构建交通事故的模拟场景。曹存根等^[6]在2008年开发的PNAI系统,能够完成计算机叙事和动画创作。2012年张岳^[7]提出一种面向中文短信的动画情节规划的方法。

本文首次将计算机动画生成用于古典诗词。在计算诗学研究基础上,我们运用自然语言处理技术分析一首诗(词)的具体内容和意境,然后为其量身定制地选取恰当的场景脚本和动画素材,并自动为这些素材进行时间和空间规划,再根据其风格特点挑选合适的文本展示字体和背景音乐,最后运用计算机动画技术(flash动画),自动合成一段声情并茂、视听交融、动静交错的多媒体动画短片。

1 古典诗词动画生成系统架构

本文所构建的古典诗词动画生成系统的整体架

构及主要处理流程如图1所示。其中,句法语义分析模块先对诗词的标题及正文进行分词和词性标注,再利用古汉语辞典等对其进行浅层语义分析;特征提取模块主要针对后续模块的需求抽取诗词中的各种特征信息并发给相应的模块;4个判定模块分别用于确定诗歌的题材、风格和时间,为后续的动画模板、元件、乐曲和字体等的选取以及场景规划提供依据;场景划分模块可以使生成动画的场景更富于变化,有助于更好地展现诗词的意境;元件补充模块利用特征共现等为元件不足的场景自动补充合适的元件;脚本生成模块负责基于Flash Actionscript 3.0的动画脚本编写。

2 基于SVM的古典诗词题材、风格与时间判定

SVM (support vector machine, 支持向量机)^[8]是在统计学习理论基础之上发展起来的一种机器学习算法,较传统算法而言,很好地克服了“维数灾难”和“过拟合”的问题。本文在对古典诗词进行题材、风格和时间分类判别时,主要采用SVM分类器。

在诗词题材判定方面,本文结合相关文献及语

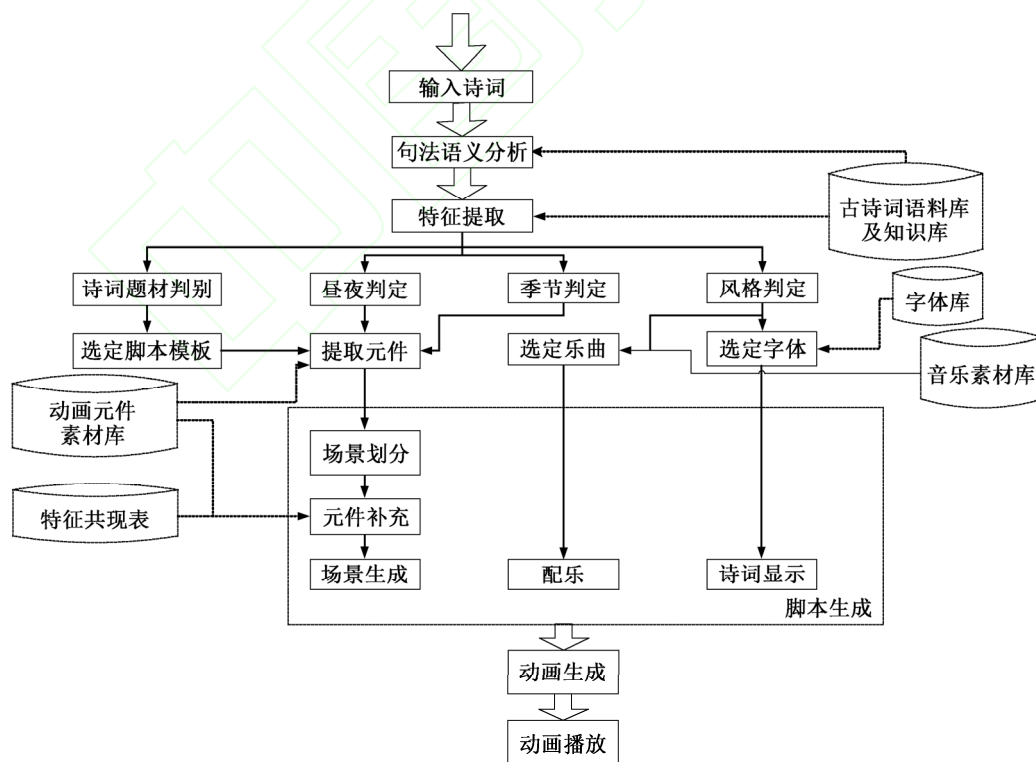


图1 古典诗词动画自动生成系统总体流程

Fig. 1 Flow chart of intelligent multimedia display system for classic Chinese poetry

料库分析将诗词题材分为写景、咏物及抒情 3 类, 并提取标题及正文中的相关句法语义信息作为特征, 利用 SVM 分类器对诗词进行分类, 自动判别诗词题材。我们针对不同诗词题材设计了几种不同的 flash 动画脚本模板, 例如写景诗的模板大致分为 3 层: 最下层为底色层, 用于呈现总体情境背景, 如白天或夜晚、远景等; 中间为背景层, 用于放置较大元件, 即中远距离的景物; 最上面为前景层, 用于放置较小元件, 即画面中的主要焦点景物。

在风格判定方面, 借鉴我们在宋词风格判定方面的已有成果^[9], 本文将其扩展应用于诗词的风格判定, 即针对婉约和豪放两类来选取基于词和语义相结合的特征, 并采用 SVM 分类器进行分类。我们针对不同风格设计了多种风格的配乐及显示字体, 如对于婉约风格的诗词选择轻柔的古琴曲配乐并用较娟秀的字体显示。

为了提高动画元件素材提取的正确性以及生成动画的合理性, 对诗词的昼夜和季节进行判定, 即确定诗词中每一单句的具体时间信息, 并用做后继场景划分和元件补充的重要依据。我们将季节判定与昼夜判定看作是两个文本分类问题, 采用基于 SVM 和改进的 Tri-Training 算法来分别构建相应的分类器。

下面主要以昼夜和季节判定问题为例具体说明我们所提出的方法。

2.1 训练样本获取

本文将诗词中的每一单句作为分类判定的基本单位, 将季节判定分为春、夏、秋、冬和季节特征不明显 5 类, 将昼夜判定分为昼、夜和昼夜特征不明显 3 类。由于目前尚无现成的相关训练语料可用, 我们主要基于《全唐诗》^[10]和《诗学含英》^[11]来构造训练语料库。

本文使用分词工具 segtag^① (专门针对古典诗词训练)对《全唐诗》语料库中的 4 万多首古诗进行自动分词和词性标注, 并采用向量空间模型作为文本表示模型。去除停用词后, 我们只选用最能直接反映时间信息的名词和时间词作为特征, 从而大大降低了特征维度。特征权重选择基于 TF-IDF, 排除一些常用词带来的影响。

由于训练语料数据量巨大, 逐句人工判定标注

费时费力, 我们借助《诗学含英》中提供的 40 个类别及其对应的关键词, 先人工筛选出季节或昼夜特征较明显的关键词标注其对应的类别, 再利用这些带类别标注的关键词来自动标注包含它们的那些单句。例如, 人事类中的“夜宴”、天文类中的“新月”等都归属于昼夜中的“夜”这一类, 故而诸如“夜宴江楼月满身”、“新月上宫城”之类诗句的昼夜类别就自动标注为“夜”。若多个关键词的标注不一致, 则提交人工判别。这样我们就得到一个较可靠的季节和昼夜判定所需的小规模带标注语料。

2.2 SVM 分类器的构建

本文的 SVM 训练采用目前比较流行的 SVM 工具 LIBSVM^②, 基于传统的协同训练^[12]思想, 将 2.1 节中的带标注语料作为初始样本集, 采用改进的 Tri-training 算法^[13], 在大规模的不带标注的《全唐诗》诗句语料库上进行半监督学习, 分别训练得到大规模的昼夜判定和季节判定的熟语料库, 继而在其基础上构建相应的 SVM 分类器。

Tri-training 算法使用三个分类器, 如果其他两个分类器对某个未标记样本的分类结果一致, 那么就将该样本加入第 3 个分类器的已标记样本集中。在实践中, 我们发现仅用两个分类器做判定, 分类器过少, 容易引入噪声, 影响分类器的整体性能, 因此我们对算法进行改进, 将算法中的分类器增加至 4 个。实验表明, 这样既可降低未标记样本分类结果的错误率, 又可较好地保持算法的执行效率。算法流程如下, 其中 x 表示待分类的样本, $K_i(x)$ 表示用第 i 个分类器判定样本 x 的结果。

算法: 改进的 Tri-training (Qua-Training)

输入: 未标记样本集 U , 已标记样本集 L , 某种分类算法 K

输出: 协同学习后得到的带标注样本 new_L , 即将 4 个分类器的训练集合并得到 new_L

迭代: 1) 随机抽取样本 L , 分成 4 份, 得到 4 个训练集 L_1, L_2, L_3, L_4 ; 分别用于同一种学习算法训练出分类器 C_1, C_2, C_3, C_4 ;

2) 对于任意一个由算法 K 训练出的分类器 C_i 维护其对应训练集 L_i , 将满足 $\{x|x \in U, \text{且 } K_i(x) \neq K_m(x) \neq K_n(x), l, m, n \neq i\}$ 的样本加入到 L_i 中, 遍历 U 后, 得到更新后的 L_i' ;

3) 对于每个 C_i , 若 $|L_i'| > |L_i|$, 则对 C_i 利用新的训练集 L_i' 重新训练, 得到新的分类器 C_i' ;

① <http://121.192.180.171/mysoft/segtag.rar>, 2014

② LIBSVM—A Library for Support Vector Machines, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, 2012

4) 对于每个 C_i , 若不满足 $C_i = C_i'$, 转 2) 继续迭代;
终止条件: 4 个分类器都不再更新

3 动画脚本的生成

3.1 场景的自动划分

从输入的诗词中提取到对应的动画元件素材后, 该如何将这些元件组合起来构成动画? 简单堆砌在一个动画场景中或者只是将同一诗句的元件放在同一动画场景中展示, 显然是不行的。合理的场景划分既可以使生成的动画富于场景变化, 又有助于更好地展现诗词的意境。

本文主要依据古典诗词单句的季节和昼夜判定结果以及古典诗词体裁特点, 提出一种基于规则的动画场景自动划分方法。季节和昼夜都可以作为场景划分的依据, 而且根据常识可知, 季节的优先级应当高于昼夜。例如, 同一联中相邻的两个单句都属于“夜”这一类, 但是它们的季节信息不同, 显然应当将它们划分在不同的场景中。至于一首诗词中的每个单句, 哪些该合并在一个场景中, 哪些该划分在不同场景中, 就需要依据诗谱或词谱制定明确的划分规则。

以格律诗为例, 我们首先以联为单位, 对同联中季节或昼夜特征不明显的单句进行归并处理, 具体的方法如下。其中, J1~J5 依次对于应春、夏、秋、冬和季节特征不明显 5 类, Z1~Z3 依次对应于昼、夜和昼夜特征不明显 3 类。

输入: 各个单句的季节判定结果 $\{aj|aj=J1 \text{ or } J2 \text{ or } J3 \text{ or } J4 \text{ or } J5 \text{ 且 } j=1, 2, \dots, n\}$ 和昼夜判定结果 $\{bj|bj=Z1 \text{ or } Z2 \text{ or } Z3 \text{ 且 } j=1, 2, \dots, n\}$

输出: 根据对仗原则处理好的判定结果 $\{aj'|aj'=J1 \text{ or } J2 \text{ or } J3 \text{ or } J4 \text{ or } J5 \text{ 且 } j=1, 2, \dots, n\}$ 和 $\{bj'|bj'=Z1 \text{ or } Z2 \text{ or } Z3 \text{ 且 } j=1, 2, \dots, n\}$

```
for(i=1;i<=n;i++) {
    if (ai=J5)
        if ((i+1)/2==0) ai-1= ai ;      else
ai= ai-1 ;
        if (bi=Z3)
            if ((i+1)/2==0) bi-1= bi ;
    else bi= bi-1 ;
}
```

然后, 根据优先级依次将属于同一类的诗句合并, 具体方法如下。

输入: 处理好的季节判定结果 $\{aj'|aj'=J1 \text{ or } J2 \text{ or } J3 \text{ or } J4 \text{ or } J5 \text{ 且 } j=1, 2, \dots, n\}$ 和昼夜判定结果 $\{bj'|bj'=Z1 \text{ or } Z2 \text{ or } Z3 \text{ 且 } j=1, 2, \dots, n\}$

输出: 基于季节判定的古诗场景划分结果
for(i=1;i<=n;i++)

if(ai'=ai+1'&&bi'=bi+1')
合并第 i 句与第 $i+1$ 句

3.2 基于古典诗词特征的共现分析

本文针对古典诗词的特点, 提取了最能直接反映时间信息的名词和时间词作为特征, 并将诗词中的每一个单句看做特征共现分析时的窗口单元。对于同时出现在一个单句中的特征词语, 认为它们共现了一次。共现频率常用来分析计算特征词语间的关联度^[14]。对特征词语间共现频率采用 Equivalence 指数(简称 E 指数)^[15]来表示, 如式(1)所示, 其中 c_{ij} 表示词 i 和词 j 的共现频率, c_i 和 c_j 分别表示词 i 和 j 单独出现的次数。 E 指数消除了文本总数 N 的限制, 而且分子分母乘以相同的数对结果也没有影响。如果两个词总是同时出现, 那么它们的共现频率 E 值等于 1, 相反, 如果两个词毫无联系, 那么 E 值为 0。

$$E_{ij} = \frac{c_{ij}^2}{c_i \cdot c_j} \quad (1)$$

根据上述原则, 我们对《全唐诗》数据库中的 4 万多首古诗进行了特征词语共现分析, 利用 E 指数计算得到特征词语间的共现频率。在 11 万余对共现特征词语中, 本文只选取其中共现频率在万分之一以上的共现特征词语, 形成特征共现表。

共现分析得到的数据主要用于诸如元件补充等动画生成过程的优化上。在系统完成自动场景划分之后, 一个场景中可能只有一个单句。有些单句可以提取到的动画元件素材很少, 仅基于这些元件素材来生成该场景的动画, 场景可能过于空洞单薄, 无法很好地呈现对应诗句的意境。为此, 我们借助特征共现表, 为元件素材过少的场景提供其他合适的备选元件。例如, 某个场景中只提取到“雪”这一动画元件素材, 借助特征共现表, 可以得到与“雪”共现频度极高的备选元件“梅花”、“枯杉”、“松”等, 从而扩充了该场景中的元件素材。

4 动画素材库的建立

动画素材库是动画自动生成中不可或缺的重要基础。本文的动画素材库分为 3 个部分, 一是用于画面生成的动画元件库, 二是用于动画配乐的音乐素材库, 三是用于呈现诗词文本的字体库。

为了统一生成动画的画风, 本文构建的所有动画元件均来源于网络上收集的水墨风格图片素材,

并逐一经过人工剪切加工后,制成 flash 动画元件。为了便于动画生成过程中古典诗词意境的展示以及动画元素素材的自动提取,对每个动画元件均标注了属性信息,主要分为 3 类。

1) 动画元件的时间信息,即元件可适配的季节与昼夜类别。

2) 动画元件的位置信息,即元件可适配的脚本模板及层次。

3) 动画元件的语义信息,即元件可适配的特征词语,不过由于特征词语集合很大,难以逐一列举,我们利用《同义词词林》^[16]对特征词语进行了同义词类归并。

本文所构建的音乐素材主要来源于网络上收集的古典乐曲音频,并根据乐曲的风格特点区分为豪放和婉约两类,以便自动用于不同风格的古典诗词配乐。本文所构建的汉字字体库中的字体也相应地分为豪放和婉约两类。用户从动画中古典诗词的显示字体可以直观地看出该首诗词的风格。

5 实验结果及分析

5.1 题材、风格与时间判定结果分析

本文中古典诗词题材和风格判定的语料主要来自相关书籍及古诗网站上人类专家分类结果,时间判定的语料即 2.2 节中训练得到的大规模熟语料,其中,不同题材古诗共 1678 首,不同风格古诗共 1305 首,不同季节古诗单句共 34975 句,不同昼夜古诗单句共 22083 句。我们将上述语料分别按 2:1 的比例分为训练集与测试集,分别训练得到相应的 SVM 分类器。对这些 SVM 分类器分别进行封闭测试和开放测试,评价指标选择常用的准确率(Precision)和召回率(Recall),实验结果如表 1 所示。

从表 1 可以看出: 1) 题材分类器对咏物类判定准确率较低,主要原因在于咏物类的诗词中往往包含许多景物描写,而且不少咏物类诗词常采用借物抒情的手法,导致一些咏物类诗词与写景类或抒情类的诗词区分度不大; 2) 风格分类器对于诗词豪放和婉约风格的判定准确率及召回率基本能够满足自动动画生成的需要; 3) 季节分类器除“季节特征不明显”类和“夏”类的个别数据外,在开放测试和封闭测试中的准确率与召回率大都比较理想; 4) 与季节分类器相比,昼夜分类器的测试结果波动较大,但总体来看,测试的平均指标能够稳定在 87%以

表 1 各 SVM 分类器测试结果
Table 1 Testing results of each SVM classifier

分类	封闭测试		开放测试	
	准确率/%	召回率/%	准确率/%	召回率/%
写景	88.49	90.10	80.87	86.26
题材 咏物	78.67	78.32	76.44	76.11
抒情	86.08	84.34	84.07	77.27
风格	豪放	96.58	94.95	94.25
	婉约	92.15	94.63	87.93
季节	春	97.69	93.39	95.92
	夏	97.97	86.38	92.43
	秋	99.52	99.03	99.13
	冬	97.64	95.32	95.95
	季节特征不明显	87.18	97.55	80.87
昼夜	昼	96.69	84.28	98.57
	夜	97.67	95.87	89.08
	昼夜特征不明显	83.96	97.30	76.59

上。上述实验结果表明, SVM 分类器可以较好地满足古典诗词单句的题材、风格与时间判定要求,达到预期的效果。

5.2 场景划分结果分析

为了测试场景自动划分方法的有效性,随机选择《全唐诗》中 500 首古诗进行自动场景划分测试实验。经逐一人工评判,其场景划分的准确率达到 78%。我们从中选取 4 首作为示例,详见表 2,其中 S1~S8 分别表示古诗中的每一个单句。

表 2 中,对第 2 首古诗,系统自动将“春眠不觉晓,处处闻啼鸟。”划分为一个白天场景,而“夜来风雨声,花落知多少?”则划分为一个夜晚场景。对第 4 首古诗,“窗含西岭千秋雪”一句描绘的是远处的高山积雪,系统将其判为冬季,从而实现从近处黄鹂、白鹭等到远处的雪山、江水等的场景切换。

不过,由于场景划分方法过于简单,难免存在一些划分错误。一方面,目前只考虑季节和昼夜的分类,在划分一些抒情诗时容易出现错误;另一方面,由于古诗单句的字数较少,特征提取时也存在错误或遗漏,甚至可能出现提取不到特征词语的情况,这些都影响了分类器判定的准确性,从而导致场景划分错误。

表 2 古诗场景划分测试结果
Table 2 Testing results of scene classification

序号	古诗	SVM 判定结果		场景划分结果
		季节	昼夜	
1	渡远荆门外, 来从楚国游。	J5, J5	Z3, Z3	场景一: S1, S2, S3, S4
	山随平野尽, 江入大荒流。	J5, J5	Z3, Z3	
	月下飞天镜, 云生结海楼。	J5, J5	Z2, Z3	场景二: S5, S6
	仍怜故乡水, 万里送行舟。	J5, J5	Z3, Z3	场景三: S7, S8
2	春眠不觉晓, 处处闻啼鸟。	J1, J1	Z1, Z1	场景一: S1, S2
	夜来风雨声, 花落知多少?	J5, J5	Z2, Z2	场景二: S3, S4
3	好雨知时节, 当春乃发生。	J1, J1	Z1, Z1	场景一: S1, S2
	随风潜入夜, 润物细无声。	J5, J5	Z2, Z2	场景二: S3, S4, S5, S6
	野径云俱黑, 江船火独明。	J5, J5	Z3, Z2	
	晓看红湿处, 花重锦官城。	J1, J1	Z1, Z1	场景三: S7, S8
4	两个黄鹂鸣翠柳, 一行白鹭上青天。	J1, J5	Z1, Z1	场景一: S1, S2
	窗含西岭千秋雪, 门泊东吴万里船。	J4, J5	Z1, Z3	场景二: S3, S4

5.3 系统性能测试

为了测试上述算法思想的可行性和有效性, 我们搭建了一个古诗动画自动生成原型系统。图 2 给出了该系统生成的两个动画实例。需要说明的是, 对于同一首古诗, 该系统每次生成的动画可能有所不同。

按写景、抒情、咏物 3 类, 我们从《全唐诗》中各选取 20 首广为流传的五言或七言的绝句或律诗作为候选测试数据集, 随机抽取信息学院的 27 名本科生作为被试。每位被试按类别从候选测试数据集中各随机选取 3 首, 观看由该系统自动生成的动画。以被试对动画的满意度作为衡量系统性能的

一个重要指标。为了尽量减少满意度评价结果的主观随意性, 我们将其分解为场景划分准确性、风格情感一致性和总体质量 3 个评价指标, 采用 5 分制打分。表 3 给出 27 位被试的评测结果。

上述测试结果表明: 该系统的古诗动画自动生成质量还是比较令人满意的, 基本能够表现出古诗

表 3 系统性能评测结果
Table 3 Systemic testing results

平均生成时间/s	被试满意度(0~5 分)		
	场景划分准确性	风格情感一致性	总体质量
1.691	3.26	3.18	3.44



图 2 古诗动画生成系统示例
Fig. 2 Examples of intelligent multimedia display system for classic Chinese poetry

内容和意境, 达到了预期的效果; 并且该系统的动画生成效率较高, 可以较好地满足用户批量、高效地生成古诗动画的需求。

6 总结与展望

对古典诗词开展数字化研究具有重要而深远的意义。以计算机动画等形式为载体的多媒体展示, 极大地丰富了古典诗词的流传方式, 并能帮助读者更好地感受诗词的意境、领悟其中的精髓, 从而激发读者对古典诗词的热爱。

本文首次尝试将计算诗学、计算机辅助动画自动生成、自然语言处理技术、机器学习等相关学科领域的现有研究成果和技术综合应用于古典诗词的动画自动生成, 并初步构建起一个古诗动画自动生成原型系统。实际测试结果表明, 该系统生成的动画能够较合理地展现古诗的内容及意境, 用户接受率达到 70% 以上, 且动画生成速度快, 基本达到预期的效果。今后将在以下方面进一步开展研究工作。

1) 全唐诗熟语料库的加工和相关知识库的建立有待于进一步完善, 尤其亟需编撰一部古典诗词机读语义词典。

2) 提供人机交互式的动画编辑接口, 用户可对自动生成的动画进行修改和完善, 直至制作出满意的作品。

3) 改进并细化与古诗意境紧密相关的关键特征的自动提取及辨识, 如增加人物、场所、方位等特征, 提升动画生成质量。

4) 现阶段主要侧重理论方法上的探讨, 若要投入实际应用, 系统在动画、音乐、字体等素材的质量和数量上仍有待提高, 在意境展示方式上还可扩充, 如增加动画对象的声音素材库、添加古诗的自动朗诵等等。

参考文献

[1] 周昌乐. 心脑计算. 北京: 清华大学出版社. 2003: 195 - 203
 [2] 周昌乐, 郑旭玲, 何中市, 等. 中国古典诗词楹联的计算机化研究. 心智与计算, 2012(2): 75 - 82

[3] Coyne B, Sproat R. WordsEye: an automatic text-to-scene conversion system // Proceedings of the SIGGRAPH 2001 Annual Conference on Computer graphics. Los Angeles, 2001
 [4] 陆汝矜, 张松懋. 从故事到动画片: 全过程计算机辅助动画自动生成. 自动化学报, 2002, 28(3): 321 - 348
 [5] Akerberg O, Svensson H, Schulz B, et al. CarSim: an automatic 3D text-to-scene conversion system applied to road accident reports // Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics. Budapest, 2003
 [6] 曹存根, 李良君, 李志豪, 等. 智能动画创作系统 PNAI 的研究进展. 系统科学与数学, 2008(11): 1407 - 1431
 [7] 张岳. 面向中文短信的三维动画情节规划. 计算机工程及应用, 2012(10): 187 - 193
 [8] Vapnik V N. Statistical learning theory. New York: Wiley, 1998
 [9] Zheng Xuling, Zhou Changle, Zeng Hualin. Song Ci style automatic identification. Journal of Donghua University: English Edition, 2010, 27(2): 181-184
 [10] 曹寅, 彭定求, 等. 御定全唐诗. 扬州诗局刻本, 1706
 [11] 刘文蔚. 诗学含英. 香港: 香港银河出版社, 2001
 [12] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training // COLT' 98 Proceedings of the eleventh annual conference on Computational learning theory. 1998: 92-100
 [13] Zhou Z H, Li M. Tri-training: exploiting unlabeled data using three classifiers. IEEE Transactions on Knowledge and Data Engineering, 2005(17): 1529-1541
 [14] Ding Y, Chowdhury G G, Foo S. Bibliometric cartography of information retrieval research by using co-word analysis. Information Processing and Management, 2001(37): 817-842
 [15] 王曰芬, 宋爽, 熊铭辉. 基于共现分析的文本知识挖掘方法研究. 图书情报工作, 2007, 51(4): 66 - 69
 [16] 梅家驹, 竺一鸣, 高蕴琦, 等. 同义词词林. 上海: 上海辞书出版社, 1993: 106 - 1081