

基于程序设计主观题智能阅卷算法的应用研究

丁卫平¹, 管致锦^{1,2}, 陈建平¹

(1. 南通大学 计算机科学与技术学院, 江苏 南通 226019;

2. 南京航空航天大学 信息科学与计算机学院, 江苏 南京 210003)

摘要:在程序设计语言考试中, 程序设计主观题智能评阅是一项具有实用价值的应用。文中在分析主观题阅卷算法的基础上, 针对程序设计主观题的特征, 提出了基于阅卷参考点和程序理解相结合的智能阅卷算法, 并进行了相关实例分析。设计了 VC++ 程序设计题智能阅卷系统, 通过实际教学使用表明该智能阅卷算法具有较好的正确性和客观性。

关键词:程序设计题; 自动评阅; 可信度; 网络考试

中图分类号: TP301; G434

文献标识码: A

文章编号: 1673-629X(2007)11-0205-04

Research and Application of Intelligent Assessment Algorithm Based on Programming Subjective Questions

DING Wei-ping¹, GUAN Zhi-jin^{1,2}, CHEN Jian-ping¹

(1. School of Computer Science and Technology, Nantong University, Nantong 226019, China;

2. College of Information Science and Technology, Nanhang University, Nanjing 210003, China)

Abstract: The intelligent assessment of the programming subjective questions is practical application in the programming language examination. On the base of analyzed actuality of the assessment algorithm of the subjective questions, the intelligent assessment algorithm combining the program understood and reference point established together is put forward, and the corresponding example is analyzed, aiming at the character of the programming subjective questions. Then the intelligent assessment system of the VC++ programming language is designed. After the system has been applied, the actual result shows that it can achieve the better correctness and objectivity.

Key words: program designing questions; intelligent assessment; reliability; network examination

0 引言

随着计算机辅助处理、多媒体技术以及计算机网络技术的飞速发展和推广应用, 在计算机基础教育中采用计算机实现考试无纸化已成为目前发展的必然趋势^[1], 其涉及到多个重要组成部分, 如题库的建立与维护^[2]、试卷的自动生成^[3]、试卷的自动批阅等。其中试卷的自动批阅对于大规模的考试(如计算机等级考试)来说意义非常重大, 涉及到计算机领域的人工智能、模式识别以及自然语言理解等方面的理论和知识。

目前一些国际性的标准化考试已实现了计算机的

自动批阅, 江苏省作为全国教育大省, 计算机等级考试处于领先地位, 考试的方法和手段也应该与发展要求相适应。江苏省计算机等级考试每年举行两次, 每次都有几十万的学生参加不同语种的考试, 试卷采用人工阅卷的方法, 需要付出极大的人力、物力, 而且批阅的周期长, 容易受主观判断影响, 误差较大。因此, 对计算机等级考试中主观题实现自动阅卷是目前迫切的需要。

1 主观题阅卷算法现状

自从 20 世纪 60 年代以来, 国外许多专家和学者就致力于基于任意文本答案的主观题计算机自动批改技术的研究, 同时也相应出现了各种不同的自动阅卷系统: 如 Ellis Page 等人^[4]开发的 Project Easay Grade (PEG) 系统, 它依靠统计方法来评定文章的质量和写作风格; Callear D 等人^[5]开发的 Automated Text Marker (ATM) 系统是一种针对非多项选择、有明确答案且答案简短类试题的计算机辅助评分方法, 它可以对用

收稿日期: 2007-01-30

基金资助: 江苏省现代教育技术研究课题(2005-R-203, 2007-R-4618); 南通大学教学研究重点资助课题(A0615); 南通大学高等教育研究课题(07GJ002)

作者简介: 丁卫平(1979-), 男, 江苏常州人, 讲师, 硕士, 研究方向为智能辅助教学系统、数据挖掘等; 管致锦, 博士, 教授, 硕导, 研究方向为量子可逆计算; 陈建平, 教授, 硕导, 研究方向为数字信号处理等。

自然语言书写的答案内容进行评判,并且能够应用在各种具体的学科上。

国内针对各类非标准化试题自动批改技术的研究虽然刚刚起步,但有些学者专家也提出了一些较好的观点:如张量等人^[6]针对计算机基础上机考试中中文录入题,根据学生所录入字数的多少以及错录、漏录等提出了字符串匹配的方法;李辉阳等人^[7]提出了基于关系的带权匹配技术来实现计算机辅助教学中简单论述题的正误判定等;王邯等人^[8]针对当前计算机水平考试和等级考试中普遍采用的程序填空类试题,提出了将学生答案与标准答案在语义上进行匹配来实现计算机自动批改填空题的方法等等。

目前由于基于任意自然语言的自动批改系统实现具有相当的难度,真正通用的系统尚未出现,但可以开发出针对特定领域、特定需求的主观题自动批阅系统。所以笔者针对计算机等级考试这一特定领域中程序设计主观题的特征,进行了批阅算法的设计和实现,通过实际系统使用结果表明利用该批阅算法所设计的阅卷系统是有效的和可行的。

2 程序设计主观题智能阅卷算法

2.1 程序设计主观题特征

文中所研究的程序设计题是计算机考试中主观题的一种,由于程序设计语言不同于一般的自然语言,其具有如下一些特征:

(1)计算机程序设计语言是一些比较成熟的语言,语句比较精练,语法、词法约束和限制非常严格,且程序的基本结构都比较固定。

(2)程序设计题基本上指出(或者可以分析推断出)考查的是某些固定知识点的理解程度与应用能力,部分题目还明确指出了用什么样的方法来实现这个程序,甚至编程中必须用哪些语句都规定好了,因此目标程序的描述比较明确,功能叙述也较完备。

(3)考试中中学生编写程序设计题答案的代码量相对比较小,一般在 30 行左右,这样给实现机器智能阅卷提供了较好的便利条件。

程序设计题的批阅所要解决的问题是如何使机器能模仿人工分析、推理,自动理解学生程序,使用某种策略进行公正、合理的评分。但是学生程序题的答案也是多样性的,必须要从学生提交的程序答案中抽取、过滤掉一些对评分无用的信息,通过一些分析手段以及推理方法寻找学生程序和题目需要完成目标程序之间的差异,这种差异就可以作为相关的评分标准。因此对程序设计主观题进行智能阅卷是目前研究的热门且具有挑战性的课题。

2.2 智能阅卷算法思想

目前所见到的程序设计题大多数自动阅卷方法是通过运行学生所编写的程序生成相应的输出文件,然后和标准文件进行比较,根据比较结果给出一个相应的分数,也就是说,衡量的依据只是一个输出文件而已,这显然也是不正确的。笔者在大量研究和实验的基础上^[6-10],提出了基于阅卷参考点和程序理解相结合的方法,来进行程序设计题阅卷分数的确定。该方法在实际阅卷过程中能给出较公正、合理分数,并且能较好地考查学生的实际编程能力。

1)相关概念:

(1)执行可信度。

在程序的执行过程中,一个扩展名为 exe 的文件不一定就是真正的可执行程序,因为它可以通过人为的方式改变一个文件的扩展名而实现,所以阅卷时不但要检查可执行程序生成与否,还要检查其合法性。检查的结果可以分为三种情况:可执行程序不存在时为 0;存在但不合法时为 0.1;存在并且合法时为 1。这个数值称之为执行可信度。

(2)程序可信度。

根据题目的要求事先写一个参考程序,该程序尽可能具有较大的代表性,一般是正常思维情况下的实现方法,另外在该程序中要以注释的方式添加其他实现方法的程序代码,以扩大答案的覆盖面。在考查程序的可信度时通过程序理解方式进行,即设立 4 种匹配模式:规模匹配(size)、结构匹配(struct)、深度匹配(profound)和语句匹配(node)。对学生提交的程序中源程序所包含的知识信息提取出来,再按照策略将这些信息组织成以层次形式表示的信息空间,通过逐层抽象得到程序的语义理解功能。相关匹配点的计算方法参见参考文献[10],在文中的举例部分也给出了其计算方法,通过程序理解得出该系数是衡量程序最大相似性的一个重要指标,称之为程序可信度。

(3)输出可信度。

考生编写的程序运行时,根据预先提供的测试数据,自动生成相应的输出文件。测试数据一般有多组,每组对应一组输出结果,比较不等时按相应权重进行给分。

3 算法设计

智能阅卷算法如图 1 所示。

其具体实现思想如下:

(1)确定评阅试题有以下考查点:可执行文件、编码、输出、其他因素等,它们权值分别为 exep, codep, outp 和 otherp,初始化该题总分 $ts =$ 该题分数,实际得

分 $score = 0$ 。

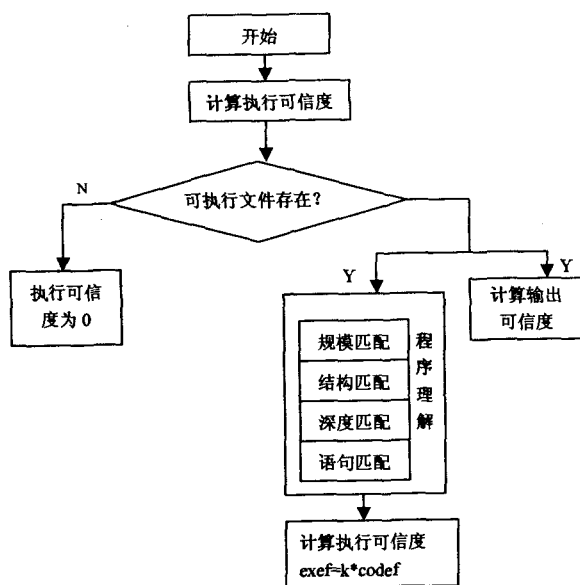


图 1 智能阅卷算法

(2) 计算程序 4 种匹配度, 从而得出程序的执行可信度 $exef$ 、程序可信度 $codef$ 。

计算执行文件得分: $score += ts * exef * exep$

(3) 进行如下判断:

$codef \geq 0.6 \&\& codef < 1.0$? 若是则计入输出得分, $score += ts * outf * outp$

$score < ts$? 若是则取代码分配分数和剩余分数中较小者记为 $bets$, 然后计入代码得分, $score += bets * codef$

$codef \geq 0.98 \&\& outf < 0.6 \parallel outf = 1 \&\& codef < 0.6$?

若是则说明输出结果不可信, 要扣除分数, $score = ts * notp$, 最后得到该题的总分。

以下面一道简单的程序设计题为例进行上述智能阅卷算法的说明和分析。标准代码与学生代码如图 2 所示。

$$\begin{cases} y = -x + 2 & x \geq 0 \\ y = 2 & x < 0 \end{cases}$$

标准参考代码	学生一代码	学生二代码
<pre>#include<iostream.h> void main({float x,y; cin>>x; if(x>=0) y=-x+2; else y=2; cout<<y<<endl; }</pre>	<pre>void main({float x; y; cin>>x; if(x>0) y=x+2; else y=2; cout<<y<<endl; }</pre>	<pre>void main({float x; float y; cin>>x; if(x<0) y=x+2; else y=2; cout<<y<<endl; }</pre>

图 2 标准代码与学生代码

学生一程序阅卷得分情况分析:

第一步: 计算执行可信度:

设立各个匹配在评分中的权值, 分别为: $\lambda_{size} = 0.1, \lambda_{struct} = 0.3, \lambda_{profound} = 0.5, \lambda_{nodes} = 0.1$ 。由学生程序、模板程序、学生程序标准化后、模板程序标准化后系统依赖图分析可得各个匹配度值为: $D_{size} = 1, D_{struct} = 1, D_{profound} = 0.840, D_{node} = 0$

所以本次匹配的综合匹配度:

$$D = \lambda_{size} * D_{size} + \lambda_{struct} * D_{struct} + \lambda_{profound} * D_{profound} + \lambda_{nodes} * D_{node} = 0.82$$

可以得出执行可信度、代码可信度: $codef = D_{node} = 0.82$; $exef = k * codef = 0.82$

第二步: 由上分析得出输出可信度 $outf = 1$

第三步: 设立权值 $exep = 0.1, codep = 0.6, outp = 0.3, otherp = 0.1$

第四步: 计算得分(初始化总分 ts = 该题的总分, $score = 0$)

(1) 计算可执行文件得分: $Score += Ts * Exef * Exep = 0.82$

(2) 由于 $codef \geq 0.6 \&\& codef < 1$, 则输出得分为 3.82

(3) 由于 $Score < ts$, 则代码得分为 7.10

所以该题的得分为: 7.10。

同样可以计算出学生二程序阅卷得分:

$D_{size} = 1, D_{struct} = 1, D_{profound} = 0.926, D_{node} = 0$
程序 2 得分为: 6.317。

4 程序设计主观题智能阅卷系统设计

4.1 阅卷系统基本结构和主要功能

基于上述提出的智能阅卷算法思想设计了以 VC++ 为例的程序设计语言等级考试阅卷系统, 该系统使用 VB 和 MFC 混合编程实现, 较好地解决了 Window 底层处理和 ActiveX 控件设计, 系统的框架结构如图 3 所示。

在实现自动阅卷系统的过程中, 将系统根据其功能分为三个模块:

1) 数据库模块: 该模块包括了标准答案、考生答案和分数库, 是自动阅卷系统记录数据的主要部分。

2) 评阅分析模块: 该模块主要完成对用户程序的分析功能, 并按上述设计的主观题阅卷策略驱动用户程序, 得到运行结果。

3) 评分模块: 该模块主要将考生结果与标准结果进行比较, 得出阅卷结果, 并保存至数据库中。

系统管理员或者教师通过自己的用户名和密码进入系统阅卷得分后台, 阅卷系统首先计算规模匹配、结构匹配、深度匹配、语句匹配等各个匹配度得到程序可

信度,然后可以得出阅卷各参考点的得分,即执行可信度得分、输出可信度得分、程序可信度得分。然后计算出每位学生相关程序设计题的总分情况,即总得分=执行得分+输出得分+程序得分+其他得分,从而给出对批阅答案一个全面的评价和较合理的分数。

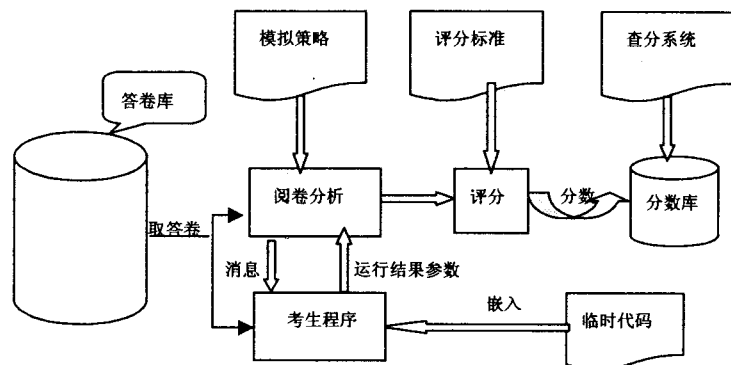


图3 程序设计主观题智能阅卷系统框架结构图

4.2 阅卷系统误差分析

在实际使用过程中,进行计算机自动阅卷与人工阅卷方法的比较,计算其实际误差率。定义计算机自动阅卷方法总分表示为 $A(\text{SUM})$,人工阅卷方法总分表示为 $P(\text{SUM})$,试卷中该题的实际总分表示为 $T(\text{SUM})$,误差率 η 的计算公式如下:

$$\eta = \frac{|P(\text{SUM}) - A(\text{SUM})|}{|T(\text{SUM})|} \times 100\%$$

随机抽取了两套试题,应用本系统的自动评阅方法以及人工阅卷的方法进行误差率 η 计算,实验结果如表1所示。

表1 自动阅卷与人工阅卷方法误差率 η 比较

答卷题号	答卷一	答卷二
	η	η
1	2.13%	2.35%
2	1.34%	3.98%
3	2.00%	2.39%
4	1.98%	1.60%
5	2.16%	2.34%
6	0.96%	1.02%
7	1.10%	1.62%

由上表可以看出该智能阅卷系统的自动评阅方法与人工阅卷方法相比,实际的误差率相对较小,在一定的允许误差范围之内,表明该阅卷系统具有较好的正确性和客观性。

目前已成功地将该阅卷系统应用了三届计算机专业和非计算机专业的计算机等级考试不同语种的期末考试和等级考试考前模拟练习。从学生参加江苏省高

校计算机二级考试的情况来看,2005年秋季江苏省二级考试 VC++ 语种学生们取得了较好的成绩,平均通过率为 68.78%,超出省二级平均通过率(42%)。平均优秀率为 28%,其中优秀人数 23 人,占通过人数的 28.8%,占有所有等级考试不同语种优秀人数的 51.

68%(以上数据源于学校教务处)。可以看出采用该阅卷系统后,计算机等级考试程序设计语言课程的教学质量得到了较大幅度的提高。

5 结束语

主观题的智能阅卷是计算机等级考试系统中基本功能与必然选择。针对程序设计主观题特征,设计了算法和系统,在实际教学应用中取得了较好的教学效果。这对江苏省程序设计语言考试全面实现自动评阅具有重要的应用和研究价值,同时对其他学科考试的自动评阅有着重要借鉴和推广意义。

参考文献:

- [1] 万玛宁,关永.人工智能技术在计算机辅助教学中的应用研究[J].微计算机信息,2006,26(2):257-260.
- [2] 丁卫平,管致锦,顾翔.智能计算机辅助教学在重点课程建设中的研究[J].电子电气教学学报,2006,28(5):96-100.
- [3] 武卫平,管致锦,沈学华.基于遗传算法的智能组卷应用研究[J].电子电气教学学报,2005,27(2):93-95.
- [4] Rudner L. An Overview of Three Approaches to Scoring Written Essays by Computer[EB/OL]. 2002-11-24. <http://ericae.net/pare/getvtn.asp>.
- [5] Callear D, Jerrams-Smith J, Soh V. CAA of Short Non-MCQ Answers[C]//Fifth International Computer Assisted Assessment Conference. [s.l.]:[s.n.], 2001.
- [6] 张量,詹国华.开放式智能化计算机考核阅卷系统的设计与实现[J].计算机工程与应用,2001,37(10):108-110.
- [7] 李辉阳,韩忠愿.有限领域简述文字的自动判读及其在CAI中的应用[J].计算机工程与应用,2002,38(8):76-78.
- [8] 王邯,冯刚,肖俊.网络教育中C程序设计自适应考试系统的设计和实现[EB/OL]. 2002-11-20. <http://www.ccc.edu.cn/elearning/2001nh/3-6.doc>.
- [9] 乔善平.基于智能移动Agent的C语言考试系统[J].计算机工程与科学,2004,26(4):101-105.
- [10] 李永浩.基于程序理解的编程题自动评分系统研究与应用[D].哈尔滨:哈尔滨工业大学,2004.