

填空题自动评分系统的研究与实现

Design and Implementation of Auto - grade system for Blank testing

梁振球 (广东技术师范学院 计算机与网络中心 广州 510665)

摘要:本文对当前大部分在线考试系统中的对填空题进行自动评分时的局限性进行了分析,并给出了改进的填空题自动评分的思路及具体的实现方法。最后对填空题自动评分系统的应用效果进行分析和展望。

关键词:在线考试系统 填空题 自动评分

1 问题的提出

网络远程教育是基于网络的远程学习,是随着现代信息技术的发展而产生的一种新型教育方式。其中计算机辅助测试(CAT)与 Internet 技术的有机结合,使基于网络的现代远程考试成为一种新的趋势。由于借助了计算机和 Internet,使得远程考试在许多方面优越于传统考试,随着技术的不断发展与完善,远程考试系统正在逐渐取代传统考试模式成为考试的主流。

目前的考试题型分两大类,一类是标准化题型,如选择题、判断题等,另一类是主观题,如简答、论述题等。标准化题型的评分利用计算机来实现非常容易,而对主观题的评分要用计算机来实现则非常困难,因需要综合人工智能、语言学等多项科学,目前虽然对这方面的一些研究,但不仅实现的过程复杂,而且评分的准确性也不容乐观。

目前大多数在线考试系统的题型仅有单选、多选、判断等客观题,这是由于这些题型的答案是唯一的,使用计算机能快速准确地实现自动评分。而仅有少数在线考试系统支持填空题,但也只能对答案是唯一的试题进行评分,例如“ $1+1=?$ ”,但在实际应用中,大部分的填空题会有多个答案,而且同一个答案不同的人会有不同的表达方式,仅仅以标准化的方式进行简单的比较,显然是不准确的。极少数题库系统尝试把填空题按主观题的方式来评分,但又存在耗时、复杂、准确性不高的情况。

由于填空题是很常见的一种题型,因此,如不能实现对答案复杂多样的填空题进行高准确度的自动评分,在线考试系统的功能将大打折扣。

2 填空题自动评分算法的思路

填空题应是一种特殊的标准化试题,其特殊性体现在答案是复杂多变的。因此我们应该从利用人的灵活性和计算机的高速处理能力的角度出发,在把一到填空题添加到题库系统中之前,就同时把题目的各种可能的答案由教师列出并保存起来,评分时由计算机按照设定的步骤,根据保存的答案及相关的参数进行评分。评分准确度与教师给出的答案的详细程度直接相关,就像一些题目通常是一题多解一样,给出的解题方法越多越详细,评分的准确度就越高。

首先填空题答案的多样性一方面来自答案出现的位置的多样性,分为固定位置和不固定位置两种情况;另一方面来自学生答案与参考答案的比较方式,可分为完全匹配与模糊匹配两种情况。此外,答案不同的表述方式也是重要的原因之一。下面以两个具有普遍性的填空题为例进行分析:

例 1:表中的行可称为_____,列可称为_____。

答案是:记录(或元组);字段(或属性),这是答案的位置固定但可能的答案不唯一的情况。

例 2:表间的关系有_____,_____,_____。

答案是“1对1;1对多;多对多”,但写成“一对一,一对多,多对多”也是对的。同样写成“1:1, 1:N, M:N”这也是对的,这是答案的位置不固定,且每个答案也不唯一的情况,再加上文字的不同(“1”和“一”及“多”和“N”),这样学生的答案就有很多种可能了,如下面都是正确的答案序列:

1 对 1; 1 对多; 多对多
 1 对 1; 多对多; 1 对多
 1 对多; 1 对 1; 多对多
 1 对多; 多对多; 1 对多

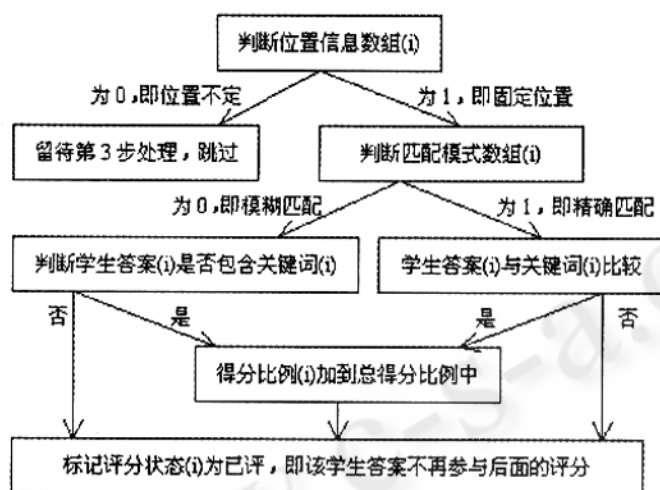


图 1 对位置要求固定的答案进行处理

因此,在进行填空题的自动评分前,教师不仅要指定每道题的各个答案的各种关键字及分值,更要进一步指定每个答案出现的次序究竟是固定的或是任意的,同时还要指定每个关键字是匹配时是要求与考生的答案一字不差的精确匹配或是只需考生的答案中包含标准答案即可的模糊匹配模式。对一些答案有多种表述方式的题目,教师应尽可能地多设定参考答案,以进一步提高评分的准确性,全部设定后把这些信息存入数据库。

在进行评分时,自动评分模块先从数据库中读取这些信息,由于标准答案信息要与每个学生的答案进行比较,即要循环重复使用,为提高速度先以数组的形式装入内存,再与每个考生填写的实际答案按照相关参数自动进行匹配及计算分数,即可实现准确评分。教师对答案的设定越细致,则自动评分的效果就越好。下面给出该算法的详细设计与实现方法。

3 算法设计

第一步:装入各信息数组

由于参考答案要反复与各考生的答案进行比较,故应预先建立相关的信息数组,包括答案关键词数组、

关键词应出现的位置信息数组、关键词对应的得分比例数组、关键词对应的评分状态数组、关键词对应的匹配模式数组、最后在评分时再逐个建立学生答案数组。这样做的好处是把评分过程变为简单的顺序比较,无需进行复杂的遍历。

第二步:对标准答案出现位置要求固定的学生答案进行处理。

如题目要求考生答案出现的次序不能颠倒,只要考生的答案不在正确的位置上,即认为是错误的。同时在比较答案后把评分状态标记为已评,则该位置上的学生答案不可再参与下面的位置不定的答案匹配,以防有的学生不懂而全部填上一个答案来碰运气。如前述的例 1,若考生的答案为“字段(或属性)、记录(或元组)”,由于次序颠倒,得 0 分。

第三步:对标准答案出现位置不要求固定的学生答案进行处理。

在这里对未被标记为“已评”的学生答案逐个进行判断,只要这个答案与未被标记为“已评”的任一标准答案相吻合,即可认为这个学生答案可得分,同时对该位置上的评分状态标记为“已评”,不再参与后面的评分。

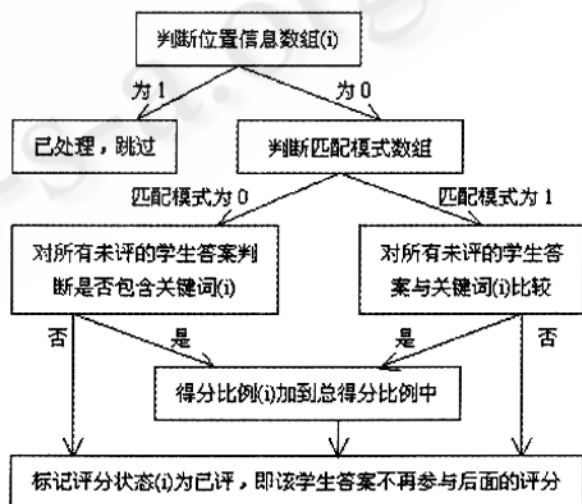


图 2 对位置不要求固定的答案进行处理

第四步:计算该题成绩,清理评分标记,准备对下一考生进行评分。

循环结束后,即完成了这题的评分过程,应立即用现在该题的得分比例乘上该题的分值,即可得到考生

在这一道填空题的实际得分并保存到数据库中。接着仅需对评分状态标记数组初始化为未评、得分比例还原为 0,其余的答案关键词数组、位置信息数组、得分比例数组、匹配模式数组等均保留,然后读取下一考生的答案并跳转到第二步。

4 应用效果分析

在填空题自动评分模块调试通过后,为检查实际的评分效果及运行效率,利用一次学生在线作业提交的数据来进行分析测试。这次作业中填空题部分共 20 个空,其中固定位置的答案有 12 个空、位置不固定的有 8 个空,共有 83 人提交了答案。在配置为 CPU P4 3.0、512M 内存、60G 的 7200 转速硬盘的计算机上运行了填空题自动评分模块,共耗时 3 秒,在运行效率上较为理想,而且用自动评分的方法的得出的成绩与教师人工评分得出的成绩误差率仅为 6%,可见该自动模块较好地实现了对答案复杂多样的填空题进行高准确度的自动评分。

5 结束语

通过给出各种不同的评分要求,不同的学生答案进行验证,本填空题自动评分系统均可进行高准确度的自动评分,证明该自动评分是可靠的,可以适合各种复杂的多答案多可能性的填空题自动评分的需要,比现行大多数系统只能对单一的答案进行简单的比较评分是一大进步,把该评分模块应用于在线测试系统中,将使在线测试系统可以支持更为复杂的填空题,使其的功能更为完善,更加符合考试的实际需求。

参考文献

- 1 江效尧、蒋东升,基于 Web 的通用题库系统设计与实现[J],计算机工程,2002 年 6 月第 28 卷第 6 期.
- 2 陈锦辉、王景皓编著,XML 与 JAVA 程序设计大全[M],北京:中国铁道出版社,2002.
- 3 闪四清,Microsoft SQL Server 2000 实用教程[M],北京:人民邮电出版社,2000.