

Г. О. Евструпов,

Национальный исследовательский университет «Высшая школа экономики», г. Москва

СИСТЕМЫ ОЦЕНИВАНИЯ В ЗАДАЧАХ С АВТОМАТИЧЕСКОЙ ПРОВЕРКОЙ НА ОЛИМПИАДАХ ПО ПРОГРАММИРОВАНИЮ

Аннотация

В статье рассказывается об основных системах оценивания, используемых в задачах с автоматической проверкой решений на крупных личных олимпиадах школьников по информатике и программированию, таких как всероссийская олимпиада школьников и международная олимпиада школьников. Отмечаются преимущества и недостатки различных систем оценивания, даются рекомендации по применению той или иной системы к различным задачам, составленные на основе опыта автора в подготовке и проведении школьных олимпиад по информатике.

Ключевые слова: олимпиады по информатике, программирование, автоматическая проверка, система оценки.

В отличие от большинства других предметных олимпиад, на основных крупных олимпиадах по информатике и программированию сложилась традиция оценивать не непосредственную работу участника, т. е. текст написанной им программы, а эффективность и правильность работы этой программы [5]. С этой целью жюри олимпиады заранее готовит некоторый набор тестовых данных, на котором решение участника запускается с использованием так называемой *тестирующей системы*, после чего оно оценивается исходя из результатов работы на тестах. Традиционно оцениваются такие параметры, как корректность полученного ответа, время работы программы и объем использованной оперативной памяти.

Система оценивания в каждой задаче должна быть известна участнику, для этого она, как правило, описывается в условии задачи.

Подготовленный жюри набор тестов и система оценивания напрямую влияют на результаты олимпиады, поэтому **при выборе системы оценивания следует придерживаться следующих базовых положений:**

- чем ближе решение участника к оптимальному решению задачи, тем больше баллов оно должно набирать;
- требуется заранее продумать все возможные варианты решения задачи (в том числе самые

неоптимальные) и постараться добиться того, чтобы каждый из них набирал хоть какое-то количество баллов;

- следует избегать несправедливых ситуаций, когда участник, допустивший незначительную ошибку в правильном решении (например, неправильно форматирующий ответ), не имеет возможности узнать об этом и набирает низкие баллы или не набирает их вовсе;
- недопустимо, когда решения, идейно одинаковые, но записанные на разных языках программирования, получают разные баллы;
- хорошо, если участники набирают по задаче много разных вариантов баллов, что позволяет жюри избежать ситуации, когда много участников оказываются с одним и тем же суммарным баллом, — подобная ситуация может осложнить определение призеров и победителей олимпиад, а также вручение призов.

Несмотря на то что эти положения кажутся простыми и логичными, соблюсти их в случае использования автоматической проверки решений весьма не просто. Ниже для различных систем оценивания будет рассмотрено, как они помогают выполнить то или иное положение. Еще одним важным аспектом, о котором пойдет речь ниже, является доступность результатов проверки участнику во время соревнования.

Контактная информация

Евструпов Глеб Олегович, ст. преподаватель факультета компьютерных наук Национального исследовательского университета «Высшая школа экономики», г. Москва; адрес: 125319, г. Москва, Косыгинский пр., д. 3; телефон: (495) 772-95-90, доб. 22913; e-mail: gevstropov@hse.ru

G. O. Evstropov,
National Research University High School of Economics

SCORING SYSTEMS USED FOR AUTOMATIC EVALUATION IN PROGRAMMING COMPETITIONS IN INFORMATICS

Abstract

The article describes different scoring systems used in largest school programming competitions such as Russian Olympiad in Informatics and International Olympiad in Informatics. For every scoring systems it's pros and cons are given, complemented by author's recommendations regarding their suitability in some particular cases basing on his experience in preparing of programming contests.

Keywords: olympiads in informatics, programming, automated check, scoring.

Классическая потестовая оценка «втемную»

При использовании данного варианта оценивания участнику до конца соревнования недоступны никакие результаты проверки его решения. Каждый подготовленный жюри тест оценивается независимо от остальных и стоит некоторое заранее фиксированное количество баллов. Описание системы оценивания в условии задачи, как правило, состоит из фраз вида: «Решения, верно работающие при $n \leq 100$, набирают не менее 40 баллов», что позволяет участнику сделать выбор, на какую задачу потратить время и каких результатов ожидать от тех или иных решений.

Данная система оценивания еще десять—пятнадцать лет назад являлась самой распространенной в мире и использовалась практически на всех школьных олимпиадах, несмотря на критику и следующие ее *недостатки*:

- Очень высокий штраф за небольшие ошибки, вызванные невнимательностью, непониманием каких-то несущественных деталей условия или формата ввода-вывода. Зачастую правильные решения с подобными незначительными ошибками набирали близкое к нулю количество баллов. За примерами можно обратиться к результатам всероссийских олимпиад до 2010 года. Так, на заключительном этапе всероссийской олимпиады 2009 года участник из Москвы Дмитрий Аникушин набрал только 12 баллов вместо полных 100 баллов по задаче именно из-за ошибки в формате ввода-вывода, которая не выявилась на тестах из условия, и не стал из-за этого абсолютным победителем олимпиады (<http://neerc.ifmo.ru/school/archive/2008-2009/ru-olymp-roi-2009-standings.html>).
- Наоборот, в корне неправильное решение, основанное на хорошо оптимизированном коде либо стандартных методах оптимизации (таких, как метод отжига, комбинация эвристик и т. п.), может набрать балл, близкий к максимальному, из-за неполноты набора тестов, невозможности жюри предугадать все возможные решения участников или специфичной структуры контрtests.

Сейчас данная система оценивания используется достаточно редко, как правило, на отборочных этапах различных олимпиад [1], однако иногда ее можно встретить и на всероссийских олимпиадах. Положительной чертой данной системы является необходимость для участников подвергать свои решения тщательному тестированию, в том числе самостоятельному автоматическому тестированию на своем компьютере прямо во время олимпиады. Стоит сказать, что это умение и раньше встречалось только у участников высокого уровня, а сейчас оно иногда отсутствует даже у членов команды, представляющей Россию на международной олимпиаде.

Рекомендуется использовать потестовую оценку «втемную» только в задачах с простым форматом ввода-вывода, легко поддающихся самостоятельному тестированию, в которых помимо этого имеется множество различных решений. Удачным примером

недавнего использования такой системы оценивания можно считать задачу «Массовый прогноз» всероссийской олимпиады 2013 года.

Потестовая оценка с частичными результатами

Данная система оценивания отличается от предыдущей тем, что участнику доступны протоколы проверки на некоторых заранее выбранных тестах, при этом в условии задачи никак не регламентируется и не описывается, какие именно тесты будут показаны в протоколе. Такая система применялась на международной олимпиаде по информатике с 2007 по 2009 год.

Этот подход позволяет жюри решить проблему потери участником всех баллов по невнимательности, описанную в предыдущем пункте, и при этом оставляет возможность «подловить» участника на крайних случаях или на тестах сложной структуры.

Главной *проблемой* данного подхода является низкий уровень формальности при выборе тестов с открытым протоколом проверки, что приводит участника к необходимости судить о полноте доступных ему тестов по различной косвенной информации, такой как номера претестов, их количество, время работы программы и тому подобное.

Представляется, что именно неформальность этого подхода стала причиной, по которой подобная система оценивания сейчас практически не используется. В несколько измененном виде она применяется на платформе для онлайн-соревнований Codeforces (<http://codeforces.com/>), но данная платформа не имеет ориентации на школьную аудиторию.

Оценка по группам с полным показом результатов

Начиная с 2010 года правила проведения международной олимпиады кардинально изменились [3]. Изменения в правилах многих других олимпиад (всероссийской олимпиады школьников, международной Жаутыковской олимпиады и др.) последовали за ней сразу или с небольшим отставанием.

На международной олимпиаде 2010 года в Канаде была предложена новая система оценивания: задачи разбиваются на подзадачи, каждая из которых имеет свои ограничения на входные данные и оценивается независимо от остальных подзадач, причем баллы за подзадачу начисляются только при прохождении всех тестов, соответствующих этой подзадаче. Участник может прямо во время тура увидеть полный протокол проверки своего решения.

Поскольку *на сегодняшний день это самая распространенная система оценивания в олимпиадных задачах по программированию*, то остановимся подробнее на ее плюсах и минусах.

Начнем с *плюсов*:

- Проверка стала более честной в том смысле, что каждая подзадача либо полностью решена, либо полностью не решена. Решения, которые работают в ограничениях подзадачи, часто (но не всегда) не получают частичных баллов. Под «честностью» понимается тот факт, что раньше начисление некоторого количества баллов за решения с ошибками в различных крайних

случаях целиком и полностью оставалось на усмотрение жюри, теперь же участник может заранее точно оценить баллы для любого своего решения, в том числе разбирающего какой-то частный случай.

- Жюри имеет возможность не ограничивать себя в количестве тестов и их структуре, что повышает общее качество проверки.
- Участник во время тура точно знает, получил ли он за задачу те баллы, которые ожидает от своего решения. Это избавляет от необходимости тратить нервы и время на перепроверку сданных задач.

Минусами подобной системы оценивания, часто обсуждаемыми в олимпиадном сообществе, являются:

- Частичная потеря школьниками навыков тестирования своих программ, а особенно — неумение автоматизированно тестировать свое решение, сравнивая результат работы на большом количестве небольших тестовых данных с результатом работы наивного решения (такая техника называется *стресс-тестированием*).
- Ступенчатость результатов, которая иногда приводит к невозможности провести четкие границы призеров и победителей. Например, на последней всероссийской олимпиаде по информатике (2015 год) неаккуратная работа жюри с группами тестов привела к ситуации, когда впервые с 2000 года не был выявлен абсолютный победитель олимпиады, — первое место поделили два участника.

Рекомендуется использовать систему оценивания по группам для задач, в которых имеется большое количество неправильных решений, каждое из которых ошибается или работает слишком долго только на тестах определенной структуры.

Также при использовании группировки тестов следует очень внимательно отнестись к выставлению ограничения по времени на работу программы. Следует убедиться, что оптимальное решение с худшей асимптотикой, чем предполагается для данной группы, работает ощутимо медленнее, чем самое неоптимальное решение с правильной асимптотикой. Примером задачи с хорошим подбором групп тестов может служить задача «Teams» с международной олимпиады 2015 года в Казахстане.

Задачи с открытыми тестами

Данный способ оценивания принципиально отличается от всех описанных выше, так как жюри вообще не требует от участника программу для запуска в тестирующей системе. Участнику сразу выдаются файлы, содержащие тестовые данные, которые он должен обработать и ответы по которым

должен сдать в тестирующую систему. При этом неважно, каким способом были получены ответы (если, конечно, не были нарушены правила олимпиады): участник может как найти решения для некоторых тестов вручную, так и написать решающую задачу программу.

Данная система оценивания достаточно редко применяется на всероссийской олимпиаде — в силу своей специфичности и того, что мало какие задачи подойдут под такой формат. В частности, в туре олимпиады точно не стоит делать более одной подобной задачи, иначе участники просто не справятся с правильным распределением своего времени. Последними задачами с открытыми тестами на заключительных этапах всероссийской олимпиады стали задачи «Несчастливые номера» (2008 год) и «Съезд кинозвезд» (2014 год).

В заключение стоит отметить, что организаторы олимпиад часто не ограничивают себя использованием одной из описанных выше систем оценивания в чистом виде, комбинируя их или используя разные системы в разных задачах. Так, например, на проводимой в Москве Открытой олимпиаде по программированию (<https://olympiads.ru/zaoch/>) популярна комбинированная схема, разработанная Е. В. Андреевой. Суть данной схемы заключается в том, что первые несколько групп тестов каждой задачи оцениваются только целиком, при этом результаты проверки доступны участникам сразу, а последняя подзадача проверяется «втемную» и, как правило, по тестам [2, 4]. Данная схема позволяет использовать лучшие стороны как оценки по группам (нельзя случайно набрать много баллов эвристическим решением или случайно не набрать ничего правильным решением по невнимательности), так и закрытой потестовой оценки (разнообразные баллы, требование от участника умения тестировать свое решение).

Литературные и интернет-источники

1. *Кирюхин В. М.* Методические рекомендации по проведению школьного и муниципального этапов всероссийской олимпиады школьников по информатике в 2012/2013 учебном году, <http://methodist.lbz.ru/lections/6/files/mrzd2013.pdf>
2. Московские олимпиады по информатике. 2002–2009 / под ред. Е. В. Андреевой, В. М. Гуровца, В. А. Матюхина. М.: МЦНМО, 2009.
3. Правила IOI. <http://www.ioi2010.org/rules.shtml>
4. Разбор задач заочного этапа X Открытой олимпиады школьников по программированию. <https://olympiads.ru/zaoch/2015-16/zaoch/analysis.pdf>
5. *Mares M.* Perspective on Grading Systems // *Olympiads in Informatics*. 2007. Vol. 1.