

АКАДЕМІЯ НАУК УКРАЇНИ
ИНСТИТУТ ПРОБЛЕМ МОДЕЛИРОВАНИЯ
В ЭНЕРГЕТИКЕ

На правах рукописи
УДК 621.391

КОНДРАТЮК Диана Лукинична

РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДА
ПОВЫШЕНИЯ НАДЕЖНОСТИ
ВЫЧИСЛИТЕЛЬНЫХ УСТРОЙСТВ,
РЕАЛИЗУЮЩИХ ЭВРИСТИЧЕСКИЕ АЛГОРИТМЫ

Специальность 05.13.13
Вычислительные машины, комплексы,
системы и сети

А В Т О Р Е Ф Е Р А Т
диссертации на соискание ученой степени
кандидата технических наук

Киев — 1993

11827910

Работа выполнена в департаменте кибернетики и вычислительной техники Севастопольского приборостроительного института.

Научный руководитель: кандидат технических наук, доцент Островский Владимир Ильич

Официальные оппоненты:
доктор технических наук, Ю.М. Коростиль,
доктор технических наук, профессор Ю.Г. Савченко

Ведущее предприятие: Институт Кибернетики АН Украины

Защита состоится 30 сентября 1993 г. в 10 час. 00 мин. на заседании специализированного совета К016.61.01 по присуждению ученых степеней в Институте проблем моделирования в энергетике Академии наук Украины по адресу:

252680, Киев – 164, ул. Генерала Наумова, 15.

С диссертацией можно ознакомиться в библиотеке Института проблем моделирования в энергетике Академии Наук Украины.

Автореферат разослан 27 июля 1993 г.

Ученый секретарь
специализированного совета,
кандидат технических наук

Э.П.Семагина

ЛННБ України ім.В.Стефаника



00815721 (0)

ЛННБ ім. В. Стефаника
АН України

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы. Надежность вычислений традиционно включает две основные составляющие — надежность технических средств и надежность программного обеспечения. При этом предполагается, что реализуемые аппаратно, программно или аппаратно — программно алгоритмы обладают следующими тремя свойствами:

1) определенностью, т.е. однозначностью интерпретации и понятностью для возможных исполнителей, благодаря чему процесс выполнения алгоритма носит механический характер;

2) результативностью, т.е. способностью приводить в тех случаях, для которых он создан, к получению результата после конечного числа шагов;

3) массовостью, т.е. пригодностью для решения любой задачи из некоторого класса задач.

Расширение сферы применения ЭВМ привело к возникновению проблем, при решении которых в той или иной степени используются модели рассуждений человека и выполняемых им операций при решении соответствующих задач. К числу таких проблем относится проблема распознавания образов и ряд других задач искусственного интеллекта, перевод с одного языка на другой, ряд логико — комбинаторных проблем и др. Методы решения подобных задач принято называть эвристическими или эвристиками.

Эвристический алгоритм определяется как алгоритм со следующими свойствами:

1. Он обычно находит хорошие, хотя не обязательно оптимальные решения.

2. Его проще (быстрее, с меньшими затратами) реализовать, чем любой известный точный алгоритм, если он вообще существует.

Атрибуты "хороший" и "обычно" в свойстве 1 имеют различный смысл в различных задачах.

С большой степенью приближения к истине можно сказать, что эвристические алгоритмы не обладают третьим свойством в определении алгоритма — массовостью. Иначе говоря, эвристические алгоритмы не для всех задач дают достоверные или точные решения. Поэтому можно говорить о вероятности получения правильного решения с помощью данного алгоритма и как следствие этого — о надежности вычислительного устройства (системы), реализующего данный алгоритм.

В диссертации исследуется общий метод повышения надежности вычислительного устройства, реализующего эвристический алгоритм, основанный на введении структурной избыточности. Этот метод позволяет повысить вероятность правильного решения задачи в том случае, когда нет алгоритма, дающего достоверное решение, т.е. с вероятностью, что оно правильное, равной единице. Наиболее типичным примером такой проблемы является проблема распознавания и классификации. К этой проблеме сводится большое число практически важных задач от автоматического ввода текстовой и числовой информации в ЭВМ до обнаружения полезных ископаемых. Во всех этих задачах получение правильного решения не является гарантированным, а зависит от случайного шума, поэтому говорят о вероятности ошибки при их решении, обусловленной не отказом технических средств или неточностью в написании программы, а несовершенством алгоритма.

Назовем ряд конкретных типов устройств, ориентированных на решение указанных задач. Это устройства контроля изделий по внешнему

виду — в диссертации рассмотрена задача контроля качества печатного монтажа при изготовлении печатных плат и специализированное устройство "Глаз", ориентированное на решение этой задачи, так называемые читающие автоматы — устройства распознавания буквенных и цифровых символов, реализуемые в самых различных формах, и предназначенные для автоматического ввода данных и программ в ЭВМ, специализированные устройства управления промышленными и военными объектами.

Задачи, решаемые этими устройствами, широко распространены как в промышленности, так и в военном деле, в научных исследованиях, и разработка методов и средств их эффективного решения в настоящее время представляется актуальной.

Цель диссертационной работы. Наряду с практической важностью поиска эффективных способов решения задач указанного класса не меньшее значение имеет и чисто научный интерес к методам, позволяющим повысить достоверность решения задачи в тех случаях, когда алгоритма ее решения в традиционном понимании этого термина нет, а есть эвристики; чаще всего для решения одной и той же задачи изобретено несколько эвристик. Ясно, что важнейшим путем повышения достоверности решения таких задач является поиск алгоритмов и улучшение эвристик. Этот процесс, по-видимому, будет развиваться, и следует полагать, что на этом пути будут получены основополагающие результаты в каждом отдельном направлении.

В данной работе сделана попытка найти некоторый универсальный подход, позволяющий уже в настоящее время повысить достоверность решения задач за счет использования числа

эвристик и их надлежащего подбора.

Цель данного исследования — разработка механизма объединения нескольких эвристических алгоритмов решения задачи для повышения достоверности результатов, проверка предложенных средств на нетривиальных задачах, имеющих практическое значение.

Предметом исследования является комплексное рассмотрение некоторых задач, решаемых эвристическими методами, их практической важности и путей решения. Построена математическая модель алгоритма, содержащего избыточные модули и предложены методы оценки важности модулей алгоритма, а также создана реализующая эти методы программа. На примере двух актуальных задач проверена эффективность предложенного метода повышения достоверности решения таких задач. Разработан и программно реализован оригинальный алгоритм автоматического исправления орфографических ошибок в русско- и украинскоязычных текстах, возникающих при вводе текстовой информации в ЭВМ с помощью специальных устройств, называемых читающими автоматами.

Метод исследования основан на сочетании теории цифровых схем, теории надежности, методов алгоритмизации и эвристического программирования. Теоретической базой рассматриваемых в работе алгоритмов и методов являются: теория дискретных устройств, теория распознавания образов, прикладная математика, методы алгоритмизации.

Разработанные алгоритмы написаны на языке Паскаль. При построении системы совместно работающих алгоритмов обнаружения дефектов печатного монтажа использовались методы автоматизации визуального технологического

контроля в электронном приборостроении.

Разработка алгоритма автоматической коррекции орфографических ошибок в украинско- и русскоязычных текстах потребовала довольно глубокого изучения морфологии русского и украинского языков, механизмов словоизменения и словообразования в этих языках.

Научная новизна. В диссертационной работе представлены следующие научные результаты, полученные в процессе исследования:

— сделано обобщение развитого в рамках теории избыточных цифровых систем метода резервирования с адаптивным голосованием на эвристические алгоритмы, что позволяет создавать более надежные вычислительные устройства;

— разработана математическая модель, отражающая модульную структуру алгоритма с избыточностью и предложен графо-аналитический метод вычисления функции надежности;

— построен алгоритм выявления значимых (важных) модулей в алгоритме, что позволяет выявлять наиболее ответственные этапы вычислительного процесса;

— предложен алгоритм обнаружения и автоматической коррекции ошибок в русско- и украинскоязычных текстах, основанный на морфемном анализе языка, что позволило создать оригинальную программную систему, обеспечивающую более надежный ввод текстовой информации в ЭВМ и отличающуюся компактностью представления словаря.

Практическая ценность. Теоретические результаты и разработанные алгоритмы реализованы в виде программ на языке Паскаль. Эффективность предложенного способа повышения надежности устройств, реализующих эвристические алгоритмы,

продемонстрирована на реальных задачах путем имитационного моделирования.

Метод найдет применение при решении широкого класса задач, относимых к искусственному интеллекту, обработке текстов, информационным системам.

Программа упорядочения элементов по значимости имеет широкую область применений и уже применяется не только к алгоритмам, но и к механическим и электронным системам.

Реализация результатов работы. Результаты диссертационной работы использовались в ряде хозяйственных и бюджетных разработок в департаменте кибернетики и вычислительной техники Севастопольского приборостроительного института (надежность математического обеспечения, системы машинного зрения). Методика и программа оценки важности элементов сложной системы передана для использования в КБ "Полет" г.Омск, ЦКБ "Коралл" г.Севастополь. Пакет автоматической коррекции орфографических ошибок передан для испытаний и пополнения словаря ряду заинтересованных организаций, в частности, институту Украинизмов Академии Наук Украины (г.Львов).

Апробация работы. Основные положения диссертационной работы докладывались и обсуждались на:

— Всесоюзной конференции "Методы и средства повышения эффективности и качества программных систем" (Севастополь, 1990 г.);

— Третьей Всесоюзной научно-технической конференции "Живучесть и реконфигурация информационно-вычислительных и управляющих систем" (Севастополь, 1991 г.);

— Второй Всесоюзной конференции "Перспективные информационные технологии в анализе

изображений и распознавании образов" (Ташкент, 1992 г.);

— Второй Международной конференции "Проблемы украинизации компьютеров" (Львов, 1992 г.);

— семинаре "Техническое и программное обеспечение ЭВМ, систем, комплексов и сетей". Научного Совета по комплексной проблеме "Кибернетика" при Президиуме АН Украины.

Публикации. По результатам исследований опубликовано восемь научных работ.

Структура и объем работы. Диссертационная работа состоит из введения, четырех глав, заключения, списка использованной литературы и приложения. Работа содержит 141 страницу машинописного текста. В приложении приведены листинги программ, иллюстрации их работы, а также акты внедрения.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационной работы, определена цель проводимых исследований, приведены основные материалы по научной новизне, теоретическим и практическим результатам, полученным соискателем при решении поставленных задач.

Показано, что расширение сферы использования средств вычислительной техники выдвигает новые требования: появился класс задач, решение которых опирается не на математические результаты, а на менее строгие посылки, называемые эвристиками.

На примере проблемы распознающих устройств выявлены особенности эвристических алгоритмов. Показано, что целый ряд задач, относящихся к различным областям науки и техники, может быть

отнесен к этому классу, и нет оснований полагать, что в ближайшее время не возникнут новые задачи, обладающие указанными свойствами.

Основная идея, положенная в основу повышения достоверности решения задач, для которых нет алгоритмов, обеспечивающих достоверное решение — использование нескольких различных алгоритмов решения одной и той же задачи с последующим принятием обобщенного решения, учитывающего достоверность результата каждого из алгоритмов. При этом учитывается тот факт, что с точки зрения входа — выхода эти алгоритмы различаются тем, что достоверность работы каждого из них зависит от входных данных. Иначе говоря, особенность рассматриваемого класса задач состоит в том, что любой алгоритм их решения имеет в пространстве исходных данных области, на которых он дает правильное решение и области, на которых он дает решение неправильное. Поэтому в некоторых задачах этого класса достоверность работы алгоритма может быть увязана с выдаваемым им решением, которое связано, очевидно, с входными данными.

Используя параллельно совокупность алгоритмов, мы можем надеяться, что области исходных данных, на которых каждый из них дает достоверное решение, покроют все пространство исходных данных. Проблема, таким образом, сводится к определению "областей достоверных решений" для каждого алгоритма. Предложен один из подходов к решению этой проблемы в случае задачи распознавания образов. Предлагается использовать общий подход, не опирающийся на специфику решаемой задачи, основанный на использовании механизма адаптации путем изменения весов пороговой функции.

В первой главе описывается предложенная нами

математическая модель системы алгоритмов (программ) решения одной и той же задачи, обосновывается выбор параметров пороговой решающей функции, позволяющий определить наиболее вероятное выходное значение для совокупности алгоритмов. Рассчитаны зависимости оценок вероятности ошибок избыточной системы алгоритмов от вероятностей ошибок входящих в нее алгоритмов и от числа этих алгоритмов.

Рассмотрены два варианта механизмов адаптации пороговой решающей функции, максимизирующих достоверность решения задачи с учетом вероятности ошибок для каждого конкретного результата, построена математическая модель процесса адаптации, а также процедура выбора наиболее достоверного решения избыточной системы алгоритмов.

Во второй главе предложена математическая модель эвристического алгоритма, отдельные блоки которого реализуются избыточной структурой из нескольких дублирующих друг друга модулей. По аналогии с принятыми в теории надежности технических систем вводятся понятия структурной функции и функции надежности алгоритма. Основное назначение вводимых понятий и математической модели — разработка методики выявления "узких" мест в так называемой схеме соединения модулей. Такие узкие места представляются модулями, недостоверность которых оказывает наиболее существенное влияние на достоверность результата решения всей задачи. Такие модули называются важными или значимыми по аналогии с блоками технических систем. Важность модуля может быть вычислена по функции надежности. Проведенный во второй главе анализ известных методов вычисления

функции надежности (раздел 2.2) показал, что, несмотря на наличие описанных в литературе нескольких методов вычисления функции надежности рассматриваемого класса структур, актуальной остается разработка методов с дифференцированной ориентацией: с одной стороны, нужны эффективные, хорошо формализованные алгоритмы, ориентированные на использование ЭВМ, с другой стороны, не отпала потребность в ручных методах анализа функции надежности, незаменимых при исследованиях, обучении студентов, общении специалистов.

Предложены две модификации метода вычисления функции надежности. Первая ориентирована на ручные вычисления при небольшом числе модулей (до восьми), она основывается на графическом представлении структурной функции в виде диаграммы Вейча, хорошо известной в теории дискретных устройств. Метод основан на свойствах монотонных булевых функций и соответствующих сокращенных нормальных булевых форм. Это позволяет формализовать его до степени алгоритма. В результате получается вторая модификация метода: разработаны и реализованы в виде программ на языке Паскаль два алгоритма вычисления функции надежности, один из них — алгоритм 1 является интерпретацией в терминах булевой алгебры известного метода включения — исключения. Алгоритм 2 является более быстродействующим, чем алгоритм 1 и особенно эффективен, если число переменных в структурной функции не превосходит числа разрядов в машинном слове. Вычисление функции надежности в нем сведено к нахождению покрытия структурной функции совокупностью непересекающихся интервалов

наименьшего ранга.

Как уже отмечалось, вычисление функции надежности не является в данной работе самоцелью. Разработанные алгоритмы и метод используются для упорядочения модулей алгоритма с избыточностью по отношению "важнее". Такое упорядочение позволяет:

1. Выделить модули, совершенствование которых или дублирование приводит к наиболее эффективному повышению достоверности решения задачи.

2. Более планомерно распределять ресурсы на тестирование реализующей алгоритм программы: более важные модули должны проходить более глубокое тестирование.

В качестве критерия важности выбран критерий Бирнбаума. Метод вычисления важности по Бирнбауму описан в разделе 2.5, а реализующая его программа представлена в приложении 1.

Демонстрация возможностей разработанных методов осуществляется в следующих двух главах — третьей и четвертой.

В третьей главе рассматривается использование предложенной методики для повышения надежности работы читающих устройств, часто называемых также читающими автоматами. Такие устройства относятся к числу периферийных устройств ЭВМ, и их основное назначение состоит в восприятии и автоматическом распознавании букв и цифр с последующим их вводом в память ЭВМ, иначе говоря, это устройства ввода информации с документов, автоматизирующие ввод с клавиатуры. При всем многообразии технической реализации читающих устройств, общими основными показателями их качества являются вероятность ошибочного распознавания символа (буквы или цифры), вероятность отказа от

распознавания (т.е. анализируемый символ не может быть отнесен ни к одному из символов алфавита) и быстродействие, измеряемое общим числом считанных символов. Легко видеть, что читающие устройства относятся к рассматриваемому в данной работе классу устройств.

Существует по крайней мере два направления применения предложенных методов для снижения числа ошибок и отказов от распознавания в читающих автоматах. Первое направление состоит в построении системы различных алгоритмов распознавания, оценке достоверности результата каждого из них и вычислении весов пороговой функции для принятия решения. Наряду с таким общим резервированием возможно и поэлементное резервирование отдельных блоков каждого алгоритма, например, блока выделения признаков.

В данной работе исследуется второе направление, суть которого заключается в следующем. По оценкам лингвистов избыточность естественных языков (для различных языков она различна) лежит в пределах 65—75 процентов. Для русского и украинского языков она составляет примерно 70 %. Поэтому ошибочно распознанная буква в слове может быть исправлена, а нераспознанная буква восстановлена, если найти способ эффективного использования избыточности языка. Анализ существующих систем обнаружения орфографических ошибок (раздел 3.1) показал, что:

1) в настоящее время нет программных продуктов, осуществляющих и обнаружение, и исправление ошибок в русско- и украинскоязычных текстах;

2) все известные программы учитывают знание о языке на уровне не выше флективных форм;

3) перспективные разработки в рассматриваем-

мой области должны быть направлены на снижение требований к памяти. Это особенно важно для рассматриваемого класса устройств, а также для построения издательских систем, где требования к словарному запасу особенно высоки.

В диссертации показано, что ошибки читающих устройств являются частным случаем орфографических ошибок, обусловленных слабым знанием языка или ошибками оператора, работающего с клавиатурой. На примере разработки оригинального алгоритма автоматического обнаружения и исправления орфографических ошибок и отказов в русско- и украинскоязычных текстах исследуется эффективность изложенных в гл. 1, 2 методов. Алгоритм (раздел 3.2) в первую очередь направлен на учет перечисленных факторов.

Научную основу предлагаемого подхода представляет использование результатов морфемного анализа языка для компактного представления словаря и получения знаний о структуре слов. Результаты морфемного анализа языка отображаются, в частности, в словаре морфем. В основе алгоритма обнаружения и коррекции ошибок, опирающегося на префиксальную часть словаря морфем, лежат следующие два положения. Во-первых, поскольку в слове по условию возможна только одна ошибка, то она может находиться либо в основе, либо в окончании. Основа, в свою очередь, делится на префикс (возможно не один), корень и суффикс (тоже возможно не один). Это сокращает объем необходимых комбинаторных вычислений при поиске правильного слова. Во-вторых, знания о сочетаемости каждого префикса с определенными корнями слов и основы — с определенными

окончаниями довольно часто позволяют делать заключения о правильности того или иного варианта коррекции ошибки.

В разделе 3.3 приводятся результаты испытания алгоритма на русских и украинских текстах.

В четвертой главе представлены результаты имитационного моделирования избыточной системы из шести алгоритмов для специализированного вычислительного устройства "Глаз". Основное назначение этого устройства — обработка изображений. Устройство "Глаз" представляет собой программируемый видеопроцессор и может использоваться как система технического зрения роботов и контрольно — измерительное устройство. Указанная система алгоритмов предназначена для использования устройства "Глаз" с целью обнаружения дефектов печатного монтажа при производстве изделий электронной техники. Все шесть алгоритмов решают одну и ту же задачу — обнаружение дефекта типа "короткое замыкание" и дефектов типа "разрыв проводника" в печатном монтаже платы электронного изделия. Демонстрируется работа механизма адаптации решающей функции в процессе функционирования системы.

Основные результаты работы:

1. Показано, что в практически важных случаях находят применение вычислительные устройства, алгоритмы которых не удовлетворяют важному свойству вычислительных алгоритмов — массовости. Результатом этого является тот факт, что соответствующие устройства в некоторых случаях выдают неверный результат или вообще не выдают результата.

2. На основе анализа существующих методов

повышения надежности вычислительных устройств путем введения избыточности предложен метод построения избыточных эвристических алгоритмов, обеспечивающий повышение достоверности результата.

3. Разработана математическая модель эвристического алгоритма, некоторые блоки которого реализуются избыточной структурой из нескольких дублирующих друг друга модулей.

4. Предложен метод вычисления функции надежности, являющийся алгебро-логической интерпретацией известного метода включения и исключения. Такая интерпретация позволяет получить более формализованный и более универсальный инструмент для вычисления функции надежности: метод удобен для ручных вычислений при небольшом числе модулей (до восьми), при большем числе переменных следует использовать построенные на его основе алгоритмы, различающиеся быстродействием и объемом необходимой памяти.

5. Предложенный подход экспериментально проверен на двух типах вычислительных устройств, реализующих эвристические алгоритмы:

1) для устройства типа "читающий автомат" разработан не имеющий аналогов алгоритм исправления ошибок и устранения отказов от распознавания, отличающийся значительной корректирующей способностью и компактностью используемого словаря. Показано, что этот алгоритм является частным случаем более общего алгоритма, позволяющего исправлять одиночные ошибки, возникающие при вводе текстовой информации в ЭВМ с клавиатуры, что позволяет обеспечить защиту ЭВМ от некорректных данных;

2) для системы технического зрения (видео-процессора) создана и исследована иллюстри-

рующая предложенный подход экспериментальная система алгоритмов контроля качества изготовления печатного монтажа, позволяющая повысить надежность работы устройства при решении указанной задачи.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Кондратюк Д.Л. Автоматическое исправление орфографических ошибок в текстовых данных на основе морфемного анализа. УкрНИИНТИ, № 464 — Ук90 от 12.03.90.

2. Бутаков Е.А., Кондратюк Д.Л., Манохин И.В. Выделение важных элементов при оценке надежности больших механических систем. — К.: Общество "Знание" Украины, 1991. — 20 с.

3. Запевалин А.А., Кондратюк Д.Л. Оценка важности отказов элементов живучих систем с деградацией. (Тез. докл. 3-й Всесоюзной научно-технической конференции "Живучесть и реконфигурация информационно-вычислительных и управляющих систем"). — М., 1991, — С.83.

4. Кондратюк Д.Л. Корекція орфографічних помилок в українському тексті. — Управляющие системы и машины, 1992, № 4, — С. 51 — 55.

5. Кондратюк Д.Л. Автоматизация исправления орфографических ошибок при подготовке программной документации. Всесоюзная конференция "Методы и средства повышения эффективности и качества программных систем". 21 — 22 мая 1990 г.: Тез. докл. — Севастополь, 1990. — С. 42 — 44.

6. Кондратюк Д.Л. Структурное резервирование как метод повышения надежности алгоритмов распознавания (Тез. докл. конф. "Перспективные

информационные технологии в анализе изображений и распознавании образов". Ташкент, — 1992. — С. 14 — 15.

7. Кондратюк Д.Л. Корекція орфографічних помилок в українському тексті. Друга Міжнародна конференція "Проблеми українізації комп'ютерів" (Тези доп.), Львів. — 1992. — С. 47.

8. Кондратюк Д.Л. Методы введения избыточности при построении высоконадежных баз данных. (Отчет о НИР, рук. Е.А.Бутаков). № гос.рег. 80035802, СПИ, Севастополь. — 1985. — С. 116 — 121.

AB 27.918

AB 27.918