

ИНТЕЛЛЕКТУАЛЬНАЯ ПОДСИСТЕМА ДЛЯ РЕШЕНИЯ КРИМИНАЛИСТИЧЕСКИХ ЗАДАЧ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Введение

Среди информационных систем [1–11] выделяют класс систем [1, 3, 8], которые получили название *фактографических систем (ФС)*. Принято [8], что *автоматизированные информационные системы формирования фактографических данных (АИСФФД)* являются разновидностью ФС: в них имеется человек-оператор и реализован алгоритм генерирования фактографических данных. Для повышения эффективности работы *автоматизированных средств обеспечения информационной безопасности (АСОИБ)* они содержат средства интеллектуальной подсистемы и АИСФФД.

Авторами ведутся исследования по трем направлениям: 1-е — разработка интеллектуальной подсистемы, являющейся составной частью информационной системы, реализующей АСОИБ; 2-е — разработка [2, 4, 5, 8, 9, 11] средств морфологического анализа текста электронного документа (ЭлД) для АСОИБ; 3-е — разработка [3, 8, 10] средств динамического тестирования испытуемого для обеспечения возможности принимать решение о подготовленности человека-оператора (*лица, принимающего решения (ЛПР)*).

Далее основное внимание будет уделено не самому материальному объекту (документу), а именно информации, содержащейся в этом объекте или передаваемой по каналам связи. Это объясняется тем, что такая информация позволяет идентифицировать исполнителя ЭлД (например, установить нарушителя, злоумышленника). В рамках 1-го направления была предложена схема ИС АСОИБ, представленная на рис. 1. ИС АСОИБ имеет 2 основных режима работы.

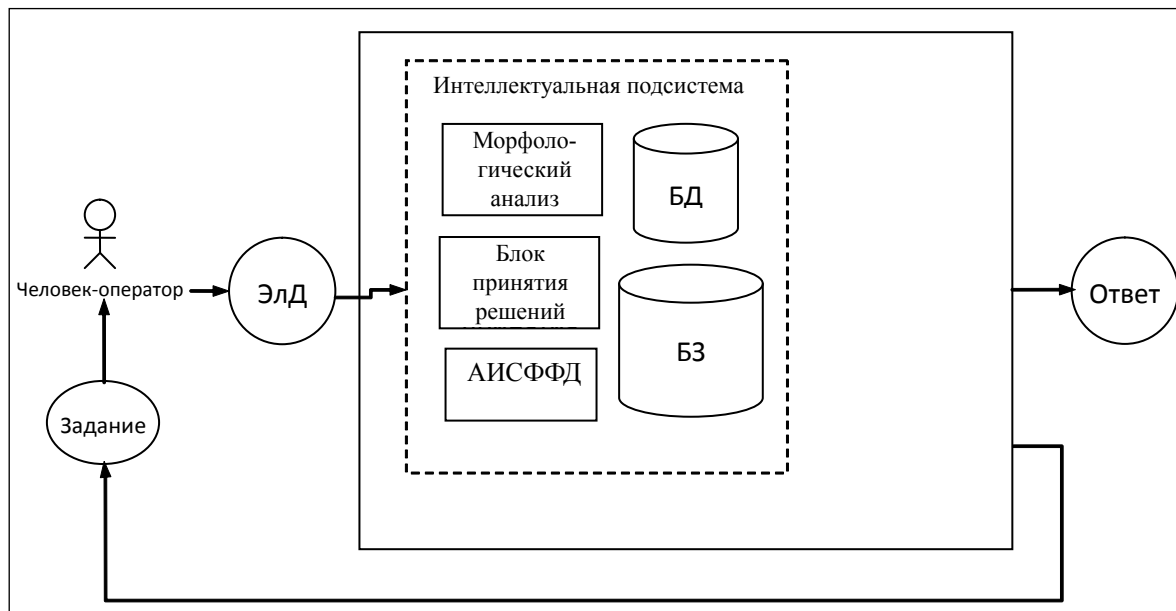


Рис. 1. Схема ИС АСОИБ

В первом режиме человек-оператор создает ЭлД, который поступает в ИС АСОИБ. В интеллектуальной подсистеме выполняется морфологический анализ и принимается решение об исполнителе ЭлД. В БД и в базе знаний (БЗ) хранятся все необходимые данные для морфологического анализа и принятия решения. Окончательный ответ поступает на выход АСОИБ.



Во втором режиме человек-оператор получает тестовое задание и формирует свой ответ на него в виде ЭЛД, который поступает в ИС АСОИБ. В интеллектуальной подсистеме по результату выполненного задания испытуемым (человеком-оператором) либо принимается решение о прекращении испытания и происходит последующая выдача окончательного ответа на выход АСОИБ, либо с помощью АИСФФД формируется следующее новое задание.

Идентификация исполнителя текста и морфологический анализ

Для реализации интеллектуальной подсистемы 1-го режима работы АСОИБ предлагается [2] подход к идентификации исполнителя текстов с помощью исследования частотных грамматических характеристик текста и выявления синтаксических особенностей. Подход применим для АСОИБ, когда имеется один или более текстов, для которых точно известен исполнитель, и требуется установить, является ли он исполнителем каких-либо других текстов. Предлагаемый подход подразумевает выполнение нескольких этапов. **Первый** этап предполагает установление (выявление) признаков (таких, как часть речи, род, число, падеж и т. д.) для каждого слова текстового документа. **На втором** этапе выделяются все пары, а затем все тройки идущих подряд слов текстового документа в рамках каждого его предложения. При этом в записи выделяемых пар и троек конкретные словоформы заменяются соответствующими им грамматическими признаками, получаемыми по результатам этапа 1. **На третьем** этапе выполняется оценка частотных характеристик (частота употребления служебных слов и т. п.). **На четвертом** этапе алгоритм машинного обучения (или комбинация нескольких алгоритмов) подстраивается под признаки, полученных на этапе 3 для всех текстовых документов с известным исполнителем. **На пятом** этапе выполняется идентификация исполнителя по электронному текстовому документу. На выходе получается решение о принадлежности данного документа конкретному исполнителю и вероятностная оценка достоверности этого решения.

При одношаговой процедуре [5] анализ текста происходит последовательно от слова к слову. Контекст употребления учитывается в момент анализа слова, в результате чего ему приписывается набор морфологических признаков (НМП), являющийся конечным результатом для данного слова. При двухшаговой процедуре на первом шаге для каждого слова вне зависимости от порядка следования по тексту (что позволяет распараллелить вычисления) и без учета контекста употребления находят наборы морфологических признаков (*бесконтекстный морфологический анализатор* – БКМА). Для омонимов может быть найдено несколько возможных НМП. Затем, на втором шаге, за счет рассмотрения контекста употребления слова и найденных НМП для каждого слова осуществляется выбор единственного НМП, который приписывается слову и является конечным результатом работы алгоритма для данного слова.

Рассмотрим кратко *контекстно-зависимый морфологический анализатор* (КЗМА).

Для построения *двухшагового КЗМА* (ДКЗМА) может быть использован, например, метод скрытых марковских моделей (СММ). Данный метод применим для реализации второго шага ДКЗМА. В работе рассмотрено применение СММ на втором шаге ДКЗМА, причем в качестве ограничения контекст будет учитываться только в рамках одного предложения. Другие подходы к построению ДКЗМА, алгоритм работы БКМА и вопросы, связанные с построением одношагового КЗМА, в данной работе рассмотрены не будут.

Экспериментальное исследование описанного ДКЗМА показало положительные результаты на неразмеченном текстовом корпусе из 4000 документов средним объемом 300 слов каждый. Точность определения НМП составила порядка 92 %. Если провести обучение отдельно для различных известных исполнителей, то подход позволяет установить, какому из M исполнителей наиболее вероятно принадлежит анализируемый текст (или тексты) ЭЛД.

Предполагается [4, 8], что на основе морфологического анализа определяют части речи слова и его морфологические признаки (падеж, число, род и т. д.). Результаты такого анализа слов предложения являются основанием для выявления связей между словами.



Для выполнения морфологического разбора необходимо (как минимум) наличие словаря лемм (канонических форм слов) и лексем (всех возможных словоформ). В таком словаре каждой лемме ставятся в соответствие все возможные словоформы, образованные на ее основе (каноническая форма также является словоформой), каждой из которых, в свою очередь, ставится в соответствие вектор морфологических признаков (МП). Для достижения более высокой эффективности разбора требуется наличие словарей имен собственных, аббревиатур и сокращений русского языка. Повышение эффективности достигается за счет увеличения значения оценки вероятности детерминированного морфологического разбора, описанного ниже, для слов, относящихся к данным группам.

Целью морфологического разбора слова t является определение вектора признаков:

$$\mathbf{m}_i = \left[m_i^{pos}, m_i^{case}, m_i^{tense}, m_i^{gender}, m_i^{plural}, m_i^{person} \right], \quad (1)$$

где $m_i^{pos} \in \{1, 2, \dots, 13\}$ – часть речи, $m_i^{case} \in \{0, 1, \dots, 6\}$ – падеж, $m_i^{tense} \in \{0, 1, \dots, 4\}$ – время глагола, $m_i^{gender} \in \{0, 1, 2, 3\}$ – род, $m_i^{plural} \in \{-1, 0, 1\}$ – форма множественного ($m_i^{plural} = 1$) или единственного ($m_i^{plural} = -1$) числа, $m_i^{person} \in \{0, 1, 2, 3\}$ – лицо. Если часть речи не обладает тем или иным МП, то его значение принимается равным нулю.

Рассматривается неполный набор существующих признаков, однако при необходимости расширения пространства признаков вектор (1) может быть успешно дополнен соответствующими компонентами. Далее предполагаем, что появление слова в некоторой форме в тексте предложения является случайным событием.

Для определения вектора признаков в словаре лексем выполняется поиск словоформы, совпадающей по написанию с t , и полагается, что слово t может обладать такими же признаками, как и найденная словоформа. В силу омонимии (в том числе в случаях, когда формы двух и более склонений совпадают) для слова t может быть выбрано несколько вариантов разбора. Введем множество детерминированного морфологического разбора, соответствующее слову t :

$$M^d = \begin{cases} \{(\mathbf{m}_i, C_i)\}_{i=1}^n, & \text{если найдены слова, сходные с } t \text{ по написанию,} \\ \emptyset, & \text{если сходные с } t \text{ по написанию слова не найдены,} \end{cases} \quad (2)$$

где C_i – количество слов в словаре лексем, сходных с t по написанию и имеющих одинаковые признаки \mathbf{m}_i , n – количество различных вариантов разбора. Под детерминированным морфологическим разбором понимается определение признаков слова на основе имеющихся в словаре лексем словоформ.

Априорная вероятность того, что слово t имеет признаки \mathbf{m}_i , оценивается по формуле:

$$P_i^d = \frac{C_i}{\sum_{j=1}^n C_j}, \quad i = 1, \dots, n. \quad (3)$$

В случаях, когда $M^d = \emptyset$, применяется процедура эвристического морфологического разбора, основанная на следующей эвристике:

Чем большую долю среди k первых лексем из словаря в порядке возрастания метрики $D_g^{hp}(t, t_g) = \min(|t|, |t_g|) - y_g$ (где k – натуральное число, являющееся входным параметром алгоритма, $g = 1, \dots, d$; d – число лексем в словаре, t_g – лексема из словаря, y_g – максимальное количество символов, идущих подряд от конца и одинаковых для слов t и t_g , а $|t|$ и $|t_g|$ – соответствующие длины слов) составляют лексемы, имеющие одинаковые признаки, тем выше оценка вероятности того, что t будет ими обладать.

По аналогии с (2), множество эвристического морфологического разбора имеет вид:

$$M^h = \{(\mathbf{m}_i, C_i)\}_{i=1}^e, \quad (4)$$

где e – количество различных \mathbf{m}_i среди выбранных k лексем.



При этом оценка априорной вероятности того, что слово t обладает признаками m_i , имеет вид:

$$P_i^h = \frac{\xi_i}{\sum_{j=1}^e \xi_j}, \quad i=1, \dots, e, \quad (5)$$

$$\xi_i = \sum_{m_i} y_g, \quad (6)$$

где $\sum_{m_i} y_g$ обозначает сумму всех y_g , относящихся к словоформам (из выбранных k лексем), обладающим признаками m_i . Формула (5) является общим случаем формулы (3), так как при наличии одинаковых по написанию слов, имеющих разные признаки, $\xi_i = \sum_{m_i} y_g = C_i$.

Для морфологического разбора предложения требуется выполнить описанную процедуру для каждого его слова t_z , $z=1, \dots, s$, где s — число слов в предложении. Таким образом, слову t_z соответствует множество морфологического разбора:

$$M_z = \begin{cases} M_z^h, & \text{если } M_z^d = \emptyset \\ M_z^d, & \text{если } M_z^d \neq \emptyset \end{cases}, \quad (7)$$

каждый элемент m_{iz} которого имеет оценку вероятности:

$$P_{iz} = \begin{cases} P_{iz}^h, & \text{если } M_z^d = \emptyset \\ P_{iz}^d, & \text{если } M_z^d \neq \emptyset \end{cases}. \quad (8)$$

Полученные множество M_z и оценки вероятностей P_{iz} реализуют процедуру морфологического анализа предложения без учета контекста употребления слов. Исключение контекста слов из анализа не позволяет разрешить синтаксическую омонимию, а следовательно, отрицательно сказывается на его эффективности.

В целях повышения эффективности морфологического разбора для АСОИБ посредством анализа контекста употребления слов предлагается использование СММ. Применение СММ для анализа отдельно взятого предложения из корпуса выполняется следующим образом. В СММ слова в предложении выступают в качестве наблюдаемой последовательности (в порядке следования), а в качестве скрытых состояний модели выступают варианты морфологического разбора каждого слова соответственно. Для АСОИБ предлагается метод подготовки начальных данных, используемых для СММ в задаче определения частей речи слов с учетом контекста их употребления. Задача сужена до определения частей речи слов для большей наглядности, однако все дальнейшие рассуждения справедливы и для расширенного пространства признаков.

На рис. 2 представлен фрагмент СММ, построенной для некоторого предложения. Для нахождения наиболее правдоподобной последовательности скрытых состояний необходимо знать оценки вероятностей $P_i(m^{(i)})$ и $P(m^{(i+1)}|m^{(i)})$, где $m^{(i+1)}$, $m^{(i)}$ — некоторые конкретные значения анализируемого признака для слов t_{i+1} и t_i соответственно. $P_i(m^{(i)})$ показывает, какова оценка априорной вероятности того, что слово t_i обладает признаком $m^{(i)}$. $P(m^{(i+1)}|m^{(i)})$ показывает, какова оценка вероятности того, что слово t_{i+1} обладает признаком $m^{(i+1)}$ при условии того, что t_i обладает признаком $m^{(i)}$.



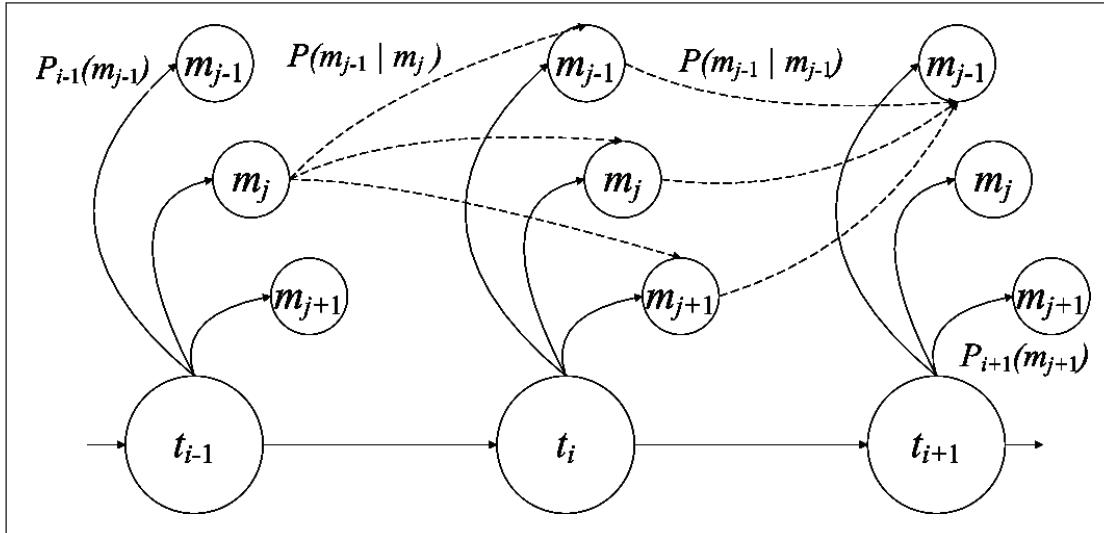


Рис. 2. Фрагмент СММ для некоторого условного предложения

$P_i(m^{(i)})$ показывает, какова оценка априорной вероятности того, что слово t_i обладает признаком $m^{(i)}$. $P(m^{(i+1)} | m^{(i)})$ показывает, какова оценка вероятности того, что слово t_{i+1} обладает признаком $m^{(i+1)}$ при условии того, что t_i обладает признаком $m^{(i)}$. Согласно принятому допущению, обладание слова тем или иным признаком «часть речи» является случайным событием. Тогда оценка вероятности того, что слово t_{i+1} обладает признаком $m^{(i+1)}$ при условии того, что t_i обладает признаком $m^{(i)}$, имеет вид:

$$P(m^{(i+1)} | m^{(i)}) = \frac{\left(\frac{w}{b}\right) \sum_k b_k P(m_k^{(1)} = m^{(i)}) P(m_k^{(2)} = m^{(i+1)})}{\sum_j w_j P(m_j = m^{(i)})}, \quad (9)$$

где $m^{(i+1)}$ и $m^{(i)}$ соответствуют некоторым конкретным значениям признака для слов t_{i+1} и t_i соответственно, а $P(m^{(i)} m^{(i+1)})$ показывает, какова оценка априорной вероятности того, что в предложении могут встретиться два слова t_i и t_{i+1} , идущие подряд и обладающие признаками $m^{(i)}$ и $m^{(i+1)}$ соответственно, w — общее количество слов в корпусе (в том числе совпадающих по написанию), w_j — общее количество слов, имеющих j -е написание, b — общее количество биграмм в корпусе (в том числе совпадающих по написанию), b_k — общее количество биграмм, имеющих k -е написание, m_j — часть речи слова с j -м написанием, $P(m_j = m^{(i)})$ — оценка вероятности того, что часть речи $m^{(i)}$ слова с j -м написанием правильно определена, $m_k^{(1)}$ и $m_k^{(2)}$ — соответственно части речи первого и второго слова, образующих бигramму, $P(m_k^{(1)} = m^{(i)})$ — оценка вероятности того, что часть речи $m^{(i)}$ первого слова биграммы правильно определена, $P(m_k^{(2)} = m^{(i+1)})$ — оценка вероятности того, что часть речи $m^{(i+1)}$ второго слова биграммы правильно определена. Под биграммой понимается пара слов, находящихся в предложении и идущих подряд. При этом в рамках описываемого метода знаки препинания внутри предложения игнорируются.

Рассмотрим кратко случай, когда слово является первым в предложении, а следовательно, не имеет предшествующего слова. Введем дополнительное обозначение: s — общее количество предложений; s_p — позиция в предложении p -го слова. В приведенных выше рассуждениях оценки вероятностей вычислялись с учетом написания слов. Далее речь пойдет о конкретных случаях употребления слова. Т. е. слову с некоторым написанием соответствует множество случаев его употребления. Для решения задачи генерации правил морфологического разбора используются выражения (9), (10) и (11).



$$P(m^{(i)} | s_i = 1) = \frac{\sum_{p \in W} \chi(s_p) \cdot P(m_p = m^{(i)})}{s}, \quad (10)$$

$$\chi(v) = \begin{cases} 1, & v = 1, \\ 0, & v \neq 1. \end{cases} \quad (11)$$

где m_p — часть речи p -го слова корпуса. Необходимый специальный генератор можно частично заимствовать из [3]. В совокупности с оценками вероятностей корректности морфологического разбора полученные выражения позволяют полностью задать параметры СММ. Результатом морфологического разбора слов предложения с учетом контекста их употребления будет являться решение задачи определения наиболее правдоподобной последовательности состояний СММ, соответствующей наблюдаемой последовательности. Решение этой задачи может быть получено, например, с помощью алгоритма Витерби [8].

Исследование эффективности морфологического анализа. Для проведения исследований вручную были отобраны 50 предложений из текстов статей заданного корпуса. Далее отобранные предложения были вручную размечены, т. е. каждому слову были приписаны морфологические признаки либо метка, указывающая, что при оценке эффективности его следует пропустить (для случаев, когда часть речи слова не была представлена в словаре лексем). Всего подготовленная выборка насчитывала 813 слов, из которых 655 словам были приписаны морфологические признаки.

В исследовании оценивались следующие показатели эффективности: **РЕ** — доля случаев по выборке, при которых морфологические признаки слова определялись корректно без учета контекста его употребления; **РС** — доля случаев по выборке, при которых среди наиболее вероятных вариантов (т. е. для которых $P_{\tilde{e}}$ из формулы (8) равен $\max P_{\tilde{e}}$) классификации МП слова присутствовал корректный вариант; **РД** — доля случаев по выборке при определении части речи без учета контекста употребления анализируемого слова, при которых часть речи была определена корректно; **РМ** — доля случаев по выборке при определении части речи с учетом контекста употребления анализируемого слова, при которых часть речи была определена корректно. Результаты экспериментальных исследований сведены в таблице 1.

Таблица 1. Результаты экспериментальной оценки показателей эффективности

Показатель эффективности	Значение показателя эффективности (в процентах)
РЕ	71
РС	94
РМ	90
РД	94

По итогам проведенного исследования можно сделать вывод о достижении достаточных значений показателей эффективности для использования предложенных методов в составе интеллектуальной подсистемы АСОИБ в криминалистике. Исследование проводилось с использованием сокращенного грамматического словаря, что явилось причиной снижения эффективности по рассмотренным показателям относительно достижимых на практике значений. Таким образом, имеет место тенденция к повышению эффективности методов при дальнейшем наполнении грамматического словаря. На данном этапе ведется разработка методов автоматизированного наполнения грамматического словаря на основе интеллектуального анализа текстовых документов, по итогам которой ожидается повышение показателя **РМ** до 98 %.



Динамическое тестирование человека-оператора (ЛПР)

Для реализации 2-го режима работы АСОИБ используется (на основе последовательного анализа [7, 10]) динамическое тестирование испытуемого (человека-оператора) в АИСФФД интеллектуальной подсистемы. Для того чтобы начать применять последовательный анализ А. Вальда для заданного набора признаков объекта задания и испытуемого, необходимо знать оценки вероятностей α , β , p_0 и p_1 (эти 4 числа выбираются (см. [7]) из практических соображений), а также правило принятия решений, по которому определяется дефектность (или годность) конкретного изделия из партии и правило принятия решений, по которому определяется дефектность (или годность) всей партии изделий.

Предположим, что все признаки (характеристики) объекта задания представлены одним обобщенным признаком, являющимся заданным признаком. Тогда по аналогии с приемочным контролем (опираясь на [7, 10]) для тестирования ЛПР (испытуемого) введем:

p — вероятность дефекта для заданного признака объекта задания; при $p \geq p_1$ предпочитают забраковать испытуемого (т. е. полагают, что он не подготовлен, так как допускает слишком много ошибок), а при $p \leq p_0$ предпочитают принять решение о подготовленности испытуемого;

α — вероятность принять решение о неподготовленности испытуемого для работы с АСОИБ (для заданного признака объекта задания) в случае, когда он подготовлен (т. е. $p \leq p_0$);

β — вероятность принять решение о подготовленности испытуемого для работы с АСОИБ (для заданного признака объекта задания) в случае, когда он не подготовлен (т. е. $p \geq p_1$);

роль партии изделий выполняет набор реализаций значений для заданного признака объекта задания, сформированных испытуемым (набор объектов заданий);

роль изделия из партии выполняет конкретизация (реализация значений) испытуемым заданного признака объекта задания;

d_m — число дефектов среди первых m реализаций испытуемым значений для заданного признака объекта задания.

Пусть $p_0 < p_1$. В соответствии с введенными обозначениями справедливо [7, 10]:

$$a_m = h_0 + s \cdot m, \quad r_m = h_1 + s \cdot m,$$

где

$$s = m \frac{\ln \frac{1-p_0}{1-p_1}}{\ln \frac{p_1}{p_0} - \ln \frac{1-p_1}{1-p_0}}; \quad h_0 = \frac{\ln \frac{\beta}{1-\alpha}}{\ln \frac{p_1}{p_0} - \ln \frac{1-p_1}{1-p_0}}; \quad h_1 = \frac{\ln \frac{1-\beta}{\alpha}}{\ln \frac{p_1}{p_0} - \ln \frac{1-p_1}{1-p_0}};$$

a_m — нижняя граница, при пересечении которой по текущему заданному признаку объекта задания принимается решение о том, что испытуемый подготовлен;

r_m — верхняя граница, при пересечении с которой принимается решение о том, что испытуемый не подготовлен.

При последовательном анализе испытуемый выполняет трудоемкое исследование характеристик объектов заданий выборки при тестировании до тех пор, пока выполняется условие [7, 10]: $a_m < d_m < r_m$. Тестирование немедленно прекращают (и принимают решение), когда d_m в первый раз нарушает это условие:

если $d_m \geq r_m$, то забраковать испытуемого;

если $d_m \leq a_m$, то испытуемый подготовлен.

Значения a_m и r_m на практике вычисляют заранее, до проведения тестирования испытуемого. При этом если приемочное число a_m и браковочное число r_m не целые числа ($a_m > 0$, $r_m > 0$), то их можно заменить [7]: a_m на $\lfloor a_m \rfloor$; r_m на $\lceil r_m \rceil$, где $\lfloor x \rfloor$ — наибольшее целое, такое, что $\lfloor x \rfloor < x$, а $\lceil x \rceil$ — наименьшее целое, такое, что $x < \lceil x \rceil$.



Исследователи А. Вальд и Дж. Вольфовиц доказали теорему (о минимуме средних размеров выборки, т. е. числа наблюдений), основной смысл которой сводится к тому, что «из всех критериев одной и той же мощности последовательный критерий отношений вероятностей требует в среднем наименьшее число наблюдений». Исследования эффективности показали, что последовательный критерий дает выигрыш не менее 47 % необходимого числа наблюдений по сравнению с обычным наиболее мощным критерием.

Таким образом, применяя при тестировании последовательный анализ, можно в некоторых случаях существенно (примерно до двух раз) сократить число испытаний (объем выборки). При использовании предлагаемого метода тестирования значительно снижается нагрузка на высококлассных операторов (лиц, принимающих решения), а также повышается объективность оценок уровня подготовленности испытуемых и значительно расширяется перечень характеристик, по которым появляется возможность принять решение об испытуемом за отведенное (ограниченное) на тестирование время. Отметим, что в случае очень длительного испытания (что возможно при последовательном анализе), когда ресурсы ограничены, применяют так называемую *усеченную схему проведения последовательного анализа* [7. С. 90]. В этом случае заранее задается ограничение на максимально возможное число испытаний (проверок) N_{\max} . Как только выделенные ресурсы будут исчерпаны (выполнено максимально возможное число проверок), испытания прекращаются и принимается решение по результатам этих проверок.

Результаты тестирования. Применим интеллектуальную подсистему и АИСФФД [8] для тестирования [10] испытуемых. Для проведения экспериментов по тестированию была использована специальная эталонная выборка объектов (≈ 700 заданий). Были подготовлены экспериментальные выборки заиндексированных образцов объектов, характеризующие каждого испытуемого в роли ЛПР. На основе эталонной выборки были выявлены все ошибки в признаках, допускаемые испытуемыми. Для каждого испытуемого была известна оценка показателя \hat{p} . Результаты последовательного анализа представлены на рис. 3 и 4.

Тестирование испытуемого (рис. 3) показывает, что принято решение о его подготовленности (уровень его ошибок достаточен для выполнения задания), а другое тестирование (рис. 4) показывает, что он не подготовлен (уровень его ошибок недостаточен для выполнения задания). Отметим, что этот испытуемый в обоих случаях характеризуется показателем $\hat{p} = 0,3$, что полностью согласуется с результатами его тестирования и требованиями к его подготовленности.

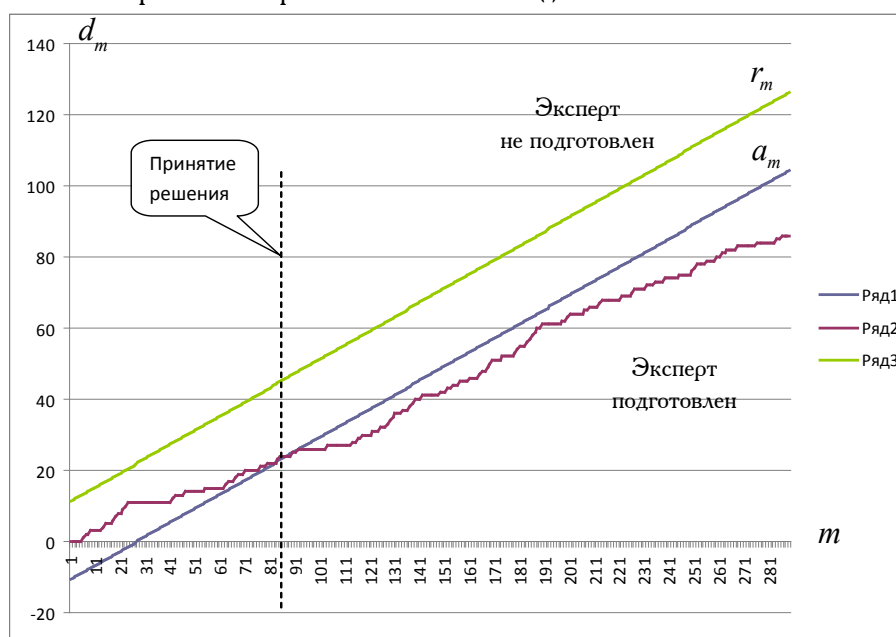


Рис. 3. $\alpha=0,01$, $\beta=0,01$, $\rho_0=0,35$ и $\rho_1=0,45$ ($\hat{p}=0,3$)



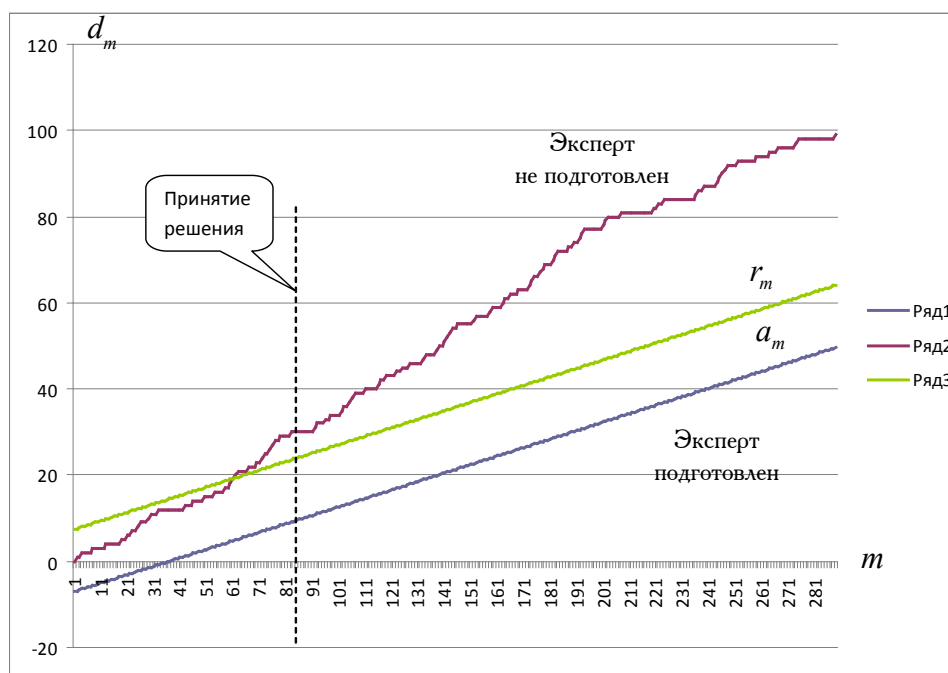


Рис. 4. $\alpha=0,01$, $\beta=0,01$, $\rho_0=0,15$ и $\rho_1=0,15$ ($\hat{\rho}=0,3$)

Выводы

Кратко представлена концепция построения АСОИБ. Представлен вариант возможной схемы информационной системы для реализации АСОИБ.

Разработан алгоритм морфологического анализа для принятия решения об исполнителе электронного документа. Выполнена реализация этого алгоритма в виде программного обеспечения, которое может быть использовано для АСОИБ. Экспериментальная проверка показала что, алгоритм работоспособен. Разработка эффективных правил принятия решения и программного обеспечения в области морфологического анализа электронного документа представляется весьма актуальной и полезной для решения задач информационной безопасности.

На основе последовательного анализа разработан вариант динамического тестирования испытуемого (человека-оператора) в АИСФФД интеллектуальной подсистемы АСОИБ. Экспериментальная проверка и исследование алгоритма динамического тестирования испытуемого показали что, алгоритм работоспособен. Полученные результаты помогают решить задачу обеспечения информационной безопасности.

По результатам проведенных исследований были успешно получены различные охранные документы Роспатента.

СПИСОК ЛИТЕРАТУРЫ:

1. Кулик С. Д. Учет человека-оператора в контуре АФИПС // Безопасность информационных технологий. 2003. № 2. С. 30–39.
2. Кулик С. Д., Ткаченко К. И., Лукьянов И. А. Идентификация исполнителя текстов по частотно-грамматическим характеристикам и синтаксическим особенностям // Безопасность информационных технологий. 2011. № 1. С. 108–110.
3. Кулик С. Д., Ткаченко К. И. Разработка генераторов для обеспечения информационной безопасности // Безопасность информационных технологий. 2010. № 1. С. 87–89.
4. Кулик С. Д., Ткаченко К. И., Лукьянов И. А. Методы морфологического анализа слов русского языка в системе фактографического вопросно-ответного поиска по законодательной и нормативной документации // Сборник трудов XX Междуна-



родной конференции «Информатизация и информационная безопасность правоохранительных органов», 24–25 мая 2011 г. М.: Академия управления МВД России, 2011. С. 420–424.

5. Кулик С. Д., Лукьянов И. А. Применение скрытых марковских моделей для контекстно-зависимого морфологического анализа текста // Тезисы докладов. X Всероссийская научная конференция «Нейрокомпьютеры и их применение» НКП-2012. Москва, 20 марта 2012 г. М.: МГППУ, 2012. С. 52–53.

6. Войтовецкий А. С., Сергеев М. С. Методика построения систем для задач проектирования систем розничного кредитования // Научная сессия МИФИ. 2010. Сборник научных трудов. М.: МИФИ, 2010. Т. 3. – С.141.

7. Вальд А. Последовательный анализ. М.: ГИФМЛ, 1960. – 328 с.

8. Кулик С. Д., Николец Д. А., Ткаченко К. И., Лукьянов И. А. Методы и средства повышения эффективности информационных систем (нейронные сети, криминалистика, формирование фактографических данных, морфологический анализ). Том 2: Системы. Изд-во «Радиотехника» (Деп. в ВИНТИ 05.05.2011, № 207-В2011; Библ. указат. № 7 (473)). М., 2011. – 223 с.

9. Кулик С. Д., Ткаченко К. И., Лукьянов И. А. Свидетельство на базу данных Российской Федерации № 2010620310 «QASE.Morphology.DB v.1.0» (QASEMDB) / С. Д. Кулик, К. И. Ткаченко, И. А. Лукьянов (Россия). Заявка № 2010620134; Заяв. 01.04.2010; Зарегистр. 28.05.2010; Оpubл. Бюл. № 3 (72). Ч. 2. С. 476 (Роспатент).

10. Кулик С. Д., Ткаченко К. И. Использование методов последовательного анализа в задаче тестирования экспертов-почерковедов // Материалы 3-й Международной научно-практической конференции «Теория и практика судебной экспертизы в современных условиях» (г. Москва, 25–26 января 2011 г.). М.: МГЮА, 2011. С. 539–541.

11. Кулик С. Д., Лукьянов И. А. Свидетельство на программу Российской Федерации № 2011614788 Система криминалистического анализа текстов «Crime TGS v.1.0» (CATGS) / С. Д. Кулик, И. А. Лукьянов (Россия). Заявка № 2011613023; Заяв. 28.04.2011; Зарегистр. 17.06.2011 (Роспатент).

