

*А.А.Копыльцов  
А.В.Копыльцов  
Санкт-Петербург, Россия*

*A.A.Kopyltsov  
A.V.Kopyltsov  
Saint-Petersburg, Russia*

## ОБРАБОТКА СЛАБО ФОРМАЛИЗОВАННОЙ ИНФОРМАЦИИ, ПОСТУПАЮЩЕЙ ОТ ТЕХНИЧЕСКИХ СИСТЕМ

## PROCESSING OF POORLY FORMALIZED INFORMATION FROM TECHNICAL SYSTEMS

**Аннотация.** Разработана совокупность моделей и алгоритмов обработки поступающей от технических систем информации, позволяющая осуществлять поддержку принятия решений в слабо формализованных областях.

**Ключевые слова:** информация; обработка информации; технические системы.

**Сведения об авторах:** Копыльцов Антон Александрович<sup>1</sup>, инженер-программист кафедры автоматизированных систем обработки информации и управления; Копыльцов Александр Васильевич<sup>2</sup>, доктор технических наук, профессор, заведующий кафедрой информатики.

**Место работы:** <sup>1</sup> Санкт-Петербургский государственный электротехнический университет (ЛЭТИ) им.В.И.Ульянова (Ленина); <sup>2</sup> Российский государственный педагогический университет им.А.И.Герцена.

**Контактная информация:** 191186, Санкт-Петербург, наб. р. Мойки, д. 48, корпус 1, ауд. 224; тел. (812) 3128624. E-mail: <sup>1</sup> antonkop@gmail.ru; <sup>2</sup> kopyl2001@mail.ru

**Abstract.** The present paper is devoted to the development of a set of models and algorithms for processing information from technical systems which allows carrying out support of decision-making in poorly formalized areas.

**Key words:** information; processing information; technical systems.

**About the authors:** Anton Alexandrovich Kopyltsov<sup>1</sup>, programming engineer of the Department of Automated Information Processing Systems; Alexander Vasilievich Kopyltsov<sup>2</sup>, Doctor of Engineering, Professor, Head of the Department of Computer Sciences.

**Place of employment:** <sup>1</sup> Saint-Petersburg State Electrotechnical University «LETI»; <sup>2</sup> The Herzen State Pedagogical University of Russia.

В настоящее время в разных областях человеческой деятельности используется большое количество технических систем. Такие системы осуществляют сбор различной информации (видео- и аудиоинформации, информации о форме и составе различных объектов наблюдения, их динамике и др.). Информация, поступающая от технических систем, разнообразна как по виду (дискретная, аналоговая), так и по частоте (одноразовая, периодическая, непрерывная), по длительности (кратковременная, продолжительная, постоянная), по размерности (одномерная, многомерная). Информация, которую снимают с приборов, часто слабо формализованная, нечеткая и содержит помехи, шумы, искажения. Обработка такой информации с целью подготовки к принятию решения осуществляется различными способами, в частности, имеются детерминированные и вероятностные методы извлечения информации, экспертные системы и системы поддержки принятия решений на основе искусственного интеллекта. Каждая из перечисленных систем имеет свои преимущества и недостатки. Однако такие системы, которые позволяли бы это делать удовлетворительно, до сих пор не созданы. Сложность решения этой проблемы состоит в том, что объем слабо формализованной информации, поступающей от технических систем, с одной стороны, существенно возрастает, а с другой стороны, интервал времени, в течение которого нужно принять правильное решение, резко уменьшается. Таким образом, разработка моделей и алгоритмов обработки слабо формализованной информации, поступающей от технических систем с целью поддержки принятия решений актуальна и своевременна.

Построена система моделей и алгоритмов обработки слабо формализованной информации, поступающей от технических систем, используемых в промышленности и социальной сфере и позволяющих осуществлять поддержку принятия решений. Эта система включает, в частности, модели классификации поступающей информации, свертки, оценивания достоверности и безопасности информации, оценивания вероятности, с которой

можно доверять вновь полученной информации, поддержки принятия решений в каждом из классов и сбора решений по всем классам и поддержки принятия решений, выработки рекомендаций.

Система моделей обработки слабо формализованной информации состоит из 14 основных модулей: 1. Информация; 2. Распознавание; 3. Классификация; 4 (4-1, ..., 4-i, ..., 4-n). Свертка; 5 (5-1, ..., 5-i, ..., 5-n). Оценивание достоверности полученной информации; 6 (6-1, ..., 6-i, ..., 6-n). Оценивание безопасности полученной информации; 7 (7-1, ..., 7-i, ..., 7-n). Установление связей вновь полученной информации с ранее полученной; 8 (8-1, ..., 8-i, ..., 8-n). Оценивание вероятности, с которой можно доверять вновь полученной информации; 9 (9-1, ..., 9-i, ..., 9-n). Поддержка принятия решения (по каждому классу), 10. Поддержка принятия решения (обобщенная); 11. Число подтверждений (связей), подтверждающих правильность принятого решения; 12. Выработка устойчивой реакции на поступающую информацию и ее закрепление; 13. Руководство к действию; 14. Опыт, память (база данных).

Предложены алгоритмы сбора, распознавания, классификации информации, поступающей от технических систем, а также модели оценивания достоверности и безопасности информации, установления связей вновь полученной информации с ранее полученной и оценивание вероятности, с которой можно доверять вновь полученной информации.

Последовательность выполнения алгоритма включает следующие этапы:

1. Информация собирается в модуле «1. Информация». Информация может быть весьма разнообразной. У человека имеется пять основных источников получения информации — зрение, слух, обоняние, вкус, осязание. Каждый из органов чувств имеет свою чувствительность, т.е. определенный диапазон, в пределах которого возможно получение информации. Что касается технических систем, то в некоторых случаях диапазон существенно шире, чем у человека, например, электромагнитное излучение (свет, гамма-излучение) или звук (инфразвук, ультразвук). Кроме того, разработаны технические системы для восприятия такой информации, которую человек не может непосредственно воспринимать, например, радиация (альфа-, бета- и гамма-излучения).

2. Информация поступает из модуля «1. Информация» в модуль «2. Распознавание», где осуществляется ее распознавание, т.е. разделение на аудио- и видеоинформацию, тактильную, вкусовую и другие виды информации. К особенностям информации при ее распознавании можно отнести то, что часто информация о наблюдаемом объекте не одного вида, а нескольких видов, т.е., например, объект можно наблюдать в видео- и аудиодиапазоне. В модуле «2. Распознавание» определяются виды информации, в которых проявляет себя объект наблюдения.

3. Из модуля «2. Распознавание» информация поступает в модуль «3. Классификация», где осуществляется классификация поступающей информации на классы  $K_1, \dots, K_i, \dots, K_n$ . Каждый из классов включает либо один вид информации (зрение, слух, обоняние, вкус, осязание), либо их какую-нибудь комбинацию по два, по три и т.д. Количество видов информации при желании может быть увеличено. Параметры (количество классов и их разнообразие) в модуле «3. Классификация» изменяются в процессе работы (наблюдения за каким-либо объектом), т.е. это самообучающаяся система. На начальном этапе работы можно использовать либо обучающую выборку, либо случайным образом сгенерированное распределение поступающей информации между классами.

4. Информация, поступающая в каждый из классов, подвергается своей, присущей данному классу, обработке по определенному алгоритму (свертке). В итоге получается новая информация, обработанная специальным образом, присущим данному классу. В основу алгоритма свертки положен экспертный метод, разработанный ранее для оценки качества программных продуктов при недостатке информации. В алгоритме оценивания качества съема информации с объекта наблюдения используется понятие свертки, суть которой

состоит в следующем. Пусть имеется объект наблюдения, с которого осуществляется съем информации нескольких типов (например, зрение, слух и др.). Пусть  $N$  — достаточно большое целое число. Можно взять  $N = 20$  или больше. Для оценивания качества съема информации с объекта наблюдения предполагаются выполненными несколько условий — условие дискретности, условие приоритета, условие нормировки. Условие дискретности можно записать в следующем виде. Пусть

$$f_i = \sum_{j=1}^{m_i} P_j K_j$$

— формула для расчета качества съема информации с объекта наблюдения, где

$$P_1 = 0, P_2 = 1/N, P_3 = 2/N, \dots, P_{N+1} = N/N=1$$

— весовые коэффициенты,  $K_j$  ( $j = 1, \dots, m$ ) — числовые значения используемых типов съема информации (показатели),  $m$  — число показателей, т.е. используемых типов съема информации, определяющих  $f_i$ . Условие приоритета означает, что вводятся отношения порядка между показателями (такой же или чуть-чуть важнее ( $\geq$ ), важнее ( $>$ ), значительно важнее ( $\gg$ )). А именно, если показатель  $K_a$  такой же или немного важнее показателя  $K_b$ , то  $P_a \geq P_b$ , если показатель  $K_p$  важнее показателя  $K_q$ , то  $P_p > P_q$ , если показатель  $K_p$  значительно важнее показателя  $K_q$ , то  $P_p \gg P_q$ , т.е. между  $P_p$  и  $P_q$  находится по крайней мере одно некоторое  $P_r$ ,  $P_p > P_r > P_q$ . Предполагается выполненным условие нормировки

$$\sum_{j=1}^{m_i} P_j = 1.$$

Поскольку набор весовых коэффициентов  $P_1, P_2, P_3, \dots, P_{N+1}$ , удовлетворяющих условиям дискретности, приоритета и нормировки, не единственный, то получаем несколько значений  $f_i$ . Можно определить их среднее значение

$$\bar{f} = \left( \sum_{j=1}^{N_0} f_j \right) / N_0,$$

дисперсию

$$D = \left[ \sum_{j=1}^{N_0} (f_j - \bar{f})^2 \right] / N_0$$

и среднеквадратичное отклонение  $\sigma = D^{0.5}$ , где  $N_0$  — число полученных  $f_j$ . Процедура получения  $\bar{f}$  называется сверткой. Применяя свертку к показателям (числовым значениям используемых типов съема информации), можно оценить качество съема информации с объекта наблюдения.

5. После свертки в модулях «4(4-1, ..., 4-i, ..., 4-n). Свертка» осуществляется оценивание достоверности информации в модулях «5(5-1, ..., 5-i, ..., 5-n). Оценивание достоверности полученной информации» в каждом классе путем сравнения ее с ранее полученной информацией (предполагается, что есть хранилище (модуль «14. Опыт, память (база данных)») ранее полученной информации). Если информация недостоверная (достоверность ее ниже некоторого, заранее заданного для каждого класса уровня), то осуществляется переход в пункт 1, где информация подвергается повторному распознаванию, классификации и помещается в другой класс.

6. В модулях «6(6-1, ..., 6-i, ..., 6-n). Оценивание безопасности полученной информации» осуществляется проверка безопасности информации в каждом классе, путем сравнения ее с ранее полученной информацией (предполагается, что есть хранилище (модуль «14. Опыт, память (база данных)») ранее полученной информации). Если информация

представляет собой опасность (уровень опасности ее выше некоторого, заранее заданного для каждого класса уровня), то осуществляется переход в пункт 1, где информация подвергается повторному распознаванию, классификации и помещается в другой класс. Если же и при повторной классификации информация представляет собой опасность, то выдается предупреждение «Информация опасная», и далее управление системой осуществляется в ручном режиме, т.е. с участием человека.

7. Установление связей между вновь полученной информацией в каждом классе и ранее полученной информацией, находящейся в хранилище (модуль «14. Опыт, память (база данных)»), осуществляется в модулях «7(7-1, ..., 7-i, ..., 7-n). Установление связей вновь полученной информации с ранее полученной». Если количество связей меньше некоторой, заранее заданной величины, специфичной для каждого класса, то осуществляется переход в пункт 1, где информация подвергается повторному распознаванию, классификации и помещается в другой класс.

8. Оценивание вероятности, с которой можно доверять полученной информации в каждом классе, осуществляется в модулях «8(8-1, ..., 8-i, ..., 8-n). Оценивание вероятности, с которой можно доверять вновь полученной информации». Если вероятность меньше некоторой, заранее заданной величины, специфичной для каждого класса, то осуществляется переход в пункт 1, где информация подвергается повторному распознаванию, классификации и помещается в другой класс.

9. Поддержка принятия решения в каждом классе осуществляется в модулях «9 (9-1, ..., 9-i, ..., 9-n). Поддержка принятия решения (по каждому классу)», т.е. генерируются рекомендации для каждого класса (что и когда нужно делать, если получена такая информация).

10. Сбор сгенерированных решений (что и когда нужно делать) из всех классов осуществляется в модуле «10. Поддержка принятия решения (обобщенная)». В этом модуле осуществляется анализ сгенерированных решений в классах и генерация на их основе новой совокупности решений в поддержку принятия решения. Вывод рекомендаций (что и когда нужно делать). Окончательное решение (выбрать какое-то решение из рекомендованных или принять свое собственное решение) принимает человек.

11. В модуле «11. Число подтверждений (связей), подтверждающих правильность принятого решения» осуществляется сравнение принятого решения с решениями, принятыми ранее на основе сравнения с информацией, хранящейся в модуле «14. Опыт, память (база данных)». Если число подтверждений меньше некоторого, заранее заданного числа, то осуществляется переход в пункт 1, где информация подвергается повторному распознаванию, классификации и помещается в другой класс.

12. В модуле «12. Выработка устойчивой реакции на поступающую информацию и ее закрепление» осуществляется выработка устойчивой реакции на многократно поступающую повторяющуюся информацию и ее закрепление путем сравнения полученной информации с информацией, хранящейся в модуле «14. Опыт, память (база данных)». Если вновь полученная и проверенная информация отсутствует в модуле «14. Опыт, память (база данных)», то она в этот модуль записывается.

13. В модуле «13. Руководство к действию» осуществляется путем сравнения с модулем «14. Опыт, память (база данных)» выработка рекомендаций по действиям, которые нужно предпринять, и генерируются рекомендации к действиям.

14. Модуль «14. Опыт, память (база данных)» включает в себя хранилище информации, поступившей ранее. В модуле «14. Опыт, память (база данных)» на основе вновь записанной информации в этот модуль и путем сравнения с ранее записанной информацией генерируется новая информация (гипотезы, идеи), которая поступает в модуль «1. Информация».

Таким образом, построен алгоритм обработки слабо формализованной информации, поступающей от технических систем. Он позволяет перерабатывать информацию путем классификации полученной информации, оценивания ее достоверности и безопасности, установления связей вновь полученной информации с ранее полученной информацией таким образом, чтобы результат обработки информации сохранялся для его использования при последующей обработке информации с целью поддержки принятия решений. Применяя этот алгоритм для обработки информации, получаем на выходе рекомендации по поддержке принятия решений. Окончательное решение принимает человек.

#### ЛИТЕРАТУРА

1. Воробьев В.И., Копыльцов А.В., Пальчун Б.П., Юсупов Р.М. Методы и модели оценивания качества программного обеспечения. СПб., 1992.
2. Копыльцов А.А. Нечитайленко Р.А. Кластерное атрибутирование объектов информационной обработки по понятийным частным и интегральным признакам // XII Санкт-Петербургская международная конференция «Региональная информатика — 2010». Материалы конференции. (СПб., 20—22 окт. 2010 г.). СПб, 2010.
3. Копыльцов А.В. Об оценке качества программных продуктов // Проблемы информатизации (теоретический и научно-практический журнал). 1994. Вып. 3—4.
4. Хованов Н.В. Статистические модели теории квалиметрических шкал. Л., 1986.