

УДК 004.056

Е.А. Родин

## Схема адаптивной системы фильтрации сообщений на основе метода вывода по прецедентам

В автоматизированных системах обучения возникает проблема получения строгих функциональных зависимостей между входными и выходными параметрами, связанная со слабоформализуемостью объекта управления. В данной работе рассматривается подход, основанный на методе вывода по прецедентам, для создания адаптивной системы фильтрации сообщений по различным классификационным признакам.

**Ключевые слова:** система фильтрации сообщений, метод вывода по прецедентам, прецедент, адаптивная система обучения.

Исследование процессов и методов организации адаптивного информационного взаимодействия в процессе принятия решения оперативных задач в области управления на современном этапе технологического развития приобретает все более актуальное значение.

Интерес к моделям и методам рассуждений на основе прецедентов обусловлен тем, что на сегодняшний момент актуальна проблема моделирования человеческих рассуждений для интеллектуальных систем поддержки принятия решений. Понимание широких возможностей подхода, основанного на прецедентах, привело к созданию и быстрому развитию «теории прецедентов», известной как Case-Based Responding (CBR) – метод рассуждений на основе прецедентов [1–3].

Актуальность данного направления обусловлена слабой аксиоматикой вероятностного подхода с точки зрения практической деятельности: известно, что при малых объемах данных асимптотические методы теории вероятностей и математической статистики могут приводить к неточным или даже ошибочным выводам [4, 5].

К наиболее важным компонентам системы фильтрации относится система адаптивного управления (система обучения), для которой требуется разработка количественных моделей, адекватно отражающих организацию процесса обучения.

Среди известных механизмов управления, исторически появившихся в автоматизированных системах обучения и получивших развитие в современных разработках, можно выделить следующие [6–10]:

- программированное обучение;
- управление на основе стереотипов;
- динамическая разработка стратегий обучения;
- управление на базе экспертных систем и байесовского вывода;
- обучение как управление сложным объектом;
- управление с использованием механизмов нечеткой логики;
- управление на основе взаимодействия интеллектуальных агентов.

Вывод, основанный на прецедентах, представляет собой метод построения экспертных систем, которые делают заключения относительно данной проблемы или ситуации по результатам поиска аналогий, хранящихся в базе прецедентов.

Вывод по прецедентам особенно эффективен, когда [7]:

- основным источником знаний о задаче является опыт, а не теория;
- решения не уникальны для конкретной ситуации и могут быть использованы в других случаях;
- целью является не гарантированное верное решение, а лучшее из возможных решений.

Как правило, методы рассуждения на основе прецедентов включают в себя четыре основных этапа, образующие так называемый CBR-цикл [11], структура которого представлена на рис. 1.

Основными этапами CBR-цикла являются:

- извлечение наиболее соответствующего (подобного) прецедента (прецедентов) для сложившейся ситуации из библиотеки прецедентов (БП);
- повторное использование извлеченного прецедента для попытки решения текущей проблемы;
- пересмотр и адаптация в случае необходимости полученного решения в соответствии с текущей проблемой;
- сохранение вновь принятого решения как части нового прецедента.

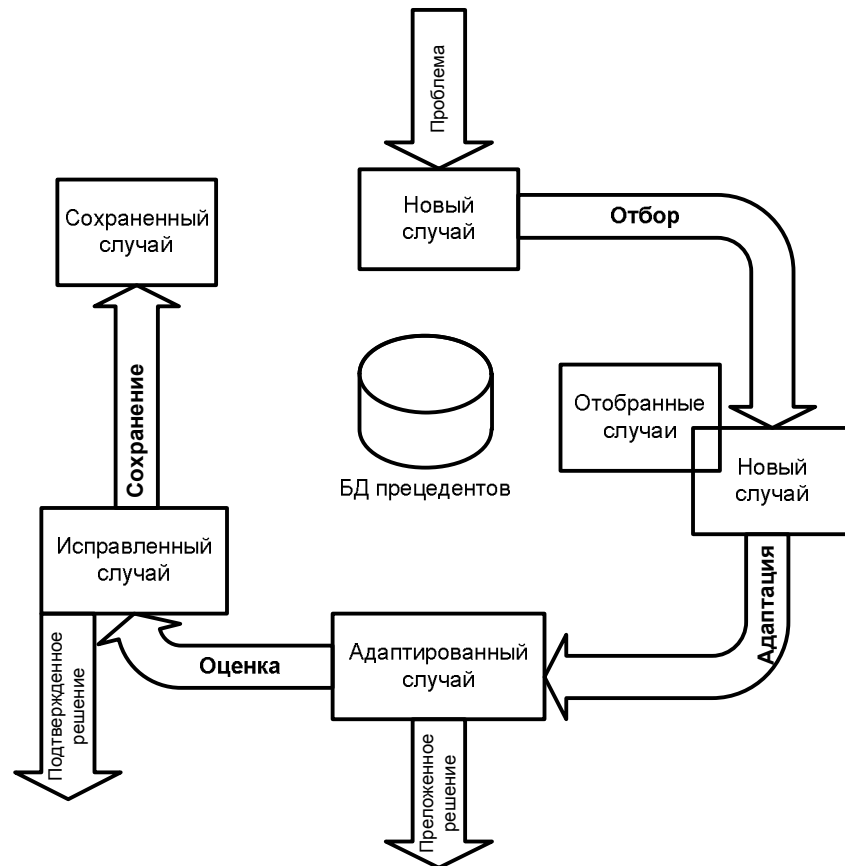


Рис. 1. Цикл вывода на основе прецедентов

Основной целью использования аппарата прецедентов в системах поддержки принятия решения является выдача готового решения для текущей ситуации на основе прецедентов, которые уже имели место в прошлом при управлении данным объектом или процессом.

Проблема выбора подходящего прецедента является одной из самых важных в таких системах. Естественно искать подходящий прецедент в той области пространства поиска, где находятся решения сходных проблем, иначе говоря, поиск должен быть организован сообразно цели.

Наиболее популярным и часто используемым является метод «ближайшего соседа» [12]. В его основе лежит тот или иной способ измерения степени близости прецедента и текущего случая по каждому признаку (будь это текстовый, числовой или булевский), который пользователь сочтет полезным для достижения цели. Вводится метрика на пространстве всех признаков, в этом пространстве определяется точка, соответствующая текущему случаю, и в рамках этой метрики находится ближайшая к ней точка из точек, представляющих прецеденты. Обычно прогноз делается на основе нескольких ближайших точек, а не одной. Такой метод более устойчив, поскольку позволяет сгладить отдельные выбросы, случайный шум, всегда присутствующий в данных.

Каждому признаку назначают вес, учитывающий его относительную ценность. Полностью степень близости прецедента по всем признакам можно вычислить, используя обобщенную формулу вида

$$y = \frac{\sum_{j=1}^n w_j * \text{sim}(x_{ij}, x_{kj})}{\sum_{j=1}^n w_j}, \quad (1)$$

где  $w_j$  – вес  $j$ -го признака;  $\text{sim}$  – функция подобия (метрика);  $x_{ij}$  и  $x_{kj}$  – значения признака  $x_j$  для текущего случая  $i$  и прецедента соответственно.

Метод может быть реализован очень эффективно, правда, требует для работы большой памяти, так как в процессе нахождения значения зависимой переменной для новой записи используется вся существующая база данных.

Выбор метрики (или меры близости) считается узловым моментом, от которого решающим образом зависит поиск подходящих прецедентов. В каждой конкретной задаче этот выбор производится по-своему, с учетом главных целей исследования, физической и статистической природы используемой информации и т.п. В некоторых методах выбор метрики достигается с помощью специальных алгоритмов преобразования исходного пространства признаков.

Основные метрики, используемые при выборе прецедентов:

- эвклидово расстояние;
- манхэттенская метрика;
- мера сходства Хэмминга;
- мера сходства Роджерса–Танимото;
- расстояние Махолонобиса;
- расстояние Журавлева.

Формально прецедент можно описать следующим образом: пусть имеются множество объектов  $X$ , множество ответов  $Y$  и существует целевая функция  $y^* : X \rightarrow Y$ , значения которой  $y_i = y^*(x_i)$  известны только на конечном подмножестве объектов  $\{x_1, \dots, x_l\} \in X$ . Пары «объект–ответ»  $(x_i, y_i)$  называются прецедентами.

Имеется множество способов представления прецедента: от записей в базах данных, древовидных структур до предикатов и фреймов. Конкретное выбранное представление прецедентов должно соответствовать общим целям системы. Проблема представления прецедента – прежде всего проблема выбора информации, которую надо включать в описание прецедентов, нахождение соответствующей структуры для описания содержания прецедента, а также определения, каким образом должна быть организована и индексирована база знаний прецедентов для эффективного поиска и многократного использования.

Согласно [13] прецедент включает:

- состояние объекта управления (ОУ) до воздействия;
- состояние управляющего воздействия;
- состояние ОУ после воздействия;
- исход и его оценку.

Состояние ОУ в прецеденте на абстрактном уровне представим следующим образом:

$$Y(t) = (S, F_{Zn}(t-1)), \quad (2)$$

где  $S$  – стереотип, построенный на множестве классификационных признаков  $Z$ ;  $F_{Zn}$  – последний изученный фрагмент знаний о предметной области;  $(t, t-1)$  – текущий и предыдущий момент времени соответственно.

Состояние среды определяется как

$$X = \{P_k\}, \quad (3)$$

где  $P_k$  – множество параметров системы, которые позволяют адаптировать управляющие воздействия.

Управляющие воздействия определяются как:

$$U(t) = P_k(t), \quad (4)$$

где  $P_k(t)$  – назначаемые параметры в конкретный момент времени.

В ситуации, когда известных параметров ОУ и окружающей среды недостаточно для полного и однозначного определения его поведения, нельзя принимать решение об управляющем воздействии на объект, зная только его входные параметры. Необходимо знать поведение объекта, т.е. осуществлять управление не по его параметрам, а по его состояниям. Если удастся сформировать на основе априорной информации обобщенные, или агрегированные образы – классы состояний ОУ с известной реакцией объекта каждого класса, то управляющее воздействие можно рассматривать как отображение ОУ из класса в класс.

Количество классов определяется количеством контролируемых модулей обработки. В соответствии с общей концепцией построения системы приема и обработки материалов можно выделить ряд классов состояний (рис. 2).

Учитывая вышесказанное, исходом является функция

$$I(t) = IP(Q_i), \quad (5)$$

где  $Q_i$  – класс конечного состояния сообщения в тракте обработки;  $IP(Q_i)$  – функция оценки качества обучения.

Таким образом, прецедент представляется как:

$$PR(t) = (Y(t), U(t), Y(t+1), I(t)). \quad (6)$$

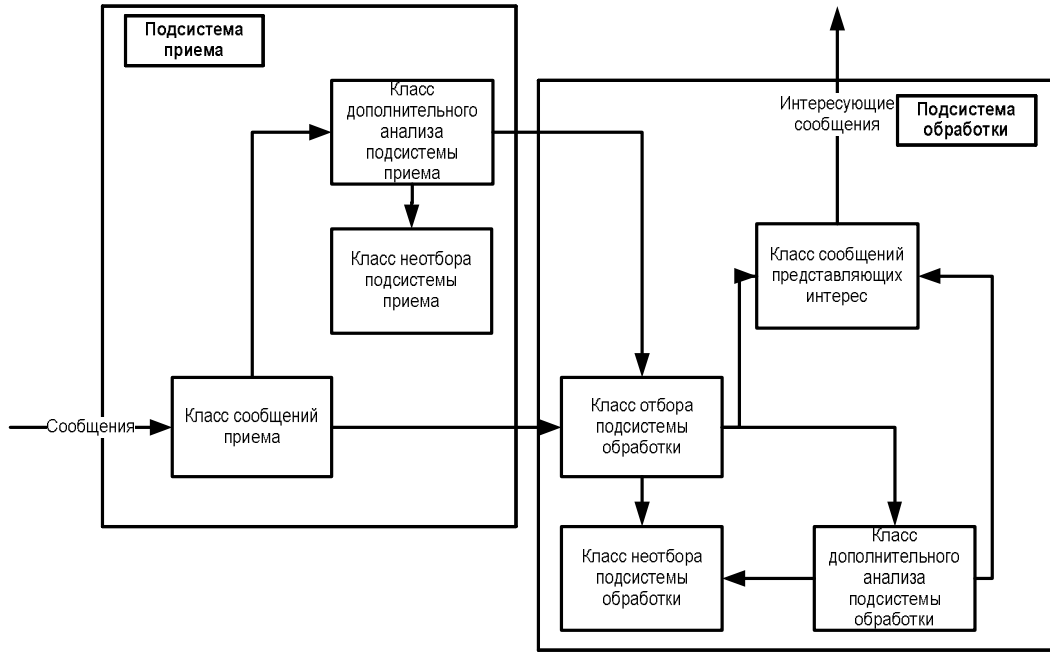


Рис. 2. Классы состояний системы фильтрации сообщений

Обобщенная схема адаптивной системы фильтрации сообщений представлена на рис. 3.

Критерием перестроения системы фильтрации является показатель эффективности выполнения поставленных задач

$$I_{\text{тек}} \rightarrow \max, \tag{7}$$

где  $I_{\text{тек}}$  – количество сообщений, представляющих определенный интерес в момент времени  $t$ , которое должно быть максимизировано с учетом пропускной способности системы фильтрации и множества классификационных параметров.

Схема применения метода вывода по прецедентам для организации модуля анализа и управления представлена на рис. 4.

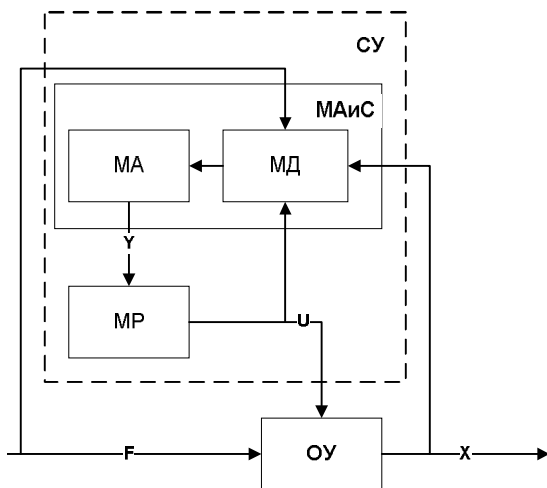


Рис. 3. Схема адаптивной системы фильтрации сообщений, где СУ – система управления; МАиС – модуль адаптации и самонастройки; МД – модуль детектирования (снятия параметров состояния); МА – модуль анализа и управления; МР – модуль регулирования (регулятор); ОУ – объект управления

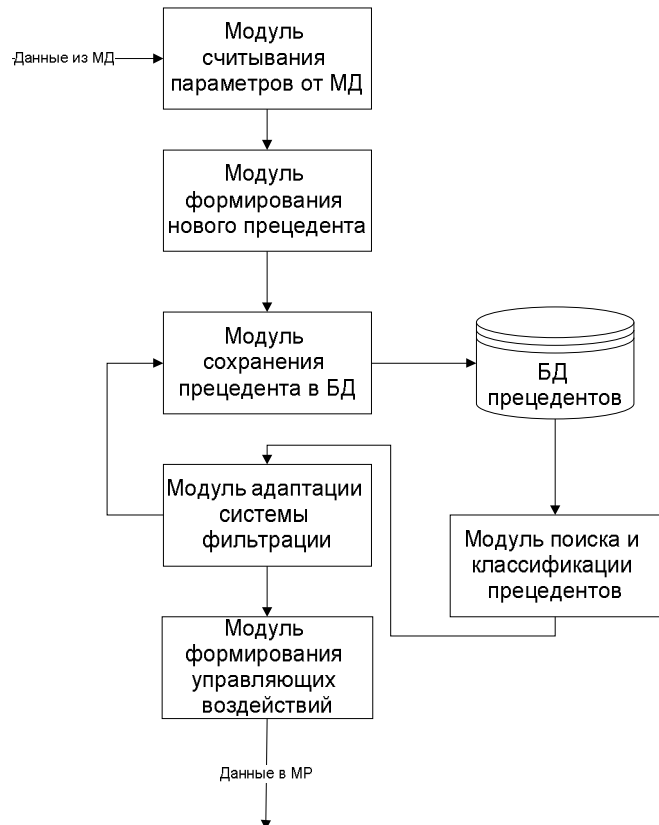


Рис. 4. Схема применения метода вывода по прецедентам

Технология вывода по прецедентам представляет собой основу для приложений, которые наращиваются за счет постоянного сбора информации, что дает возможность ее применения для создания различных адаптивных систем обучения и, в частности, системы фильтрации сообщений.

Кроме того, обратная связь, возникающая при сохранении решений для новых проблем, означает, что вывод по прецедентам по своей сути является «самообучающейся» технологией, благодаря чему рабочие характеристики каждой базы прецедентов с течением времени и накоплением опыта непрерывно улучшаются.

База данных прецедентов является по своей сути хранилищем данных (ХД), которое обеспечивает, при необходимости, возможность повторного использования прецедентов с целью формирования «оптимального» признакового пространства системы фильтрации, т.е. всегда существует возможность перестроения системы фильтрации с учетом состояния ОУ и требований, предъявляемых системой более высокого уровня с учетом предыдущего опыта – прецедентов.

#### *Литература*

1. Shank R. Dynamic Memory: A Theory of Learning in Computers and People. – New York: Cambridge University Press, 1982. – 247 p.
2. Фишберн П. Теория полезности для принятия решений. – М.: Наука 1978. – 352 с.
3. Smyth B. Adaptation-Guided Retrieval: Questioning the Similarity Assumption in Reasoning / B. Smyth, M.T. Kaene. // Artificial Intelligence. – 1998. – № 102. – P. 249–293.
4. Орлов А.И. Эконометрика: учебник для вузов. – М.: Экзамен, 2003. – 576 с.
5. Орлов А.И. Нечисловая статистика. – М.: МЗ-Пресс, 2004. – 436 с.
6. Цибульский Г.М. Модели обучения автоматизированных обучающих систем / Г.М. Цибульский, Е.И. Герасимова, В.В. Ерошин // Системотехника: Сетевой электронный научный журнал. – 2004. – № 2. – С. 153–162.
7. Петрушин В.А. Экспертно-обучающиеся системы. – Киев: Наукова думка, 1992. – 356 с.
8. Буль Е.Е. Обзор моделей студента для компьютерных систем обучения // Educational Technology & Society. – 2003. – № 6. – С. 82–98.
9. Пустынникова И.Н. Технология использования экспертных систем для диагностики знаний и умений // Educational Technology & Society. – 2001. – № 4. – С. 147–156.
10. Астанин С.В. Сопровождение процесса обучения на основе нечеткого моделирования // Открытое образование. – 2000. – № 5. – С. 37–44.
11. Варшавский П.Р. Поиск решения на основе структурной аналогии для интеллектуальных систем поддержки принятия решений / П.Р. Варшавский, А.П. Еремеев // Известия РАН. Теория и системы управления. – 2005. – № 1. – С. 97–109.
12. Anand S.S. Utilising Censored Neighbours in Prognostication / S.S. Anand, J.G. Hughes, D.A. Bell. – Denmark: Workshop on Prognostic Models in Medicine, 1999. – P. 15–20.
13. A Review of Industrial Case-Based Reasoning Tools / K.D. Althof, M. Manago, E. Auriol, R. Barlette. – Oxford: AI Intelligence, 1995. – 145 p.

---

#### **Родин Евгений Александрович**

Инженер ФГУП «НИИ Квант», г. Москва

Эл. почта: earodin@rambler.ru

Тел.: 8-903-565-58-20

Rodin E.A.

#### **An adaptive system scheme for messages filtering based on the method of precedent decision making**

In the automated training systems there is a problem of making rigorous functional dependences between input and output parameters, connected with a low formalizability of an object to be controlled. In the given work, an approach, which is based on the method of precedent decision making for creation of adaptive system filtering messages by different classification signatures, is considered.

**Keywords:** system of messages filtering, method of precedent decision making, precedent, adaptive training system.