

## ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ПОИСКА ИНФОРМАЦИИ ПУТЕМ РЕШЕНИЯ ЗАДАЧИ КЛАССИФИКАЦИИ

В последние годы наблюдается бурный процесс информатизации общества. Информационные технологии активно внедряются в различные сферы деятельности человека, позволяя значительно сократить трудоемкость работы с потоками данных и автоматизировать обработку информации. Кроме того, с развитием сетевых технологий и особенно сети *Internet* намечается акцент в пользу глобальных информационных систем, позволяющих в рамках какой-либо задачи объединить большое количество заинтересованных лиц по всему миру.

Важнейшей функцией подобных систем является оперативный поиск и предоставление информации пользователям. Однако зачастую пользователь не способен точно определить множественные критерии поиска, позволяющие обеспечить достаточную адекватность найденной информации. Одним из вариантов решения данной проблемы является динамическая классификация хранимой информации.

Примером подобной системы может служить разработанное на кафедре «Конструирование радиоэлектронных и микропроцессорных систем» Тамбовского государственного технического университета динамическое Web-приложение, включающее в себя базу данных электронных учебно-методических пособий [1]. В данном приложении задача классификации учебных пособий осуществляется при помощи функции расстояния, основанной на взвешенном Евклидовом расстоянии, для чего были выработаны эвристические правила, а также разработано алгоритмическое и программное обеспечение, которое входит в состав Web-приложения.

Рассмотрим постановку задачи классификации пособий.

Пусть множество  $O = \{O_1, O_2, \dots, O_n\}$  обозначает множество пособий из некоторой совокупности. Для множества  $O$  имеется множество векторов характеристик  $X = \{X_1, X_2, \dots, X_n\}$ , которые описывают множество  $O$ .

Объекту  $O_1$  соответствует вектор измерения  $X_1$ , объекту  $O_2$  – вектор  $X_2$  и т.д. Множество  $X$  может быть представлено как  $n$  точек в  $p$ -мерном евклидовом пространстве, где  $p$  – зависит от числа характеристик объекта.

Известны центры классов:

$$K_1 = \{k_1^1, k_2^1, \dots, k_p^1\};$$

$$K_2 = \{k_1^2, k_2^2, \dots, k_p^2\}, \dots, K_k = \{k_1^k, k_2^k, \dots, k_p^k\}.$$

Для решения об отнесении объекта  $O_i$  к какому-либо классу необходимо вычислить расстояния между данным объектом и всеми центрами обозначенных классов.

Взвешенное Евклидово расстояние

$$l_i^m = \sqrt{\sum_{j=1}^p w_j^m (x_j^i - k_j^m)^2},$$

где  $i$  – номер пособия;  $m$  – номер класса;  $j$  – индекс элемента вектора наблюдений;  $p$  – число характеристик пособия;  $w_j^m$  – вес, обозначающий важность  $j$ -го наблюдения для  $m$ -го класса [2].

Для решения задачи классификации определим исходные данные:

- центры классов и их число определены заранее;
- каждое пособие характеризуется фиксированным количеством свойств, оцениваемых в баллах.

Для каждой характеристики пособия экспертом назначаются веса, показывающие важность данного показателя при отнесении пособия к тому или иному классу.

При вынесении решения об отнесении пособия к какому-либо классу используется набор эвристических правил.

1. Пособие относится к тому классу, к центру которого оно расположено ближе в Евклидовом пространстве, т.е.

$$l_i^m = \min.$$

2. Определены граничные расстояния для каждого класса, которые позволяет отнести к ним пособия

$$l_i^m = \min \leq l_{гп}^m.$$

В том случае, когда все расстояния между пособием и центрами классов превышают обозначенные граничные расстояния, пособие не может быть отнесено ни к одному классу.

3. В случае, когда количество неклассифицированных пособий превышает некоторое определенное значение, определяется центр их дислокации. Если размер данной дислокации находится в пределах, обозначенных экспертом, выносится решение об определении нового класса.

4. Определены характеристики пособия, оказывающие влияние друг на друга. Например, «Применение новых информационных технологий»  $x_i$  будет зависеть от параметра «Год издания пособия»  $x_j$ , т.е.

$$x_i = f(x_j).$$

Вид зависимости определяется экспертом в каждом конкретном случае.

Использование данного алгоритма в рассматриваемой системе позволило значительно повысить эффективность и адекватность поиска информации об учебно-методических пособиях.

#### СПИСОК ЛИТЕРАТУРЫ

1. Осипов, Е.В. Система хранения информации с оперативным доступом // Математические методы в технике и технологиях – ММТТ-19 : сб. тр. XIX Междунар. науч. конф. : в 10 т. / Е.В. Осипов ; под общ. ред. В.С. Балакирева. – Воронеж : Воронеж. гос. технол. акад., 2006. – Т. 10. – Секция 11. – 250 с.

2. Айвазян, С.А. Прикладная статистика. Основы эконометрики. Т. 1. Основы эконометрики : учебник для вузов. – 2-е изд. / С.А. Айвазян. – М. : ЮНИТИ-ДАНА, 2003. – 432 с.

*Кафедра «Проектирование радиоэлектронных и микропроцессорных систем»*