

Е.В. ВОЛЧЕНКО, канд. техн. наук, доц., Институт информатики и искусственного интеллекта ГВУЗ "Донецкий национальный технический университет", Донецк

КЛАССИФИКАЦИЯ ОБЪЕКТОВ В АДАПТИВНЫХ СИСТЕМАХ РАСПОЗНАВАНИЯ НА ОСНОВЕ ФУНКЦИИ ВЗВЕШЕННОГО КОНКУРЕНТНОГО СХОДСТВА

В работе предложено естественное расширение области использования функции конкурентного сходства на взвешенные обучающие выборки w -объектов в адаптивных системах распознавания. Описан принцип классификации объектов методом k -ближайших соседей на основе функции взвешенного конкурентного сходства ($wFRiS$ -функции). Приведены результаты экспериментальных исследований, подтвердившие эффективность предложенного подхода. Библиогр.: 10 назв.

Ключевые слова: адаптивная система распознавания, w -объект, функция взвешенного конкурентного сходства, метод k -ближайших соседей.

Постановка проблемы и анализ литературы. Построение классификаторов в обучающихся системах распознавания невозможно без оценки степени близости классифицируемого объекта к каждому из классов системы с помощью выбранной меры [1, 2]. Используемые в системах распознавания меры близости условно можно разделить на два вида [3]:

- абсолютные, позволяющие находить конкретное численное значение – расстояние от классифицируемого объекта до объектов выбранного класса;
- относительные, позволяющие оценить, к какому из классов системы классифицируемый объект находится ближе, чем к другим классам.

При принятии решения о классификации предпочтительнее использование относительных мер близости, поскольку они позволяют непосредственно дать ответ на вопрос о классификации на основе анализа всех классов системы, выбрав ближайший из них. Одной из наиболее эффективных относительных мер близости в обучающихся системах распознавания на сегодняшний день является функция конкурентного сходства ($FRiS$ -функция), предложенная в [3, 4]. Существенными преимуществами данной меры являются: независимость от закона распределения объектов, одинаковые значения меры при анализе равноудаленных объектов, ограниченность диапазона $[-1; 1]$ и простая интерпретируемость получаемых значений. Величина

конкурентного сходства классифицируемого объекта X_s с ближайшим объектом X_1 первого класса в сравнении с ближайшим объектом X_2 второго класса вычисляется по формуле [4]:

$$F_{X_1/X_2}(X_s) = \frac{R(X_2, X_s) - R(X_1, X_s)}{R(X_2, X_s) + R(X_1, X_s)}, \quad (1)$$

где $R(X_a, X_b)$ – функция расстояния между объектами a и b .

Классифицируемый объект X_s относится к первому классу, если $F_{X_1/X_2}(X_s) \in [0, 1]$ и ко второму классу, если $F_{X_1/X_2}(X_s) \in [-1, 0)$.

Использование FRiS-функции в решающих правилах, например, в методе k -ближайших соседей [3], существенно ускоряет процесс классификации и, при этом, обеспечивает достаточно высокую эффективность классификации.

Отличительной особенностью адаптивных систем распознавания, рассматриваемых в данной работе, является возможность постоянного пополнения обучающих данных. Необходимость пополнения обучающих выборок в адаптивных системах распознавания вызвана в большинстве случаев изменениями в объектах (значениях их признаков), происходящих в процессе функционирования систем [5]. Это, в свою очередь, приводит к неограниченному росту обучающих выборок и необходимости постоянной корректировки решающих правил классификации. Именно поэтому для адаптивных систем распознавания предпочтительным является использование решающих правил, не требующих существенных временных затрат на их построение.

В предыдущих работах автора [6 – 8] для эффективного функционирования адаптивных систем распознавания была предложена, теоретически обоснована и подтверждена экспериментально идея перехода к взвешенным обучающим выборкам w -объектов. Каждый объект такой выборки кроме значений признаков описывается дополнительной характеристикой, названной весом w -объекта. В качестве веса предложено использовать:

– данные о топологических свойствах обучающей выборки (плотность объектов в некоторой области пространства признаков, расстояние между выбранными объектами и др.) [6, 7];

– показатели уверенности классификации w -объекта группой экспертов [8];

– значения, обусловленные особенностями рассматриваемых объектов, или априорно полученные от экспертов.

При этом определение классификации объектов на основе взвешенных обучающих выборок выполнялось с использованием абсолютной меры близости, что, как указано выше, может увеличивать трудоемкость классификации.

Целью данной работы является уменьшение трудоемкости классификации за счет расширения области применения функции конкурентного сходства на взвешенные обучающие выборки w -объектов в адаптивных системах распознавания.

Постановка задачи. Пусть имеется некоторая конечная взвешенная обучающая выборка w -объектов $X^W = \{X_1^W, X_2^W, \dots, X_k^W\}$. Каждый w -объект X_i^W этой выборки описывается системой признаков $\{x_{i1}, x_{i2}, \dots, x_{in}\} \in R^n$, т.е. представляется точкой в линейном пространстве признаков, и весом p_i – целым положительным числом, тогда $X_i^W = \{x_{i1}, x_{i2}, \dots, x_{in}, p_i\}$. Для каждого w -объекта известна его классификация $y_i \in V$, где $V = \{V_1, \dots, V_l\}$ – множество всех классов системы.

Имеется также некоторый объект $X_s = \{x_{s1}, x_{s2}, \dots, x_{sn}\}$, заданный только набором признаков (для единообразия присвоим ему вес равный единице, т.е. $p_s = 1$, тогда $X_s^W = \{x_{s1}, x_{s2}, \dots, x_{sn}, p_s\}$).

Необходимо выполнить классификацию объекта X_s^W на основе функции конкурентного сходства в конкурирующей взвешенной среде.

Функция взвешенного конкурирующего сходства. Для оценки близости между парой w -объектов X_i^W и X_j^W взвешенной обучающей выборки X^W в работе [9] была введена метрика

$$d_W(X_i^W, X_j^W) = \frac{R(X_i^W, X_j^W)}{p_i \cdot p_j} = \frac{\sqrt{\sum_{o=1}^n (x_{io} - x_{jo})^2}}{p_i \cdot p_j}, \quad (2)$$

и показано, что она обеспечивает корректную классификацию по взвешенной выборке w -объектов.

Можно показать, что данная метрика в отличие от FRiS-функции [10] удовлетворяет аксиомам тождества и симметричности, и, также как (1), не для всех значений признаков и весов w -объектов удовлетворяет аксиоме треугольника.

Результаты работы [9] и предыдущие рассуждения показывают, что (2) является метрикой во взвешенном пространстве признаков, поэтому выполним далее естественную замену метрики $R(X_a, X_b)$ в (1) на метрику (2) для построения функции конкурентного сходства на основе взвешенных обучающих выборок w -объектов.

Поскольку в (1) требуется определять расстояние между классифицируемым объектом X_s , вес которого $p_s = 1$ и некоторым w -объектом X_a^W , имеющим вес p_a , преобразуем (2) к следующему виду:

$$d_w(X_a^W, X_s) = \frac{R(X_a^W, X_s)}{p_a}. \quad (3)$$

Тогда функция конкурентного сходства на основе взвешенных обучающих выборок w -объектов будет иметь вид:

$$F_{X_1^W / X_2^W}(X_s) = \frac{\frac{R(X_2^W, X_s)}{p_2} - \frac{R(X_1^W, X_s)}{p_1}}{\frac{R(X_2^W, X_s)}{p_2} + \frac{R(X_1^W, X_s)}{p_1}}.$$

На основе проведенных рассуждений и выполненных преобразований получим функцию взвешенного конкурентного сходства (w FRiS-функцию) классифицируемого объекта X_s с ближайшим w -объектом X_1^W первого класса в сравнении с ближайшим w -объектом X_2^W второго класса:

$$F_{X_1^W / X_2^W}(X_s) = \frac{p_1 \cdot R(X_2^W, X_s) - p_2 \cdot R(X_1^W, X_s)}{p_1 \cdot R(X_2^W, X_s) + p_2 \cdot R(X_1^W, X_s)}. \quad (4)$$

Аналогично можно показать, что функция взвешенного конкурентного сходства (w FRiS-функцию) для классификации некоторого

w -объекта X_q^W с ближайшим w -объектом X_1^W первого класса в сравнении с ближайшим w -объектом X_2^W второго класса будет иметь вид:

$$F_{X_1^W / X_2^W} (X_q^W) = \frac{p_1 \cdot R(X_2^W, X_q^W) - p_2 \cdot R(X_1^W, X_q^W)}{p_1 \cdot R(X_2^W, X_q^W) + p_2 \cdot R(X_1^W, X_q^W)}. \quad (5)$$

Анализируя (5), сформулируем следующие свойства wFRiS-функций, основанные на свойствах, приведенных в [3, 10].

Свойство 1. Результат классификации w -объекта X_q^W не зависит от веса p_q этого взвешенного объекта. Таким образом, подтверждается принцип относительности классификации, являющийся основой функции конкурирующего сходства, поскольку результат классификации в конкурирующей взвешенной среде зависит от взаимного расположения w -объектов и весов только конкурирующих w -объектов.

Свойство 2. Функция взвешенного конкурентного сходства принимает значения из диапазона $[-1; 1]$, т.е. $F_{X_1^W / X_2^W} (X_q^W) \in [-1; 1]$.

Покажем, что

$$\frac{p_1 \cdot R(X_2^W, X_q^W) - p_2 \cdot R(X_1^W, X_q^W)}{p_1 \cdot R(X_2^W, X_q^W) + p_2 \cdot R(X_1^W, X_q^W)} \geq -1$$

или

$$p_1 R(X_2^W, X_q^W) - p_2 R(X_1^W, X_q^W) \geq -p_1 R(X_2^W, X_q^W) - p_2 R(X_1^W, X_q^W).$$

В результате получим $2p_1 R(X_2^W, X_q^W) \geq 0$, что верно для любых X_q^W и X_2^W согласно постановке задачи.

Аналогично можно показать, что

$$\frac{p_1 \cdot R(X_2^W, X_q^W) - p_2 \cdot R(X_1^W, X_q^W)}{p_1 \cdot R(X_2^W, X_q^W) + p_2 \cdot R(X_1^W, X_q^W)} \leq 1$$

или в результате преобразований: $-2p_2 R(X_1^W, X_q^W) \leq 1$ для любых X_q^W и X_1^W .

Свойство 3. Если оценивается мера сходства w -объекта X_q^W с ближайшим w -объектом X_1^W первого класса в сравнении с ближайшим w -объектом X_2^W второго класса, то справедливо следующее:

$$F_{X_1^W / X_2^W} (X_q^W) = \begin{cases} +1, & \text{если } X_q^W = X_1^W; \\ -1, & \text{если } X_q^W = X_2^W; \\ (-1; 1), & \text{иначе.} \end{cases} \quad (6)$$

Покажем, что при $X_q^W = X_1^W$ мера $F_{X_1^W / X_2^W} (X_q^W) = 1$. Заменяем в (5)

X_1^W на X_q^W и, учитывая что $R(X_q^W, X_q^W) = 0$, получим

$$\frac{p_1 \cdot R(X_2^W, X_q^W) - p_2 \cdot R(X_q^W, X_q^W)}{p_1 \cdot R(X_2^W, X_q^W) + p_2 \cdot R(X_q^W, X_q^W)} = \frac{p_1 \cdot R(X_2^W, X_q^W)}{p_1 \cdot R(X_2^W, X_q^W)} = 1.$$

Аналогично покажем, что при $X_q^W = X_2^W$ мера $F_{X_1^W / X_2^W} (X_q^W) = -1$.

Заменяем в (5) X_2^W на X_q^W и получим

$$\frac{p_1 \cdot R(X_q^W, X_q^W) - p_2 \cdot R(X_1^W, X_q^W)}{p_1 \cdot R(X_q^W, X_q^W) + p_2 \cdot R(X_1^W, X_q^W)} = \frac{-p_2 \cdot R(X_1^W, X_q^W)}{p_2 \cdot R(X_1^W, X_q^W)} = -1.$$

Отметим, что данное свойство выполняется только при условии, что объекты равны по всем значениям признаков и весу.

Классификация объектов методом ближайшего соседа на основе функции взвешенного конкурирующего сходства. Одним из наименее трудоемких, но достаточно эффективных алгоритмов классификации объектов по обучающей выборке является метод k -ближайших соседей (k NN-классификатор) [2]. Его основу составляет поиск класса, k ближайших по выбранной метрике объектов которого располагаются на минимальном суммарном расстоянии к классифицируемому объекту. В простейшем случае классификация осуществляется по одному ближайшему объекту, а в качестве метрики может быть использована FRiS-функция [10]. Соответственно, для классификации по взвешенной

обучающей выборке оценка близости w -объектов может быть выполнена на основе $wFRiS$ -функции (4).

Для существенно пересекающихся в пространстве признаков классов классификация может также проводиться по k ближайшим w -объектам каждого из классов. Отметим, что если w -объекты были построены путем объединения множества близкорасположенных объектов исходной выборки [6, 7], то классификация по одному ближайшему w -объекту фактически равноценна классификации по k ближайшим объектам исходной выборки.

Для оценки эффективности предложенного подхода был проведен ряд экспериментальных исследований. В качестве исходных данных были использованы наборы данных репозитория UCI. В качестве критерия эффективности классификации использовалась процедура скользящего контроля. Результаты тестирования показали, что ошибка классификации по выборке w -объектов на основе $wFRiS$ -функции уменьшилась в среднем на 4,3% по сравнению с классификацией по исходной выборке с использованием $FRiS$ -функции. Для некоторых наборов данных ошибка составила не более 1,5%, что совпадает с лучшими опубликованными результатами, а в некоторых случаях и превосходит их.

Выводы. В работе предложено естественное расширение области использования функции конкурентного сходства на взвешенные выборки w -объектов ($wFRiS$ -функции), что позволяет с высокой эффективностью и минимальными затратами решать задачу классификации объектов в адаптивных системах распознавания.

Автор благодарит д.т.н., проф. Н.Г. Загоруйко за постоянный интерес и поддержку исследований автора по данной тематике.

Список литературы: 1. *Загоруйко Н.Г.* Прикладные методы анализа знаний и данных / *Н.Г. Загоруйко*. – Новосибирск: Издательство института математики, 1999. – 270 с. 2. *Theodoridis S.* Pattern Recognition / *S.Theodoridis, K. Koutroubas*. – San Diego: Academic Press, 2008. – 823 p. 3. *Загоруйко Н.Г.* Количественная мера компактности и сходства в конкурентном пространстве / *Н.Г. Загоруйко, И.А. Борисова, В.В. Дюбанов, О.А. Кутненко*. // Сиб. журн. индустр. матем. – 2010. – № 13. – С. 59–71. 4. *Zagoruiko N.G.* Methods of Recognition Based on the Function of Rival Similarity / *N.G. Zagoruiko, I.A. Borisova, V.V. Dyubanov, O.A. Kutnenko* // Pattern Recognition and Image Analysis. – 2008. – Vol. 18. – № 1. – P. 1–6. 5. *Pal S.K.* Pattern Recognition Algorithms for Data Mining: Scalability, Knowledge Discovery and Soft Granular Computing / *S.K. Pal, P. Mitra*. – Chapman and Hall/CRC, 2004. – 280 p. 6. *Волченко Е.В.* Метод построения взвешенных обучающих выборок в открытых системах распознавания / *Е.В. Волченко* // Доклады 14-й Всероссийской конференции "Математические методы распознавания образов (ММРО-14)", Суздаль, 2009. – М.: Макс-Пресс, 2009. – С. 100–104. 7. *Волченко Е.В.* Сеточный подход к построению взвешенных обучающих выборок w -объектов в адаптивных системах распознавания / *Е.В. Волченко* // Вісник Національного технічного університету "Харківський політехнічний інститут". Збірник наукових праць. Серія: Інформатика і

ISSN 2079-0031 Вестник НТУ "ХПИ", 2012, № 62 (968)

моделювання. – Харків: НТУ "ХПІ", 2011. – № 36. – С. 12 – 22. **8. Волченко Е.В.** Построение обучающей выборки w -объектов на основе коллективного решения группы экспертов / *Е.В. Волченко // Штучний інтелект.* – 2011. – № 1. – С. 147–153. . **9. Волченко О.В.** Оцінка близькості об'єктів у адаптивних системах розпізнавання, що навчаються / *О.В. Волченко // Інформаційні технології: наука, техніка, технологія, освіта, здоров'я: Тези доповідей XX міжнародної науково-практичної конференції.* – Харків, НТУ "ХПІ", 2012. – Ч. IV. – С. 8. **10. Загоруйко Н.Г.** Меры сходства, компактности, информативности и однородности обучающей выборки / *Н.Г. Загоруйко, И.А. Борисова, В.В. Дюбанов, О.А. Кутненк. // Труды Всероссийской Конференции "Знания-Онтологии-Теории" (ЗОНТ-09), Новосибирск, 2009. – Том I. – С. 93–102.*

Статью представил д.ф.-м.н., проф., зав. кафедрой системного анализа и моделирования Института информатики и искусственного интеллекта ГВУЗ "ДонНТУ" А.С. Миненко

УДК 004.93'11

Класифікація об'єктів у адаптивних системах розпізнавання на основі функції зваженої конкурентної подібності / Волченко О.В. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2012. – № 62 (968). – С. 18 – 25.

У роботі запропоновано природне розширення області використання функції конкурентної подібності на зважені навчаючі вибірки w -об'єктів у адаптивних системах розпізнавання. Описано принцип класифікації об'єктів методом k -найближчих сусідів на основі функції зваженої конкурентної подібності (wFRiS-функції). Наведено результати експериментальних досліджень, що підтвердили ефективність запропонованого підходу. Бібліогр.: 10 назв.

Ключові слова: адаптивна система розпізнавання, w -об'єкт, функція зваженої конкурентної подібності, метод k -найближчих сусідів.

UDC 004.93'1

Objects classification based on the function of rival similarity in adaptive recognition systems / Volchenko E.V. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modeling. – Kharkov: NTU "KhPI". – 2012. – №. 62 (968). – P. 18 – 25.

Author proposed the extension of the function of rival similarity that is used for weighted training samples of w -objects in adaptive recognition systems. Algorithm of k -nearest neighbors expansion was described in the article, which is based on the function of rival similarity for weighed samples of w -objects. Experimental results were confirmed the efficiency of the offered approach. Refs: 10 titles.

Keywords: adaptive recognition system, w -object, k -nearest neighbors, the function of rival similarity metric.

Поступила в редакцію 06.08.2012