

УДК 51-77

**ОЦЕНКА ИНФОРМАТИВНОСТИ КОНТУРНЫХ ДАННЫХ О
ПРОТЯЖЕННЫХ ОБЪЕКТАХ СЕЛЬХОЗНАЗНАЧЕНИЯ В
ПРОЦЕССЕ СПУТНИКОВОГО МОНИТОРИНГА**

Бочаров А.А.

Аспирант ЧОУ ВО ЮУ (ИУБИП)
E-mail: a.a.bocharov1980@gmail.com

Митясова О.Ю.

Аспирант ЧОУ ВО ЮУ (ИУБИП)
E-mail: olya.mityasova@yandex.ru

Храмов В.В.

Студент Донского Государственного технического университета
(ДГТУ), гр. СЗИТ -11,
E-mail: soullegend@inbox.ru

Рассмотрены вопросы оценки информативности целевых признаков объектов в различных источниках при спутниковом мониторинге объектов сельхозназначения, для отбора наиболее перспективных из них с точки зрения достоверности распознавания.

ИНФОРМАТИВНОСТЬ, МОНИТОРИНГ, ПРОГНОЗ ЧУВСТВИТЕЛЬНОСТИ,
ДОСТОВЕРНОСТЬ РАСПОЗНАВАНИЯ

**EVALUATION OF THE INFORMATION CONTENT OF THE
CONTOUR DATA ON THE EXTENDED OBJECTS OF AGRICULTURAL
PURPOSE IN THE PROCESS OF SATELLITE MONITORING**

Bocharov A.A.

Postgraduate student of the PEI HE SU (IMBL),
Email: a.a.bocharov1980@gmail.com

Mityasova O. Yu.

Postgraduate student of the PEI HE SU (IMBL),
E-mail: olya.mityasova@yandex.ru

Khramov V.V.

Student of Don State Technical University (DGTU), gr. SZIT-11,
E-mail: soullegend@inbox.ru

Эффективность прогнозных оценок перспектив использования информации о решении, реализованных в виде компьютерных программ, в

значительной степени зависит от того, каким образом определяется информативность признаков, какова количественная характеристика их прогнозно-поисковой значимости.

Исходя из того, что любой признак исследуемого объекта должен являться его индикатором какого-либо свойства и содержать об этом свойстве определенное количество информации, для оценки информативности признаков неоднократно предлагалось использовать математический аппарат традиционной теории информации [1]. За меру информативности при этом, в общем случае, принималось “количество информации” (I), интерпретируемое следующим образом [2]:

$$I = \log \frac{\text{вероятность признака на исследуемом объекте}}{\text{вероятность признака на изображении в целом}}, \quad (1)$$

Пусть: N – общее количество объектов распознавания (элементарных ячеек, на которые делится исследуемая территория); X – количество эталонных объектов распознавания; M – общее число объектов распознавания, на которых проявлен признак; Z – число эталонных объектов распознавания, на которых проявлен признак. В этих обозначениях формула (1) приобретает вид:

$$I = \log_2 \frac{Z/X}{M/N} = \log_2 \frac{ZN}{MX}. \quad (2)$$

Способ представления известной информации зависит от смысла задачи, которую необходимо решать. Причем большое влияние на то, как эксперт представит знания, оказывает и традиционная форма представления принятая для данной предметной области, и объем, и возможности интеллектуальной системы. Однако выбор технологии интеллектуального анализа информации должен быть сделан с учетом характера предметной области и целей обработки. В рамках подготовки известной информации к анализу необходимо решить несколько важных задач:

Какие составляющие входной информации следует учитывать при создании модели предметной области?

Какой объем информации необходимо и достаточно сохранять для адекватной работы базы знаний?

Какова технология организации базы знаний?

Какие технологии обработки информации следует использовать для создания концептуальных правил данной предметной области?

Решение всех перечисленных задач, как правило, ложится на плечи эксперта. При этом опыт человека определит общее качество решения. Зачастую недостаточно квалифицированная постановка задачи дискредитирует технологии интеллектуальной обработки информации в целом. Рядовым явлением при постановке учебных задач оказалось неумение пользователей четко описать цель анализа даже для знакомой предметной области. Однако, когда цель была описана, не менее сложным, оказалось определить признаки объектов. Самой частой ошибкой стало включение бессмысленных с точки зрения цели признаков (например, при выборе ЭВМ-сервера в качестве признака добавляется «цвет системного блока»). Не менее сложны и остальные вопросы организации модели предметной области. Однако первая составляющая постановки задачи интеллектуального анализа информации оказывает чрезвычайно важное влияние на все последующие этапы работы. Будем рассматривать наиболее простые формы представления информации – признаковые представления.

Выбор наиболее подходящей меры информативности для некоторой конкретной предметной области позволяет определить с одной стороны небольшое, а с другой стороны достаточное множество признаков. При этом нет формального критерия качества сделанного выбора, а рекомендации в пользу той или иной меры носят характер мнений эксперта и имеют неточный вид.

Процедуру поиска информативных признаков можно реализовать с использованием различных методов. Среди них можно выделить следующие подходы:

Коэффициенты корреляции, такие как коэффициенты Пирсона, Фехнера, Кендалла, Спирмена и др. [3]. Они предъявляют достаточно жесткие требования к закону распределения значений и виду признаков. Например, Коэффициент корреляции Спирмена может определять монотонные зависимости (т.е. монотонно возрастающие, либо монотонно убывающие). Он может применяться для признаков измеренных как в количественной шкале, так и в порядковой. С другой стороны, коэффициент Пирсона применяется для вычисления силы линейной связи между количественными признаками

Информативность по Шеннону, которая использует меру трудности распознавания, как количество попарно одинаковых битов векторов [4].

Все предъявленные меры с той или иной точностью решают вопрос информативности. Для сравнения результатов их работы провели [5-7] численный эксперимент.

На одном и том же наборе данных был получен существенно отличающийся порядок признаков с точки зрения информативности.

Результаты решения одной и той же задачи существенно зависят от способа вычисления информативности. Это хорошо видно для случая пространства из 50 признаков, где разброс ошибки разделения объектов достиг 9.1%. Качество выделения информативной составляющей зависит от применяемой меры и ее совместимости с конкретной задачей, что не всегда легко определить при постановке интеллектуальной задачи. Таким образом, подбор способа определения наиболее важных составляющих описания носит все признаки плохо формализуемой задачи. Следовательно, имеет смысл решать задачу выбора метода вычисления информативности как задачи интеллектуального анализа информации.

Будем рассматривать наиболее простой подвид фреймовых представлений – признаковые. Объектом исследования в данной постановке задачи является пространство признаков и множество объектов, для которых известно значение признаков и имя целевого класса или значение

количественной меры цели. Такое представление некоторой предметной области задается через набор признаков в следующем виде:

$$\begin{cases} \bar{p}^1 = \{o_1^1, o_2^1, \dots, o_n^1\} \\ \bar{p}^2 = \{o_1^2, o_2^2, \dots, o_n^2\} \\ \bar{p}^m = \{o_1^m, o_2^m, \dots, o_n^m\} \end{cases} \quad (3)$$

$$C = \{c_1, c_2, \dots, c_n\}$$

\bar{p}^j – множество значений j -го признака объектов предметной области,

o_i^j – значение j -го признака i -го объекта предметной области,

C – множество значений цели, c_i – значение цели i -го объекта,

n – число рассматриваемых объектов предметной области.

Для удобства в дальнейшем \bar{P}^j будем называть вектором-признаком, а C вектором-целью. Каждый признак \bar{P}^j имеет определенный формат. Признак может быть номинальным, дихотомическим, порядковым или количественным. В выше приведенном примере вектором-целью является тип объекта «шум» или «метка», а векторами-признаками – значения коэффициентов спектра.

Из множества признаков выбираем один, например \bar{P}^k и обозначим его символом P ($P = \bar{P}^k$). Размерность векторов P и C равна n и $\bar{P} = \{p_1, p_2, \dots, p_n\}$, $\bar{C} = \{c_1, c_2, \dots, c_n\}$.

В рамках численного эксперимента P и C формировались как случайные последовательности значений. Для анализа методов вычисления информативности введена дополнительная «идеальная мера», которая определена как косинус угла между векторами \bar{P} и \bar{C} в n -мерном пространстве, где n – число рассматриваемых объектов предметной области и вычисляется по формуле:

$$\cos(\bar{P}\bar{C}) = \frac{\langle \bar{P} \cdot \bar{C} \rangle}{|\bar{P}| \cdot |\bar{C}|}$$

где $\langle \bar{P} \cdot \bar{C} \rangle$ – скалярное произведение векторов, а $|\bar{P}| \cdot |\bar{C}|$ – произведение модулей соответствующих векторов.

Для полностью коррелированных последовательностей $\cos(\overline{P_1 C_1}) = 1$, а для полностью различных $\cos(\overline{P_0 C_0}) = 0$.

Исследуется отношение некоторой мер информативности для векторов P и C к их «идеальной мере». Анализа различных методов вычисления информативности провели на основе нескольких тысяч вычислительных экспериментов и построили интервалы значений, которые принимали меры информативности.

Для 1000 пар векторов \overline{P} и \overline{C} получены результаты по мерам информативности Пирсона, Шеннона и Хемминга, которые приведены на

Анализ результатов показал, что методы вычисления информативности можно разделить на несколько групп. Условно были выделены три группы методов: с жесткими, слабыми и средними требованиями. Жесткими считаем требования, которые лежат выше «идеальной меры информативности», слабыми – те, которые лежат ниже, а средними – требования, которые совпадают с «идеальной меры информативности».

Сведение разброса значений по разным коэффициентам позволило получить дисперсию и среднее отклонение.

На основании исследования мер информативности удалось сформировать нечеткую базу правил. В качестве входных параметров используются количество объектов, количество признаков, вычисленная в результате эксперимента мера жесткости мер информативности, процент отбрасываемых признаков и эвристические представления эксперта. На основе построенной базы знаний создана система, использующая нечеткий вывод по Мамдани [5], которая определяет степень жесткости требований эксперта. В результате удается получить сокращенное подмножество рекомендуемых мер информативности. В частности для выше приведенного примера выбор меры информативности при сокращении пространства на 50%(50 признаков) требования эксперта признаны жесткими, а в качестве мер информативности следует использовать коэффициент Фехнера, меру Хемминга или Шеннона, но т.к. признаки носят количественный характер, то

наиболее пригодной признан корреляционный коэффициент Фехнера. Эта мера позволяет получить качество решения на уровне 22% ошибок.

Таким образом, исследован ряд методов вычисления информативности и показано, что они дают существенно отличающиеся решения.

Качество решения интеллектуальных задач значительно зависит от набора признаков, а следовательно необходимо решить вопрос выбора наиболее подходящей модели информативности для каждой задачи и каждого из признаков.

Разработана технология анализа мер информативности и построена база правил для выбора наиболее подходящей среди всех доступных в каждой конкретной предметной области.

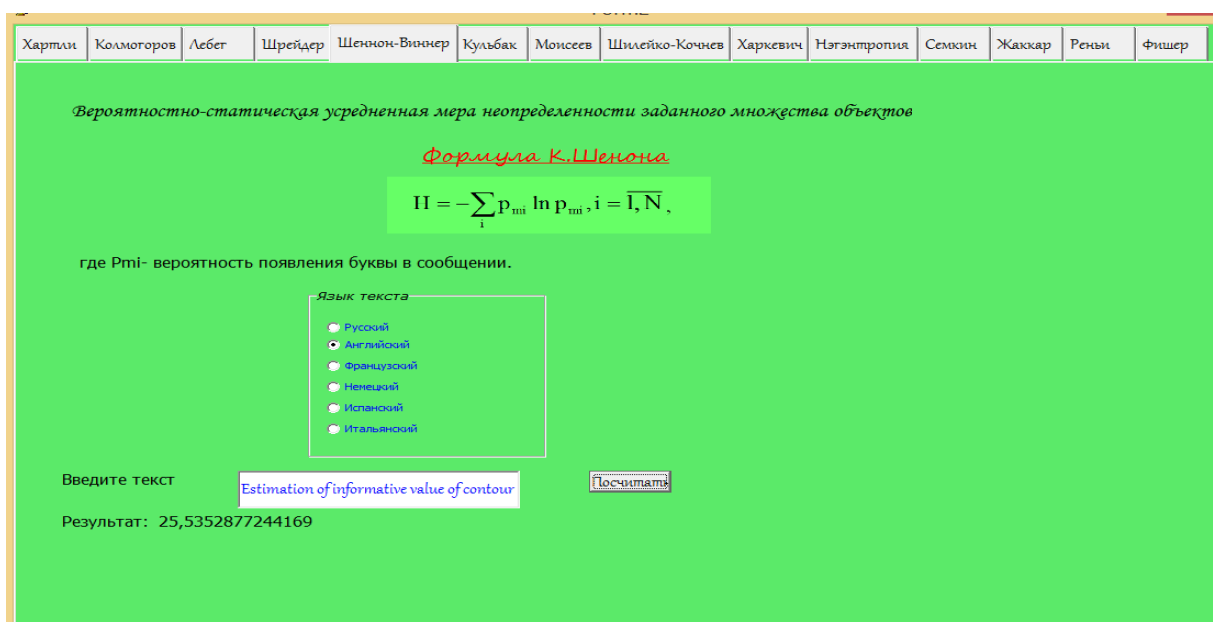


Рисунок 1 – Программное средство для оценки количества информации, получаемой от источника

На основе разработанной технологии анализа мер информативности, были проанализированы некоторые другие меры, что позволяет расширить доступное для работы множество методов вычисления близости элементов описания объектов. Было разработано программное средство для оценки количества информации в различных источниках и с использованием

различных носителей информации. Рабочее окно этого средства представлено на Рисунок 1.

Дальнейшая работа будет направлена на детализацию базы правил с целью повышения качества принимаемого системой решения.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК:

1. Загоруйко, Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: Изво Ин-та математики, 1999. – 270 с.
2. Вагин, В.Н. Достоверный и правдоподобный вывод в интеллектуальных системах/ Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В. – М.: Физматлит, 2004. – 704 с.
3. Корлякова, М.О. Анализ подходов к определению информативности признаков. // Научная сессия МИФИ-2006. Сборник научных трудов. В 16 томах. Т.3. Интеллектуальные системы и технологии./ М.О. Корлякова, Н.С. Твердохлеб.- М.: МИФИ, 2006. 256 с. С. 146-147
4. Гайдышев, И. Анализ и обработка данных: специальный справочник. – СПб: Питер, 2001. – 752 с
5. Храмов В.В. Теория информационных процессов и систем. Учебно-методическое пособие, Ростов-на-Дону, РГУПС, 2011.- 87с.
6. Гвоздев, Д.С Прикладные методы идентификации в автоматизированных системах на транспорте: монография / Д.С. Гвоздев, В.В. Храмов, С.М. Ковалев, Е.В. Голубенко; ФГБОУ ВПО РГУПС.- Ростов-на-Дону, 2015.-186 с.
7. Корлякова М.О. Выбор метода вычислений информативности признаков // Научная сессия МИФИ-2006. Сборник научных трудов. В 16 томах. Т.4. Интеллектуальные системы и технологии./ М.О. Корлякова, Н.С. Твердохлеб.- М.: МИФИ, 2008. 221 с. С. 106-110
8. Храмов В.В. Методы и модели обнаружения и распознавания протяженных объектов на земной поверхности // Системные проблемы надёжности, качества, математического моделирования и информационных технологий в инновационных проектах. Коллективная монография. Под ред. проф. Кофанова Ю.Н./ Изд. НИИВШЭ. – Москва. 2013. С.11-23
9. Kramarov S. The principles of formation of united geo-informational space based on fuzzy triangulation/ S/ Kramarov, I.Temkin, V. Khramov // 9 th INTERNATIONAL CONFERENCE ON SOFT COMPUTING, COMPUTING WITH WORDS AND PERCEPTIONS ICSCCW 2017, 22-23 August 2017, Budapest, Hungary