

Л.В. АНТОНОВ,  
А.Д. ВАРЛАМОВ

**Использование метрик при  
интеллектуальном мониторинге  
состояний крупного рогатого скота**

УДК 004.896

Муромский институт  
(филиал) ФГБОУ ВО  
«Владимирский  
государственный  
университет имени  
А.Г. и Н.Г. Столетовых»,  
г. Муром

*Приведены и описаны основные метрики, которые используются для оценки качества алгоритмов поиска больных животных, описаны их свойства и особенности применения в разных видах поиска изображений. Приведен алгоритм, основанный на анализе отклонений параметров животных.*

В настоящее время активно развивается рынок информационных систем автоматизации животноводческих хозяйств. В рамках автоматизации решается ряд задач, связанных с мониторингом производственного процесса. Задачи мониторинга, в первую очередь, решают проблемы поиска потенциально больных животных на основе выявления некоторых тенденций и закономерностей при анализе данных с датчиков [2,4]. Разработкой алгоритмов, осуществляющих подобный анализ и позволяющих найти проблемные объекты производства занимаются ученые многих стран [9]. В связи с этим становится актуальной задача анализа оценки качества разработанных алгоритмов поиска больных животных.

Для оценки качества работы алгоритма необходимо сопоставить результаты его работы с мнением эксперта(ов). При использовании интеллектуальных методов анализа данных экспертные оценки присутствуют в обучающей выборке в качестве целевых значений [5,7,8]. Если оценки экспертов полностью совпадают с результатами поиска заболевших животных, то алгоритм является идеальным, как и информационная система, основанная на нем. На практике, результаты работы различных алгоритмов в разной степени близки к результатам экспертных

оценок, но также существенно отличаются друг от друга, так как в их основе лежат разные признаки поиска больных животных. Мерами оценки результатов работы алгоритмов являются метрики – численные характеристики, полученные путем сопоставления результатов работы системы с выводами экспертов, сделанными по отношению к одному и тому же стаду [1,6].

Рассмотрим задачу с точки зрения информационного поиска: система должна из множества голов крупного рогатого скота животноводческого комплекса идентифицировать нездоровых особей. Так как всех животных предприятия можно разделить на больных и здоровых по фактическому состоянию здоровья и на больных и здоровых по "мнению" системы, можно составить таблицу сопряженности:

Таблица 1

**Общий вид таблицы сопряженности**

	<b>Больные</b>	<b>Здоровые</b>
<b>Найденные</b>	<i>SF</i>	<i>HF</i>
<b>Не найденные</b>	<i>SN</i>	<i>HN</i>

Таким образом, каждое животное попадет в одну из следующих групп:

*SF* – выявленные больные особи;

*SN* – не выявленные больные особи;

*HF* – ложное подозрение на болезнь животного;

*HN* – здоровые особи, не вызвавшие подозрений у системы.

Нужно отметить, что по одному критерию животных делит человек-эксперт (зоотехник), а по другому – система. По значениям элементов таблицы вычисляется большинство метрик, которые рассмотрим ниже.

### 1. Полнота

Полнота (*recall*) определяется как отношение количества больных особей с диагностированной немощью к общему количеству животных с недугом:

$$r = \frac{SF}{SF + SN} \quad (1)$$

Она характеризует диагностическую способность системы не упускать заболевших животных. Чем выше полнота, тем меньше нездоровых животных будет упущено при поиске. Полнота никак не учитывает количество диагностических ошибок (ложное отнесение здоровых голов к числу подозреваемых в потере здоровья), поэтому не может применяться в одиночку как мера качества.

## 2. Точность

Точность (precision) вычисляется как отношение выявленных больных особей к общему количеству животных с признаками болезни по мнению системы:

$$p = \frac{SF}{SF + HF} \quad (2)$$

Точность характеризует способность системы отбирать только действительно больных особей, однако, не учитывает количество заболевших животных, которые не выдала система. Точность, как и полнота, не может применяться в качестве единственного критерия качества при поиске. Обычно полноту и точность используют совместно.

## 3 F-мера

F-мера (F-measure) используется как единая метрика, объединяющая полноту и точность. Она вычисляется как гармоническое среднее полноты и точности:

$$F_1 = \frac{2}{\frac{1}{r} + \frac{1}{p}} = \frac{2rp}{r+p} \quad (3)$$

## 4. R-точность

Для описания этой метрики сначала дадим определение другой характеристике с учетом специфики решаемой задачи. Точность на уровне  $n$  диагностируемых животных определяется как количество больных особей среди первых  $n$  найденных голов рогатого скота, деленное на  $n$ .

R-точность (R-precision) равна точности на уровне R особей для R, равному количеству голов животных с хворью.

$$R = SF + SN \quad (4)$$

Эта метрика малочувствительна к доле заболевших особей во всем поголовье.

#### 5. Средняя точность

Средняя точность определяется следующим образом: пусть на ферме имеется R нездоровых особей. Точность на уровне i-го больного животного  $p(i)$  равна точности на уровне  $pos(i)$  особей, если i-е больное животное находится на позиции  $pos(i)$  в списке упорядоченных по степени подозрения болезни особей. Если заболевание i-го больного животного не диагностировано, то  $p(i)=0$ . Средняя точность равна среднему значению величины  $p(i)$  по всем R больным особям:

$$p\_aver = \frac{1}{R} \sum_{i=1}^R p(i), R = SF + SN \quad (5)$$

#### 6. Аккуратность

Аккуратность (ассигасу) - это отношение правильно принятых системой решений к общему числу решений. Формально вычисляется по формуле:

$$a = \frac{SF + HN}{SF + HN + HF + SN} \quad (6)$$

#### 7. Ошибка

Ошибка (error) - это отношение неправильно принятых системой решений к общему числу решений. Формально вычисляется по формуле:

$$e = \frac{HF + SN}{SF + HN + HF + SN} \quad (7)$$

#### 8. Информационный шум

Информационный шум вычисляется как отношение числа ложных заключений о болезни животного к общему количеству заключений о болезни. Формально вычисляется по формуле:

$$n = \frac{HF}{SF + HF} \quad (8)$$

#### 9 Потери информации

Потери информации - отношение числа пропущенных системой искомых больных животных к общему количеству больных особей:

$$l = \frac{SN}{SF + SN} \quad (9)$$

#### 10 Специфичность

Специфичность (specificity) - отношение числа здоровых животных, не заподозренных системой в заболеваниях, к общему количеству здоровых особей.

$$sp = \frac{HN}{HF + HN} \quad (10)$$

#### 11 Избирательность

Избирательность (selectivity) - это отношение количества животных, отнесенных системой к больным, к общему поголовью на предприятии.

$$e = \frac{SF + HF}{SF + HN + HF + SN} \quad (11)$$

Избирательность характеризует способность отбирать больных животных, исключая здоровых особей. Специфичность и избирательность практически применяются при оценке эффективности классификации только в случаях особой необходимости.

Результаты исследований.

Для исследования был взят разработанный авторами алгоритм, основанный на анализе отклонений параметров животных [3]. Применив алгоритм к подготовленной тестовой выборке данных и, рассчитав метрики, были получены величины, отображенные в таблице 2.

Таблица 2

**Таблица сопряженности для алгоритма отслеживания отклонений параметров животных.**

	Больные	Здоровые
Найденные	39	4
Не найденные	7	342

Исходя из данных величин, были рассчитаны метрики, которые сведены в таблицу 3.

Таблица 3

**Расчет метрик для алгоритма отслеживания отклонений параметров животных.**

Метрика	Значение
Полнота	<i>0,848</i>
Точность	<i>0,907</i>
F-мера	<i>0,877</i>
Средняя точность	<i>0,802</i>
Аккуратность	<i>0,972</i>
Ошибка	<i>0,028</i>
Информативный шум	<i>0,093</i>
Потеря информации	<i>0,152</i>
Специфичность	<i>0,988</i>
Избирательность	<i>0,110</i>

Полученные значения могут быть использованы для балансирования соотношения количества ложных диагнозов болезни к числу не найденных больных особей (увеличение полноты за счет снижения точности и наоборот), а также сравнения алгоритма с его аналогами.

В заключении можно отметить, что выбор метрик для интеллектуального анализа состояния здоровья животных возлагается на программиста, однако, сравнение алгоритмов диагностики состояния животных и выбор лучшего из них рекомендуется проводить с использованием описанных метрик.

## Литература

1. Агеев М., Кураленок И., Некрестьянов И. Официальные метрики РОМИП 2010. "Российский семинар по Оценке Методов Информационного Поиска. Труды РОМИП 2010". Казань, 2010, с. 172-187.

2. Антонов Л.В. Алгоритм мониторинга критических изменений параметров производственного процесса животноводческого предприятия // Алгоритмы, методы и системы обработки данных. № 29. 2014. С. 3-12

3. Антонов Л.В., Варламов А.Д., Орлов А.А. Разработка адаптивного алгоритма отслеживания отклонений параметров животных в системе управления животноводческим предприятием. Динамика сложных систем - XXI век. 2015. Т. 9. № 2. С. 44-49.

4. Антонов Л.В., Варламов А.Д. Автоматизация процесса мониторинга животноводческого предприятия на основе исследования временных рядов параметров крупного рогатого скота // Современные проблемы науки и образования. 2013. № 6. С. 49-57.

5. Баженов Р.И., Лопатин Д.К. О применении современных технологий в разработке интеллектуальных систем // Журнал научных публикаций аспирантов и докторантов. 2014. № 3 (93). С. 263-264.

6. Варламов А.Д. Основные метрики, оценивающие качество работы систем поиска изображений // Алгоритмы, методы и системы обработки данных. № 24. 2013. С. 3-12.

7. Варламов А.Д., Шарапов Р.В. Использование нейронных сетей в задачах мониторинга экзогенных процессов дистанционными методами // Геоинформатика. 2014. № 4. С. 62-68.

8. Николаев С.В., Пронина О.Ю., Баженов Р.И. Исследование методов интеллектуального анализа для формирования краткосрочного прогноза в программной среде Statistica // Экономика и менеджмент инновационных технологий. 2015. № 7 [Электронный ресурс]. URL: <http://ekonomika.snauka.ru/2015/07/9500> (дата обращения: 29.07.2015).

9. Орлов А.А., Антонов Л.В. Обзор и анализ современных информационных решений автоматизации животноводческих хозяйств // Современные проблемы науки и образования. 2013. № 6 С. 58.