УДК 519.254

Плотность интервальных оценок

С.И. Жилин[1,2]
[1]*ООО «СиСорт», г. Барнаул;* [2]*АлтГУ, г. Барнаул*

Интервальный подход к внешнему оцениванию параметров зависимостей по экспериментальным данным с неопределенностями заключается в построении интервальных брусов, содержащих в себе информационное множество, т.е. множество возможных значений параметров зависимости, совместных с наблюдениями, видом зависимости и априорными ограничениями на значения параметров [1-5]. Гарантирующий характер интервальной оценки, т.е. свойство заключать в себе все точки информационного множества является одним из достоинств интервального оценивания. Однако наряду с информационным множеством, интервальная оценка почти всегда содержит в себе и избыточные точки пространства параметров. Некоторое представление о соотношении информационного множества и его интервальной оценки может давать отношение их объемов. Вполне естественным представляется именовать эту величину *плотностью интервальной оценки*.

Однако такой интегральный показатель не позволяет описать особенности формы и локализации информационного множества внутри бруса интервальной оценки. В этом смысле более информативными могут оказаться своего рода «проекции» информационного множества на оси пространства параметров, которые могут отражать распределение объема информационного множества вдоль каждой из осей. Более формально каждая из таких «проекций» может выражаться функцией, сопоставляющей точке на оси объем сечения информационного множества плоскостью, проходящей через эту точку ортогонально оси. Такая функция, отнормированная на объем сечения той же плоскостью бруса интервальной оценки, может рассматриваться как *функция распределения плотности интервальной оценки* вдоль соответствующей оси.

В совокупности функции распределения плотности интервальной оценки дают довольно содержательное представление об устройстве информационного множества и могут быть полезны при решении следующих задач.

1. Визуальное представление информационного множества для анализа и интерпретации. Как известно, информационное множество может иметь довольно сложное строение для его непосредственной визуализации и восприятия пользователем. Особенно актуальным этот вопрос становится при размерности пространства параметров выше трех, когда непосредственная визуализация информационного множества становится невозможной. Функции распределения плотности интервальной оценки могут служить некоторым компромиссным представлением, в том числе, и визуальным (рисунок 1), — менее сложным, чем полное описание информационного множества, но более информативным, чем границы интервальной оценки.

2. Построение менее консервативных (более рискованных) интервальных оценок. В некоторых задачах имеет смысл поступиться свойством гарантированности интервальных оценок ради получения более узких интервальных оценок [3]. Функции распределения плотности интервальной оценки могут быть основой для обоснованного выбора границ таких новых оценок. Известно [6], что точки информационного множества, лежащие вблизи его границы и, в особенности, близкие к его вершинам, на практике относительно редко могут соответствовать истинному оцениваемому значению. Поэтому логично границы более рискованных оценок выбирать как процентили функций распределения плотности интервальной оценки рисунок 1).

3. Построение точечных оценок. В практических целях наряду с интервальными оценками параметров, бывают необходимы и точечные оценки, соответствующие тем или иным образом выбранным точкам информационного множества [2, 4]. К уже известным вариантам конструирования точечных оценок таким, как геометрический центр интервальной оценки, чебышевский центр, центр Оскорбина и т.п., можно добавить и алгоритмы, опирающиеся на функции распределения плотности интервальной оценки. К примеру, небессмысленными оценками могут быть точка, соответствующая максимальным значениям функций распределения плотности по каждому из параметров, либо точка, соответствующая 50%-м процентилям функций распределения плотности (рисунок 1). Выяснение свойств этих точечных оценок и их сравнительное исследование по отношению к прочим представляется интересной задачей.

Следует ожидать, что список задач, при решении которых, могут стать полезными функции распределения плотности интервальной оценки, будет расширяться.

По аналогии с одномерными функциями распределения плотности интервальной оценки могут быть определены и двумерные «проекции» информационного множества на плоскости, определяемые какой-либо парой параметров модели.

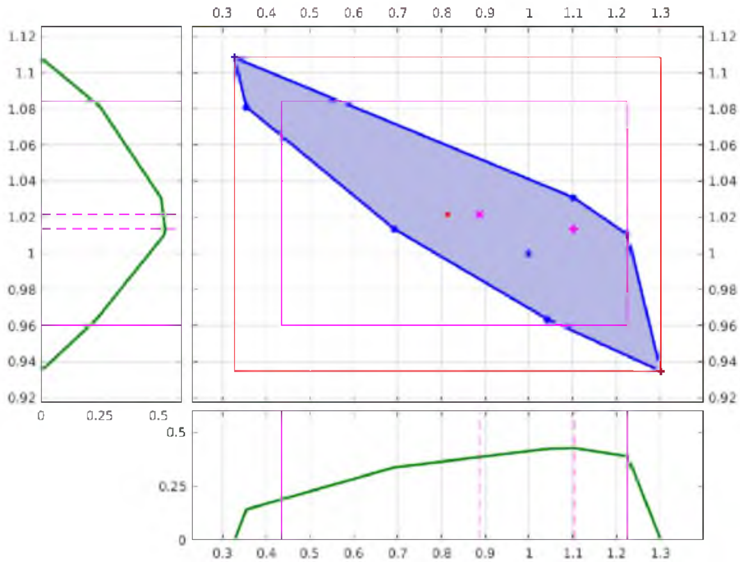

Рисунок 1 – Пример информационного множеств
для двухпараметрической линейной модели и соответствующих
функций распределения плотности интервальной оценки

Показаны гарантированная интервальная оценка (больший прямоугольник) и оценка, соответствующая 5%-м процентилям функций распределения плотности (меньший прямоугольник). Символом * показано истинное значение параметров, символом • показан геометрический центр гарантированной интервальной оценки, символами ×.и + показаны точечные оценки: соответствующая 50%-м процентилям и пикам функций распределения плотности соответственно.

Важно заметить, что задача вычисления объема выпуклых многомерных полиэдров довольно сложна и относится к классу #*P* [7]. Тем не менее, для размерностей пространства до 100 разработаны алгоритмы, позволяющие оценить объем выпуклых полиэдров [8] за практически приемлемое время с практически приемлемой точностью.

Библиографический список

1. Канторович Л.В. О некоторых новых подходах к вычислительным методам и обработке наблюдений // Сиб. мат. журнал. – 1962. – Т. 3, №5. – С. 701–709.

2. Вощинин А.П., Бочков А.Ф., Сотиров Г.Р. Метод анализа данных при интервальной нестатистической ошибке // Заводская лаборатория. – 1990. – Т. 56, №7. – С. 76–81.

3. Вощинин А.П., Сотиров Г.Р. Оптимизация в условиях неопределенности. – Изд-во МЭИ (СССР); Техника (НРБ), 1989.

4. Оскорбин Н.М., Максимов А.В., Жилин С.И. Построение и анализ эмпирических зависимостей методом центра неопределенности // Известия Алтайского государственного университета. – 1998. – №1. – С. 37–40.

5. Шарый С.П. Конечномерный интервальный анализ. – Новосибирск, 2017.

6. Шарый С.П. Интервальный анализ или методы Монте-Карло? // Вычислительные технологии. – 2007. – Т. 12, №1. – С. 103–115.

7. Dyer M.E., Frieze A.M. On the complexity of computing the volume of a polyhedron // SIAM Journal of Computing 17 (5), 1988, pp. 967–974.

8. Cousins B., Vempala S. A practical volume algorithm // Mathematical Programming Computation 8 (2), 2016, pp 133–160.

УДК 004.89

Применение машинного обучения к задачам анализа историй болезней детей с заболеваниями почек

Д.П. Налимов

АлтГУ, Барнаул

На сегодняшний день такой раздел искусственного интеллекта, как машинное обучение имеет приложения в разнообразных областях знаний. Комбинируя в себе методы математической статистики, оптимизации, теории вероятностей, алгоритмов и графов, алгебры, математического анализа и других наук, машинное обучение позволяет решать множество задач: от кредитного скоринга и построения рекомендаций до генерации изображений и музыкальных композиций. Очень важным и перспективным является применение методов машинного обучения в медицине (в частности, в доказательной медицине). Сюда относится диагностика заболеваний, прогнозирование состояния пациента, создание индивидуальной терапии, проверка эффективности лекарственных препаратов и многое другое.

При применении методов машинного обучения для диагностики заболеваний возникает множество проблем. Например, исследуемые данные неструктурированны: выписки пациентов часто отличатся от шаблона, содержат массу опечаток, неточностей, а порой представлены в рукописном виде. Признаков зачастую огромное количество: различные показатели анализов, данные осмотра пациента, результаты терапий и прочее, поэтому процесс отбора наиболее значимых призна-