Идентификация механизмов комплексного оценивания как

подход к анализу порядковых данных

на данных по банкротствам

Сергеев В.А.1, Коргин Н.А.1  
(sergeev@bureau@gmail.com)

1Институт проблем управления им. Трапезникова В.А. РАН, Москва, Россия

Мы предлагаем новый подход к анализу порядковых данных (см., например, Mariel (2021)) напоминающий подход вербального анализа решений (ВАР) (см., Например, Moshkovich, Mechitov (2013), Larichev (1997)) на основе подхода к идентификации механизмов комплексного оценивания (МКО), первоначально разработанного в Burkov, Sergeev, Korgin (2020) и Korgin, Sergeev (2021). Предлагаемый подход может быть дополнением к основным подходам, применяемым в экологической оценке (ЭО), и экспериментах дискретного выбора (ЭДВ) основанных на статистических методах (см., Например, Pecherkova, Nagy (2017), Hoyos (2010)). Мы проиллюстрируем его применение к анализу порядковых данных на основе данных оценки компаний, занимающихся недвижимостью, представленных в работе Alekseev, Alekseeva (2019).

В работе использована методология, основными понятиями которой являются: дерево свертки, матрица свертки, оптимизационный полином, анализ групп эквивалентности. Дерево свертки - бинарное дерево, листья которого являются показателями, используемыми для оценки объектов, а корень - интегральная оценка, используемая для ранжирования. Исторически для определения элементов МКО - структуры и матриц, используется экспертный подход. Рассматриваемый в данной работе подход, напротив, основан на получении элементов МКО на основе работы с набором входных данных исследуемой задачи. В этой статье мы делаем шаг вперед и используем МКО не как механизм принятия решений, а как модель, описывающую набор порядковых данных, но под именем МКО. Основным преимуществом по сравнению с нашей предыдущей работой является расширение метода для работы с неполными данными. Матрицы свертки определяют значения оценок в промежуточных узлах и в корне дерева свертки. Существует механизм, позволяющий представить каждое дерево свертки в виде ML, например: *M1 l1 M2 l2 M3 l3 l4*. Проблема идентификации может быть сведена к задаче максимизации некоторого полинома, полученного для данного двоичного дерева и набора примеров с использованием унитарного кодирования. Для задачи максимизации мы используем бесплатный для академических целей Gurobi Solver, см. Gurobi (2021).

Предлагается анализ групп эквивалентности, который позволяет выделить заведомо нереализуемые структуры, в которых есть подгруппы переменных, не укладывающиеся в допустимый алфавит. В соответствии с изложенным в Burkov, Sergeev, Korgin (2020), для реализации МКО количество групп эквивалентности должно быть меньше или равно размеру шкалы: тогда и только тогда, когда . Подробно теория изложена в работах: Burkov, Sergeev, Korgin (2020), Korgin, Sergeev (2021).

На основании Alekseev, Alekseeva (2019) используются данные оценки компаний, занимающихся недвижимостью, в зависимости от пяти параметров: дебиторская задолженность, кредиторская задолженность, нераспределенная прибыль, запасы, основные средства. Первоначально рассматривалось двести российских строительных компаний, из которых 100 ликвидированы или находятся в процессе ликвидации в связи с банкротством и 100 экономически успешных компаний, по которым дела о банкротстве не были возбуждены. В результате обработки данных осталось сорок восемь примеров, дублирующие и противоречивые примеры были удалены. Проверка всех возможных ста пяти структур полных деревьев с именованными листьями показала, что наилучший результат аппроксимации для этой задачи реализует сорок пять из сорока восьми, и не существует структуры, которая реализует все сорок восемь примеров. Наилучшая аппроксимация данных достигается по структуре: M1l3M2l4M3l5M4l1l2. По-прежнему мы не можем найти реализующиеся структуры в шкале три, но если мы изменим шкалу на пять, то окажется, что мы найдем пятнадцать структур, которые могли бы быть реализованы.

Рассматреный случай показывает, что предложенный подход к анализу наборов порядковых данных в качественных или категориальных оценках позволяет получить описание исходных данных в виде механизма. Используя анализ групп эквивалентности и повышение шкалы, мы построили несколько механизмов для описываемых данных. Также были показаны возможности анализа влияния отдельных переменных и групп переменных на конечное значение интегрального показателя и выделения целевых примеров для дальнейшей идентификации модели. Эта функциональность вполне применима к областям, где традиционно используется ЭДВ, таким как маркетинг, здравоохранение и оценка окружающей среды.

Alekseev A., Alekseeva I., Noskova A., Kylosova V., Knyazeva A. (2019). New Applications of the Annual Fiscal Accounting Analysis at the Risk Assessment. Advances in Computer Science Research, 2019, vol. 85, pp. 61-66. DOI 10.2991/cmdm-18.11.

Burkov, V., Sergeev, V., Korgin, N. (2020). Identification of Integrated Rating Mechanisms as Optimization Problem. *13th International Conference "Management of large-scale system development" (MLSD)*, IEEE, pp. 1-5.

Gurobi Optimization, LLC. (2021). Gurobi Optimizer Reference Manual. https://www.gurobi.com

Hoyos, D. (2010). The state of the art of environmental valuation with discrete choice experiments*. Ecological Economics,* volume 69, issue 8, pp. 1595-1603*.*

Korgin N. and Sergeev V. (2021) Identification of Integrated Rating Mechanisms on Complete Data Sets // Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems. Springer. 2021 Nantes, France. Proceedings, Part I. pp. 610-616

Larichev, O. and Moshkovich, H. (1997). *Verbal decision analysis for unstructured problems*. Boston: Springer.

Mariel, P. et al. (2021). Environmental valuation with discrete choice experiments: Guidance on design, implementation and data analysis. *Springer Nature*, 2021. pp. 129.

Moshkovich, H. and Mechitov, A. (2013). Verbal Decision Analysis: Foundations and Trends. *Advances in Decision Sciences*, volume 2013, 697072, pp 1-9.