

Берсенева Е.А.¹, Черкасов С.Н.¹, Мешков Д.О.¹

МЕТОДИЧЕСКИЕ ПОДХОДЫ К КЛАССИФИКАЦИИ ОБЪЕКТОВ НАДЗОРА ПО ВЕЛИЧИНЕ РИСКА

¹ ФГБНУ «Национальный научно-исследовательский институт
общественного здоровья им. Н.А. Семашко», Москва, Россия

Berseneva E.A.¹, Cherkasov S.N.¹, Meshkov D.O.¹

CONCEPTUAL APPROACHES TO CREATION RISK-FOCUSED MODEL OF SUPERVISION OBJECTS CLASSIFICATION

¹ National Research Institute for Public Health, Moscow, Russia

Резюме. В статье рассматриваются и обосновываются концептуальные подходы к созданию риск-ориентированной модели классификации объектов надзора.

Ключевые слова: риск-ориентированная модель, информационные технологии, экспертные системы.

Abstract. In article conceptual approaches to creation risk - focused model of supervision objects classification are considered and proved.

Keywords: risk-focused model, information technologies, expert systems.

Федеральным законом № 246-ФЗ от 13.07.2015 г. было введено применение риск-ориентированного подхода при организации государственного контроля (надзора) (статья 8.1.). В соответствии с данным федеральным законом в целях оптимального использования трудовых, материальных и финансовых ресурсов, задействованных при осуществлении государственного контроля (надзора) должен применяться риск-ориентированный подход. Данный подход представляет собой метод организации и осуществления государственного контроля (надзора), при котором в предусмотренных

настоящим Федеральным законом случаях выбор интенсивности (формы, продолжительности, периодичности) проведения мероприятий по контролю определяется отношением деятельности юридического лица к определенной категории риска либо определенному классу (категории) опасности.

Для осуществления данного подхода при организации контрольно-надзорных мероприятий в рамках деятельности Росздравнадзора необходимо разработать соответствующие методические подходы.

Решая задачу разработки такой методики следует учитывать как уникальные свойства объектов классификации, так и специфику задачи.

Сформулируем общие требования к методике:

1. Результат использования методики должен быть практически применим. Не допускается выработка подхода, делающего возможным только формирование теоретических заключений об особенностях тех или иных объектов.
2. Накопление результатов оценки должно повышать точность применения оценки.
3. Алгоритм оценки должен быть прозрачен, и доступен для критического осмысления и обсуждения в экспертном сообществе.
4. Алгоритм должен содержать резервы изменения параметров без утери накопленных ранее данных.
5. Алгоритм должен быть устойчив к выпадению части данных без риска существенного ухудшения результатов оценки.
6. Алгоритм должен быть устойчив к манипуляциям данными со стороны оцениваемых объектов.
7. Применение методики не должно требовать постоянного участия квалифицированных математиков. Это замечание не касается вопросов дальнейшего развития методики.
8. Применение методики не должно приводить к выборочному влиянию одних оцениваемых объектов на другие.

При выработке излагаемой в настоящей работе методики были приняты во внимание следующие особенности обследуемых объектов:

1. Объекты делятся на несколько классов (аптеки, фармдилеры, производители, и.т.д). Классы имеют свои уникальные

особенности, и их свойства в большинстве своём, не пересекаются со свойствами других классов.

2. Каждый объект является весьма сложно организованной структурой, и не может быть полностью описан с использованием существующих математических методов.
3. Заранее известен полный перечень объектов, подлежащих классификации, не ожидается появление объектов новых классов.
4. Не ожидается преобразование объектов одних классов в другие.
5. При этом ожидается изменение свойств объектов с течением времени как под воздействием внешних факторов, так и непосредственно по результатам выборочных проверок а так же других внутренних изменений.

В разработке методики нами были выделены два отдельных этапа:

1. Выработка критериев описания объектов. Выполняется ранее всех остальных мероприятий по созданию методики. Необходимость выполнения данного этапа продиктована прежде всего тем, что любой из обследуемых объектов является весьма сложной структурой, и требуется выработка новых, либо выбор из существующих, конечного числа критериев оценки объекта, а так же тем, что известно, что объекты делятся на несколько категорий. Сама выработка критериев не обеспечивает создания какого-либо правила оценки или классификации объектов, а служит исключительно для создания способа описания обозначенных категорий объектов.
2. Выработка методики оценки объектов. Любая методика должна быть применена к выработанным ранее критериям. При выработке методики, либо выборе из общепринятых, был рассмотрен максимально широкий спектр математических аппаратов и алгоритмов.

При выработке критериев необходимо использовать опыт экспертного сообщества. Основываясь на данном опыте, были выделены следующие классы оцениваемых объектов:

1. Аптеки
2. Фармдилеры
3. Производители
4. Лечебные учреждения
5. Производители медицинских изделий

Далее для каждого из классов должен быть разработан свой, уникальный набор критериев, описаны способы и источники получения данных для вычисления этих критериев.

Описывая порядок выработки методики необходимо остановиться на следующих этапах:

1. Рассмотрение существующих математических аппаратов.
2. Аргументация в пользу тех или иных из них.
3. Выбор результирующей методики.

Нами были рассмотрены следующие группы математических аппаратов:

1. Методы классификации. Характеризуются возможностью использовать меру расстояния между объектами без необходимости выполнения операций суммирования.
2. Вероятностные модели наступления событий.
3. Методы экспертной оценки.

Среди методов классификации выделяются следующие направления, различные по целям, ставящимся перед применяемыми методами, и получаемыми результатами:

1. Кластер-анализ. При кластеризации целью является выявление классов. Фактически это многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы(кластеры)(Q-кластеризация, или Q-техника, собственно кластерный анализ). Кластер — группа элементов, характеризуемых общим свойством, главная цель кластерного анализа — нахождение групп схожих объектов в выборке. Спектр применений кластерного анализа очень широк: его используют в археологии, медицине, психологии, химии, биологии, государственном управлении, филологии, антропологии, маркетинге, социологии и других дисциплинах.
2. Дискриминантный и дисперсионный анализ. В дискриминантном анализе классы предполагаются заданными — плотностями вероятностей или обучающими выборками. Задача состоит в том, чтобы вновь поступающий объект отнести в один из этих классов. У понятия «дискриминация» имеется много синонимов: диагностика, распознавание образов с учителем, автоматическая

классификация с учителем, статистическая классификация и т. д. Фактически содержанием метода является разработка и исследование статистических методов решения следующей задачи различения (дискриминации): основываясь на результатах наблюдений, определить, какой из нескольких возможных совокупностей принадлежит объект, случайно извлеченный из одной из них. На практике задача различения возникает, напр., в тех случаях, когда наблюдение признака, полностью определяющего принадлежность объекта к той или иной совокупности, невозможно или требует чрезмерных затрат средств или времени; в случаях, когда информация о таком признаке утеряна, и ее нужно восстановить, а также, когда речь идет о предсказании будущих событий на основе имеющихся данных. Ситуации первого типа встречаются в медицинской практике, например, при установлении диагноза по комплексу неспецифических проявлений заболевания. Пример ситуации второго типа - определение пола давно умершего человека по останкам, найденным при археологических раскопках. Ситуация третьего типа возникает, например, при статистическом прогнозе отдаленных результатов лечения. Методом дискриминантного анализа является многомерный статистический анализ, служащий для количественного выражения и обработки имеющейся информации в соответствии с выбранным критерием оптимальности решения. Поставленная таким образом задача о дискриминантной функции может быть перефразирована как задача одноходового дисперсионного анализа. Можно спросить, в частности, являются ли две или более совокупности значимо отличающимися одна от другой по среднему значению какой-либо конкретной переменной. Однако должно быть ясно, что если среднее значение определенной переменной значимо различно для двух совокупностей, то можно сказать, что переменная разделяет данные совокупности.

3. Группировка. При группировке решается задача, «разбиение элементов на группы независимо от того, естественны ли границы разбиения или нет». Цель по-прежнему состоит в выявлении групп однородных объектов, сходных между собой (как в кластер-анализе), однако «соседние» группы могут не иметь резких

различий (в отличие от кластер-анализа). Границы между группами условны, не являются естественными, зависят от субъективизма исследователя. То есть группировка представляет собой способ подразделения рассматриваемой совокупности данных на однородные по изучаемым признакам группы. Обычно это делается с целью изучения структуры этой совокупности либо взаимосвязей между отдельными элементами этой совокупности. С помощью группировки можно выявить влияние отдельных единиц на средние итоговые показатели. Так, например, группировка рабочих данной организации по уровню производительности труда используется с целью выявления влияния высокой производительности труда отдельных рабочих на среднюю производительность по организации и для определения резерва, кроющегося в повышении производительности труда всех рабочих до уровня передовых рабочих.

Вероятностные модели оценки риска основываются не на понятии отнесения к определенному классу, а на понятии осуществления или неосуществления выдвинутых гипотез. Рассмотренные вероятностные модели базируются на следующем математическом аппарате:

1. Алгебра событий
2. Теоремы сложения и умножения вероятностей
3. Понятия полной группы событий, зависимых и независимых событий
4. Формула полной вероятности
5. Формулы Байеса

Все рассмотренные математические аппараты значительно различны. В пользу методов классификации указывают следующие аргументы:

1. Глубоко проработанный, современный математический аппарат, находящийся «на переднем крае» математики, что повысит лояльность специалистов к методике, основанной на методах классификации.
2. Развитые средства автоматизации, позволяющие существенно ускорить процесс создания алгоритмов оценки.

3. Наличие методик самообучения, позволяющих повысить качество оценки без привлечения экспертов. Эти методики так же доступны в специализированных математических пакетах.
4. Отсутствие необходимости в численной оценке каждого из параметров исследуемых объектов.
5. Существующий тренд в сторону применения таких алгоритмов во всех областях деятельности. Это обеспечит лояльность широкого круга потребителей, в том числе не являющихся специалистами в области математической статистики.

Однако против методов классификации есть и ряд серьёзных контраргументов:

1. Все они оперируют понятиями классов. То есть, для применения описанных методов требуется тем или иным способом выделить внутри существующих групп (аптеки, дистрибьютеры, производители, медицинские организации) классы, фактически являющиеся «категориями риска». Однако на данный момент не было выявлено никаких оснований полагать, что риск возрастает «ступенчато».
2. Задачи решаемые в рамках кластерного анализа предполагают прежде всего выявление описанных выше групп методами математической статистики. Однако практика показывает, что такое выделение гораздо эффективнее, в случае необходимости, производится с использованием экспертных оценок.
3. Методы дискриминантного анализа предполагают наличие заранее определенных экспертами классов, однако требуют подготовки соответствующих обучающих выборок. При этом несмотря на большой объём работ по подготовке данных и обучению, сам алгоритм никак не поддаётся проверке на предмет устойчивости дискриминации и отсутствие фатальных ошибок.
4. Методы группировки позволяют произвести только предварительную оценку, и результаты, полученные с использованием этого метода, должны быть подвергнут дальнейшей обработке.
5. В дополнение к недостаткам каждого из методов классификации в отдельности, существует весьма существенный общий недостаток. Ни один из методов классификации не позволяет оперировать понятием полной выборки, поскольку не содержит

математического аппарата, позволяющего учесть то, что количество исследуемых объектов конечно, и известно заранее.

Вероятностные модели существенно отличаются от методов классификации подходом к оценке риска. Преимуществами данных моделей является:

1. Учёт того, что количество оцениваемых объектов конечно, и все они заранее известны.
2. Прозрачный, проверяемый алгоритм оценки риска, имеющий в своей реализации чёткую логическую структуру.
3. Отсутствие необходимости подготовки обучающих выборок и других подобных данных.
4. Отсутствие необходимости осуществлять предварительную разбивку на классы для выполнения оценки риска.

К недостаткам вероятностных моделей следует отнести:

1. Необходимость участия экспертов в процессах оптимизации алгоритмов.
2. Невозможность автоматически «подстроить» модель, в случае появления новых категорий объектов, или других изменениях базовых требований.
3. Необходимость разработки методик хранения и обработки данных именно для описанной модели.

Несмотря на наличие развитого математического аппарата, позволяющего создать повторяемые и проверяемые методики оценки, нельзя исключать из рассмотрения экспертные методики. С учётом того, что существует развитое экспертное сообщество, определены следующие роли экспертов:

1. Выработка критериев, которые будут использованы в качестве входных параметров как для математической модели, так и для осуществления экспертной оценки на основе коэффициентов(см. ниже).
2. Создание формулы экспертной оценки на основе весовых коэффициентов. Эта формула должна быть использована на равных правах с выбранным математическим методом анализа. В качестве заключения будет выбрано то, которое показывает более высокий уровень риска.
3. Накопление базы экспертных заключений на основе анкетирования. Этот метод предназначен прежде всего для

самоанкетирования учреждений, например с использованием интернета.

Рассмотрев все описанные математические методы, их преимущества и недостатки в совокупности, а так же удобство совместного с формулой экспертной оценки использования, мы пришли в заключению, что наиболее эффективным будет выбор вероятностной модели оценки риска по сравнению с методами классификации. С учётом того, что количество оцениваемых объектов конечно, и все они заранее известны, наиболее эффективно покажет себя применение формулы полной вероятности.

Расчет вероятностей: возможности.

РАСЧЕТ	ОПИСАНИЕ
Высокая (Вероятно)	Высокая вероятность достижения благоприятного исхода в течение года, или более 75% вероятности данного события.
Средняя (Возможно)	Умеренная вероятность достижения благоприятных результатов в течение года, или 25%-75% вероятности данного события.
Низкая (Маловероятно)	Некоторая вероятность достижения благоприятного исхода в среднесрочном периоде, или менее 25% вероятности данного события.

Кроме оценки вероятности наступления события нанесения вреда, требуется выработка регламента определения графиков проверок. С этой целью предлагается провести разделение объектов по степени риска возможности происхождения негативного события. Разделение предлагается осуществлять на основе стандарта FERMA. Именно этот стандарт является самым распространенным в сфере управления рисками и содержит наиболее объёмную базу наблюдений по всем отраслям экономики. Однако данный стандарт не содержит отдельных наборов наблюдений по интересующим нас

видам деятельности. Из этого следует необходимость выделения двух этапов применения стандарта FERMA. На первом должно быть произведено выделение нескольких групп риска.

Однако в данную процедуру предлагается внести изменения с целью оптимизации графиков проверок, разделив среднюю группу на две: 25-50% и 50-75%. Таким образом, получаем следующую таблицу:

РАСЧЕТ	ОПИСАНИЕ	ЧАСТОТА ПРОВЕРОК
Высокая (Вероятно)	Высокая вероятность данного исхода (риска) в течение года, или более 75% вероятности данного события.	Раз в год
Средняя (Высоко вероятно Возможно)	Умеренная вероятность достижения данного исхода (риска) в течение года, или 50%-75% вероятности данного события.	1 раз в 3 года
Средняя (Низко вероятно Возможно)	Низкая вероятность достижения данного исхода (риска) в течение года, или 25-50% вероятности данного события.	1 раз в 6 лет
Низкая (Маловероятно)	Некоторая вероятность достижения данного исхода (риска) в среднесрочном периоде, или менее 25% вероятности данного события.	1 раз в 10 лет

Поскольку описанные частоты проверок определены на основе усредненных данных по отраслям, требуется проведение второго этапа — пересчёт частоты проверок на основе данных, накопленных по результатам проведенных проверок. Предлагается следующая методика пересчёта:

1. Формируются апостериорные данные по группам риска возникновения неблагоприятного исхода проверки за весь период наблюдений.
2. Для каждого объекта фиксируется количество неблагоприятных исходов проверки за весь период наблюдения.
3. Для каждой группы ежегодно определяется среднее время возникновения неблагоприятного исхода по формуле

$$\mu = \frac{\sum_{i=1}^N X_i}{N * T}$$

Где:

μ — среднее время возникновения неблагоприятного исхода

X_i — количество неблагоприятных исходов за период наблюдения для i -го объект наблюдений

N — общее количество объектов в дано группе

T — время от начала наблюдений по методике

4. Полученное значение среднего времени возникновения неблагоприятного исхода определяет интервал проверки. На основании полученного интервала график проверок подвергается ежегодному пересмотру, при этом отсчёт интервала каждый раз ведется от даты начала наблюдения в соответствии с описанной методикой.

Описанный подход позволяет добиться прогрессивного снижения количества проверок при условии редкого выявления нарушений, но при этом обеспечивает возможность оперативной реакции в случае изменения тенденции.
