

Актуальные вопросы применения вероятностных методов в обработке информации

К. С. Миланова

Финансовый университет при Правительстве Российской Федерации (Финуниверситет), Financial University
karinamil@mail.ru

Аннотация. В ходе развития информационных технологий одним из наиболее полных и объективных в плане оценки целостной ситуации является применение в ходе изучения какого-либо процесса, явления, структуры и состава организации, предприятия, выполнения ими каких-либо задач является методология, напрямую связанная с анализом системы – или же, другими словами, системный подход к анализу ситуации. При подобном методе оценки ситуации есть множество возможных разветвлений данного способа, но на данный момент очень популярным и эффективным методом при разработке и анализе информации о каком-либо процессе, явлении, об операциях в организациях и их подразделениях является использование вероятностных методов обработки информации.

Ключевые слова: вероятностные методы; оценка модели; алгоритмы оценивания; системный анализ; модели и моделирование

I. ВВЕДЕНИЕ

Вероятностные методы активно применяются именно для объективной оценки событий, ведь даже само понятие «вероятность» позволяет сделать вывод о том, что данные методы возможно применять в ряде наук – от наук, напрямую связанных с математикой, информатикой и информационными технологиями, до ряда естественно-научных при описании какого-либо процесса или явления. Наверняка каждый из нас слышал банальную фразу о том, что «вероятность, что сегодня выпадет снег, X%». Данные методы способны объективно, непредвзято, или же адекватно, говоря иначе, оценить всю ситуацию, дают возможность сделать вывод, провести исследование, т.к. общая картина исследования будет предоставлена на некоторое время вперед. В данном исследовании основной целью ставится изучение вероятностных методов обработки информации: как историю их развития в прошлом, так и применение современных вероятностных методов в информационных технологиях. Для достижения цели исследования поставлены следующие задачи: определить значение понятия «системный анализ», «вероятность», «вероятностные методы», «обработка информации», дать характеристику развитию данных методов с исторической точки зрения; изучить сущность обработки информации в целом; привести пример использования на практике вероятностных методов в обработке информации.

Проблемой исследования стал вопрос о том, действительно ли роль вероятностных методов в обработке

информации в условиях современных информационных технологий сохранилась и является достаточно значимой, или же есть какие-либо аналоги данных методов.

II. ПОНЯТИЯ «СИСТЕМА», «ВЕРЯТНОСТЬ» И «ВЕРЯТНОСТНЫЕ МЕТОДЫ» И ИХ ЗНАЧЕНИЕ

При изучении и комплексной оценки роли методов вероятностного анализа в обработке информации считается необходимым дать определения основным понятиям на данную тематику.

Система – это термин, который подразумевает в своей сущности категорию отражения или же форму представления материи доступными пониманию средствами; другое значение системы – способ воспроизведения и отражения континуальной целостности средствами человеческого сознания, логики (мышления). Системный анализ в современных условиях научно-технического развития понимается как синтез идей и принципов общей теории систем и кибернетики, который имеет возможности использования современной вычислительной техники, а также имеет своим предметом исследования изучение и моделирование объектов сложной природы (систем).

Термин вероятность в своей сущности отображает такое число в интервале от нуля до единицы включительно, которое представляет возможность свершения определенного события. Вероятность события принято определять таким образом, что она равна отношению количества шансов на то, что событие может произойти, к суммарному (или же общему) количеству возможных исходов, при этом подразумевается, что каждая возможность одинаково вероятна. Если обратиться к примерам, то при бросании шестистороннего кубика всего в сумме может быть шесть возможных исходов, три из которых сбросят четное число. Таким образом, вероятность того, что выпадет четное число, равна $3/6$ (или $1/2$). Событие нулевой вероятности отображает такую ситуацию, которая ни в коем случае не может произойти (бросание семерки на шестистороннем кубике), а вероятность 1 соответствует событию, которое обязательно должно произойти. При бросании геометрически правильного игрального кубика результат не зависит от предыдущих результатов предшествующих бросков. В этом случае вероятность называется независимой или же безусловной. Но гораздо чаще складывается ситуация, при которой необходимо

учитывать и условную вероятность, при которой вероятность события зависит от предыдущих исходов.

Такие методы научного познания как вероятностно-статистические методы в системном подходе подразумевают процесс учета влияния некоторого множества случайных факторов, которые можно характеризовать четко определенной, а также устойчивой частотой (определенным интервалом времени, при котором данные факторы будут повторяться). Этот факт позволяет создать закономерную необходимость (говоря иначе, закон), которая «пробивается» через суммарное общее действие множества случайностей. Названные методы опираются на теорию вероятностей, которую зачастую называют наукой о случайном.

К основной задаче теории вероятностей можно отнести выяснение закономерностей, которые возникают при определенном взаимодействии некоторого большого количества множеств случайных факторов. При исследовании вероятностных методов анализа и статистики информации при системном подходе необходимо рассмотреть различные виды данных закономерностей, которые в своей сущности по существенным признакам были поделены на две основные группы: динамические закономерности и статистические закономерности. Объединенные обобщенные вероятностно-статистические методы основаны на четком разграничении динамических и статистических законов по определенному критерию, или же основанию, которое подразумевает характер вытекающих из них предсказаний. В динамических закономерностях предсказания имеют точно определенный однозначный характер (например, в классической механике наиболее точно отражается данная особенность). Динамические законы характеризуют поведение относительно изолированных объектов, которые состоят из небольшого числа обособленных частей, в которых можно абстрагироваться от целого ряда случайных факторов.

Первая группа методов представляет собой наиболее простой и примитивный расчет средних значений случайной переменной за период наблюдения. В таком варианте IP-следователь стремится изучить только фактическую среднюю величину, и оценка достоверности расчетов не является обязательной.

Вторая группа включает методы расчета характеристик выборки случайной переменной. Необходимо получить оценку точности и надежности расчета средних значений. Для оценки точности и вероятности расчетов применяются законы распределения, что обычно вызывает определенные трудности.

Для более детального изучения вероятностных методов обработки информации необходимо изучить терминологию на данную тематику.

Множественное число результатов, полученных при проведении наблюдений (или же измерений), называется выборочным множеством или выборочной популяцией.

Число наблюдений, образующих выборку, называется размером выборки и обозначается символом (N).

Известный факт, что определенная величина полностью описывается законом ее распределения, ее характеристикой. Если закон распределения и его параметры при изучении данной величины известны, то возможно определить любую характеристику случайной переменной. Поэтому, как правило, статистическая обработка информации о результатах измерения сводится к определению вида закона распределения и его параметров.

Статистическое определение закона распределения случайной переменной всегда ассоциировалось с большими объемами вычислений, но современными средствами математических вычислений (например, Excel, Mathcad, Statistica и др.) позволяют автоматизировать и значительно облегчить процесс. Техническая сторона этого вопроса хорошо описана в специальной и учебной литературе, посвященной тестированию и статистического исследования. Основное внимание уделяется использованию пакетов, таких как Mathcad и Excel для проведения соответствующих расчетов.

Статистическая информация должна соответствовать следующим требованиям:

- единообразие;
- надежность;
- выразительный.

В некоторых случаях известна форма распределения времени безотказной работы, после чего задача сводится к оценке параметров распределения с использованием известных методов.

Наиболее часто значения доверительных вероятностей принимают равными 0,90; 0,95; 0,99 или уровни значимости соответственно 0,10; 0,05; 0,01.

В результате испытаний можно получить точечные значения оценки параметров и интервальной оценки. С интервалом оценок определяется, по какому интервалу оценок с заданной вероятностью достоверности? охватывает математическое ожидание оценивается прямой параметр. Чаще всего, вид закона случайной величины заранее не известен. Тогда необходимо проводить весь комплекс статистических расчетов по сглаживанию экспериментальных данных наиболее пригодным (приемлемым) законом распределения и подбор (вычисление) его параметров.

Рассмотрим последовательность действий при определении вида закона распределения.

При определении закона распределения целесообразно придерживаться следующего порядка: подготовка данных (ошибок измерения координат); определение выборочных характеристик; построение гистограммы ошибок; проверка допустимости гипотетического закона распределения ошибок, используя определенные критерии согласия: Пирсона, Мизеса и Колмогорова и другие.

III. СОДЕРЖАНИЕ ОБРАБОТКИ ИНФОРМАЦИИ

Содержание понятия обработки информации является довольно сложным и многогранным, оно отображает очень широкий и объемный массив действий и процессов, которые осуществляются с информацией. При рассмотрении темы сущности понятия обработки информации, необходимо в первую очередь изучить и дать характеристику существенным признакам и категориям сущности инвариантной обработки. Зачастую в данном качестве выступает смысловое содержание некоторого определенного информационного сообщения, которое имеет прямое непосредственное отношение к смыслу информации, которая содержится в сообщении. Но вопреки данной особенности также существует множество видов информации, она разнородна и довольно часто имеет куда большие масштабы, что существенно усложняет ее обработку. Увеличение темпов накопления информации, которое присуще современной мировой науке, привело к тому, что её объемы, по оценкам различных ученых, увеличились от 20 до 100 раз. Данная тенденция к увеличению накопленной информации является одним из критериев, согласно которым все больше людей сталкиваются при обработке, хранении, передаче и статистическом анализе с прогрессирующими сложными проблемами адаптации к реальной ситуации.

При изучении обработки информации эксперимент над данными проводится с определенной целью извлечения из них как можно более большого объема полезной и объективно необходимой информации для нормального функционирования системы при разработке и принятии решений, связанных с управлением организацией, фирмой или предприятием. В своей сущности любая обработка статистических данных имеет преобразование приведенных для анализа данных в более удобные для использования и обработки, или же передаче ответов изучаемой системы с языка измерений, обновленной языковой модели. В зависимости от первоначальных знаний об изучаемом объекте – каковы данные объекты исследования по своей сути, их основные черты и принципы действия, в статистических исследованиях принято деление на два вида различных по характеристикам моделей: классификационные; численные.

К классификационной модели относится такая модель, которая представляет собой самую первичную форму познания. Одним из типичных классических примеров применения классификации на практике в психической деятельности человека можно отнести распознавание окружающих объектов. В такой области человеческой деятельности как научные исследования процесс познания принято начинать с постепенного грамотного соотношения исследования объекта по существенным критериям и признакам с другими однородными объектами, а также с дальнейшего выявления сходств и различий между ними. При построении и использовании определенных классификационных моделей решаются такие задачи как: кластеризация; классификация; упорядочение объектов; уменьшение размерности объектов.

Далее приведем определения каждой из приведенных выше задач.

Термин кластеризация подразумевает собой такой процесс как поиск естественных группировок объектов. Количество классов и группировок объектов в данном случае неизвестно, как и установления границ классов в области пространственных объектов. В данном случае требуется определение их на основании близости множества сходств или же существенных различий при описании множества объектов исследования кластеризации в Протоколе наблюдения.

Термин классификация, или, говоря иначе, распознавание образов, включает в своей сущности определение принадлежности каждого из множества исследуемых объектов к той или иной группе на основе значений его предельных существенных характеристик, которые, в свою очередь, обязательно должны были быть заключены в Протоколе наблюдения. При классификационном подходе системного анализа информационного потока устанавливается некоторое определенное количество классов и список характеристик каждого из установленных классов.

Само построение объектов по какому-либо определенному критерию является частью установления порядка отношений между объектами, который был установлен при использовании множества информации, которая была зафиксирована в Протоколе наблюдений. Традиционно с помощью числовых моделей решаются задачи становления закона распределения, оценки параметра или поиска экстремума.

Установление закона распределения случайной переменной заключается в нахождении математического выражения, связывающего вероятность возникновения исследуемого объекта со значениями, зафиксированными в протоколе наблюдений. Зная распределение (например, к такому можно отнести плотность вероятностей), можно получить ответы на ряд следующих вопросов: на каком промежутке расположены возможно вероятные случайные значения; при каком значении идет процесс рассеивания реализующих ее значений; параметр поиска положения распределительных установок, например, среднее, моду или медиану; какая связь между различными реализациями случайной величины. Определение связи производится при вычислении с заданной мерой зависимости; насколько широко рассеяны эти значения (находится параметр шкалы-дисперсия или стандартное отклонение).

Законы распределения отношения между переменными могут быть заданы как параметризованной формой (то есть функция с конечным числом переменных), так и не параметризованной формой (то есть в виде функционала).

Снижение размерности модели осуществляется путем выбора наиболее информативных особенностей или при помощи соединения нескольких знаков в один, для наиболее высокого уровня упрощения модели. Также данная процедура необходима не только для упрощения работы с моделью информационного массива, так и для

наиболее высокой степени прозрачности и точности моделирования. В данном случае смысл данного действия при снижении размерности модели заключается в том, что, согласно основным принятым правилам действия при работе с такого рода моделированием, любая первоначальная классификационная модель является не параметризованной, и в своей сущности идет учет многих предположений, которые, зачастую, крайне необходимо проверить на практическом применении ввиду неполного отображения реальной действительности. Совокупность определенных характеристик объекта исследуемого массива информационных потоков формируется эвристическим методом, что подразумевает собой наиболее полный анализ при формировании существенных признаков и характеристик, что приводит к негативным последствиям: чрезмерные объемы приведенной информации, избыточность данных процессов. Таким образом, мы подходим к выводу о том, что одной из наиболее значимых, основных, ключевых задач по совершенствованию классификационной модели является задача уменьшения размера данного вида модели.

Численные модели, в свою очередь, отличаются от множества классификационных моделей наличием такой особенности как измерение базовых характеристик объекта исследования в числовых масштабах, а другой вид характеристик – интеграционные, или же критерии, рассчитываются при использовании в качестве основы функции или некоторого количества функций элементарных переменных, которые являются в данном случае начальными характеристиками.

Более того, следует отметить тот факт, что численные модели чаще учитывают изменение переменных со временем, то есть измерение различных численных характеристик может проводиться в разные моменты временного промежутка, а не на каком-то определенном интервале. В связи с этим измерение может относиться к большому количеству объектов с одинаковыми объектами в разные моменты времени.

Одной из ключевых задач численных моделей, в отличие от классификационных, является решение задачи формирования закона распределения оценки параметров или же поиска экстремума.

Оценка параметров подразумевает процессы, направленные на определение значения какого-либо параметра в соответствии с Протоколом наблюдений. Если значение данного параметра необходимо для оценки долгосрочного процесса, то такая задача называется прогнозированием.

Поиск экстремума при пошаговом измерении процесса изменения наблюдаемых величин крайне необходим для обеспечения исследования экстремальных значений целевой характеристики функции или же функцию критерия.

Экономический аспект является весьма нетривиальной проблемой. С одной стороны, нас интересуют особенности возникающих свойств и прямого измерения быть не может, и очень часто измеряемые значения носят

недетерминированный (случайный) характер, что требует специальной методологии статистических наблюдений и измерений.

Если измерение возникающих свойств в большей степени творческое и реализуется косвенным измерением характеристик системы, то измерения недетерминированных факторов очень хорошо установлены и относятся к области инженерных расчетов. Локальное обучение Ограниченные выборки позволяют выявить ключевые закономерности и распространить их на все население. Статистические методы исследования случайных неопределенностей основаны на предположении о существовании распределения вероятности случайной переменной, хотя и неизвестной. Задача заключается в наблюдении случайной переменной и устранении неопределенности ее распределения с целью уточнения модели исследуемой системы на основе результатов наблюдений.

IV. ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ ВЕРОЯТНОСТНЫХ МЕТОДОВ В ОБРАБОТКЕ ИНФОРМАЦИИ

При аудите информационной безопасности одной из важных мер является анализ угроз и уязвимостей информационной безопасности. Угроза информационной безопасности – совокупность условий и факторов, создающих опасность нарушения информационной безопасности. Уязвимость – это недостаток или слабость информационной системы предприятия, программно-аппаратных комплексов и других инструментов, которые могут быть использованы для реализации-это угрозы.

Независимо от выбранной методики, аудиторы, используя количественные системы (CORAS, CRAMM) или качественные оценка рисков информационной безопасности, и понимание уязвимостей и угроз должны быть четко систематизированы и структурированы.

Если для методов, целью которых является определение соответствия информационных систем предприятия какому-либо международному стандарту информационной безопасности, данный этап аудита имеет среднее значение, то методы, использующие количественную и качественную оценку риска информационной безопасности, а также построение теоретико-вероятностной модели оценки риска информационной безопасности для конкретной компании, этот этап играет ключевую роль.

Аудит проводится в несколько этапов. На ранних этапах идентифицируются бизнес-процессы и связанные с ними информационные активы. Возле каждого информационного актива или группы активов определяется перечень угроз конфиденциальности, целостности и доступности. На основе модели информационной безопасности, является основой СУИБ предприятия. На следующих этапах проводится вероятностная оценка риска внедрения угроз с помощью экспертных оценок, а затем качественной и количественной оценки риска, после чего определяется приемлемость риска и время-работают рекомендации и

стратегии по нейтрализации угроз информационной безопасности.

В процессе разработки собственных методов аудита были построены вероятностно-теоретические и игровые модели оценки рисков и выявлена необходимость использования системного анализа для классификации и выявления угроз, структуры дерева целей функций Департамента информационной безопасности и процесса аналитической иерархии (АГП) на завершающих этапах аудита для определения оптимальной стратегии для МБ.

Для построения теоретических вероятностных моделей было использовано определение риска как продукта вероятности потери и, как следствие, риск информационной безопасности определяется как продукт финансовых потерь (убытков), связанных с инцидентами безопасности, и вероятность их реализации. Формулы расчетов риска представляют собой формулы общей вероятности совместных зависимых и независимых событий.

Когда используется игровой подход антагонистическая модель с нулевой суммой, в которой первый игрок – предприятие, второй игрок – нападающая сторона. Особенностью данной модели является сочетание стратегических игр между двумя людьми и с природой с точки зрения риска, в которых обычно используется критерий Байеса. Это происходит потому, что, если субъект (игрок 1) получает целевую стратегию выбора, злоумышленник (игрок 2) помещает атаку актива случайным образом из средней вероятности. Результатом решения данной игры должны стать рекомендации по выбору стратегии распределения бюджета. Метод МАИ также использовался для определения наилучшей стратегии инвестирования.

В данном случае выбор решения и поставщика программного обеспечения для внедрения систем информационной безопасности. Используя инструменты системного анализа было построено дерево целей и функций отдела информационной безопасности предприятия. В результате влияния внешней среды, внутренних воздействий, материальных и информационных потоков, а также целевой функции системы.

Слияние этих инструментов в единую методологию позволяет получить подробные результаты по многим показателям и разносторонний подход к проведению аудита безопасности корпоративных информационных систем и принятию решений при выборе стратегий. Важную роль играют методы системного анализа и теории систем, являющейся основой в этапы развития аудита и СУИБ в целом.

При изучении вероятностных методов обработки информации формулируется вывод о том, что роль вероятностных методов в современных условиях развития информационных технологий достаточно велика. Данные

методы применяются при обработке информации в различных науках, начиная от математических и заканчивая естественно-научными.

При изучении информации с помощью вероятностных методов данные преобразуются в удобный для обработки вид и позволяют наглядно прийти к определенным выводам. Особенно важное значение вероятностные методы обработки информации играют в экономических дисциплинах. На примере аудита информационной безопасности показано практическое значение вероятностных методов в данной дисциплине.

Таким образом, вероятностные методы обработки информации являются одной из основных моделей для работы с большим массивом данных. Такие методы широко распространены и применяются в совокупности с другими методами системного подхода и математической науки (к первым относятся методы исследования операций, исследование безопасности и эффективности функционирования систем, ко вторым, например, методы математической логики), и с другими методами научного познания – анализ, синтез, сравнение, дедукция и индукция.

Такой системный, объективный подход позволяет снизить вероятность возможных ошибок и просчетов к минимальному количеству и позволяет составить полную модель ситуации на настоящий момент времени и сделать прогноз на будущее, а роль вероятностных методов в таких условиях является существенной, и знание основ данного подхода крайне необходимо для дальнейшей работы аналитика.

СПИСОК ЛИТЕРАТУРЫ

- [1] Беляева М.А. Системный анализ технологий и бизнес-процессов в производстве: монография / М. А. Беляева; РЭУ им. Г.В. Плеханова, 2015. с. 321-353.
- [2] Вдовин В.М. Теория систем и системный анализ: Учеб. для бакалавров / В.М. Вдовин, Л.Е. Суркова и др. М.: Дашков и К, 2016. 644 с.
- [3] Комарцова Л.Г., Максимов А.В. Нейрокомпьютеры: Учеб. пособие для вузов. 2-е изд., перераб. и доп. М.: Изд-во МГТУ им. Н.Э. Баумана, 2004. 400 с.
- [4] Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. И.Д. Рудинского. М.: Горячая линия -Телеком, 2006. 452 с.
- [5] Звягин Л.С. Современные проблемы динамики социально-экономических систем в фокусе производственных отношений и управляющего развития // Теоретическая экономика. 2018. № 4 (46). С. 76-81.
- [6] Звягин Л.С. Комплексная оценка безопасности функционирования моделей экономических систем // Экономика и управление: проблемы, решения. 2017. Т. 4. № 1. С. 18-25.
- [7] Звягин Л.С. Системы распределенного моделирования и методы управления модельным временем // Экономика и управление: проблемы, решения. 2018. Т. 1. № 10. С. 4-11.
- [8] Орлов А.И.: «О перестройке статистической науки и её применений» // Журнал "Вестник статистики", 1990. No.1. С.65-71.