

МЕТОД ОЦЕНКИ ДОСТОВЕРНОСТИ КЛАССИФИКАЦИИ ИНФОКОММУНИКАЦИОННЫХ УСЛУГ

Павлов Семен

Аннотация: В работе определена область исследования, а именно – инфокоммуникационные услуги, которые предоставляет телекоммуникационная сеть. Проанализирована необходимость классификации услуг, которая проводилась с применением методов прикладной статистики. В качестве исходных данных для классификации взяты параметры предоставления услуг. Выделенные классы услуг формируют требования к сети и позволяют определить, какие характеристики сети, и какое оборудование необходимо для предоставления тех или иных услуг. Рассмотрен метод оценки достоверности результата классификации инфокоммуникационных услуг методами прикладной статистики.

Ключевые слова: инфокоммуникационные услуги – достоверность классификации – факторный анализ – кластерный анализ.

Ключевые слова классификации АСМ: C.2.0 General: Data communications.

Введение

Стремительное развитие научно-технического прогресса в области телекоммуникаций и информатики обеспечивает в настоящее время возможность предоставления обширного спектра инфокоммуникационных услуг широкому кругу пользователей. Номенклатура услуг постоянно расширяется, что затрудняет планирование развития телекоммуникационной сети для предоставления этих услуг, так как сеть должна выполнять требования каждой из предоставляемых услуг, а учесть совокупность всех этих требований – сложная задача, требующая применения средств вычислительной техники. Поэтому появилась необходимость выполнить классификацию инфокоммуникационных услуг. Для классификации выбраны следующие методы прикладной статистики: факторный и кластерный анализ.

Выделение классов услуг требуется для формирования требований к телекоммуникационной сети и сетевому оборудованию и для упорядочивания услуг. Каждая инфокоммуникационная услуга характеризуется параметрами, необходимыми для ее предоставления (например, информационная скорость доступа, коэффициент ошибок и другие), которые должны выполняться сетью. С помощью классов услуг можно определить требования к сетям для группы предоставляемых услуг, что снижает трудозатраты и повышает точность планирования и модернизации телекоммуникационной сети.

В процессе классификации возникают различного рода допущения и неточности вычислений и измерений, которые влияют на конечный результат. Поэтому возникла необходимость оценки достоверности процесса и результата классификации.

Основная часть

Под достоверностью процесса классификации будем понимать ту вычислительную погрешность, которая возникает в результате применения вычислительных методов, в частности, при применении факторного анализа. Вычислительная погрешность учтена в инструментарии, который применяется для проведения факторного и кластерного анализа [StatSoft, 2011]. Под достоверностью результата классификации будем понимать степень достоверности результата примененных статистических методов.

В результате классификации формируются неформализованные модели, дающие интегральное представление о классах подобных объектов. Неформализованные модели не идентичны рассматриваемым объектам, т.к. несут информацию о множестве подобных объектов, входящих в классы, и являются их осредненными образами. Поэтому формируемые классы услуг не будут с точностью характеризовать каждую инфокоммуникационную услугу. Каждую новую услугу можно отнести к такой неформализованной модели [Горбань, 2007].

Однако набор и характеристики инфокоммуникационных услуг постоянно меняются, меняются и их модели. Изменение моделей обусловлено не только изменением самих услуг, но и изменением критериев классификации. Существенную роль при выборе критерия играют приобретенный ранее опыт, окружающая среда и многие другие факторы (например, демографические, технические, социальные и другие).

Телекоммуникационные сети различаются между собой, различен и подход к классификации услуг (можно учесть ряд факторов, либо отсечь их), поэтому для одних и тех же услуг разными оказываются неформализованные модели.

Каждую из инфокоммуникационных услуг можно оценить с помощью показателей их предоставления, которые тоже можно рассматривать как объекты. Отличие оценки от реальных условий предоставления вызвано, во-первых, несовершенством технических средств (оборудования, среды передачи и прочих), а во-вторых, наличием общих решений в формировании требований к предоставлению услуг.

Подход, предложенный в работе автора [Павлов, 2009], позволяет не только выполнить классификацию уже существующих услуг, но и включать в сформированные классы любую новую услугу, которая может появиться в будущем. Конечно, нельзя исключить возможности добавления новых классов в будущем, однако в силу полноты выполненного исследования с достаточно большой долей достоверности можно утверждать, что такая необходимость может возникнуть очень нескоро, если не произойдет скачкообразного (революционного) изменения существующих информационных технологий.

При анализе результатов экспериментов, основными являются два понятия: характер распределения и объем выборки. Точность полученного результата характеризуется размером исходной выборки. При малых выборках (рисунок 1, а) невозможно гарантировать точность полученных результатов, так как невозможно предугадать большинство оценок любого из параметров инфокоммуникационных услуг. При больших размерах выборки (рисунок 1, б) увеличивается точность вычисления и достоверность результата. Физическая модель, построенная на основании оценок большой выборки, более точной; а построенная на ее основании математическая модель более близка к теоретическим результатам.

Учитывая, что объем выборки, достаточный для обоснованности выводов, сделанных на основании ее анализа, определяется в 300 объектов [Коваленко, 1982], можно сделать вывод, что результаты исследований являются достоверными. Если заранее неизвестен характер распределения первичных признаков объектов в этой выборке, то нельзя применять классические методы математической статистики [Коваленко, 1982], поэтому в данном случае необходимо выполнить нормализацию первичных признаков, что в результате допускает применение кластерного и факторного анализа.

Кластерный анализ, как вычислительный метод, не имеет собственной вычислительной погрешности, так как он не использует ни итеративную процедуру вычислений, ни какой-либо приближительной процедуры вычислений, а использует только пересчеты матрицы расстояний, которые представляют собой конечные формулы. Поэтому проверкой численной точности этого метода и достоверности результатов, полученных по этому методу, является то, что результаты получены с помощью подсчета трех различных формул расстояний (евклидова расстояния, расстояния Чебышева и манхэттенского расстояния). Совпадения результатов при подсчете тремя разными способами матрицы расстояний говорят о достоверности результата кластерного анализа.

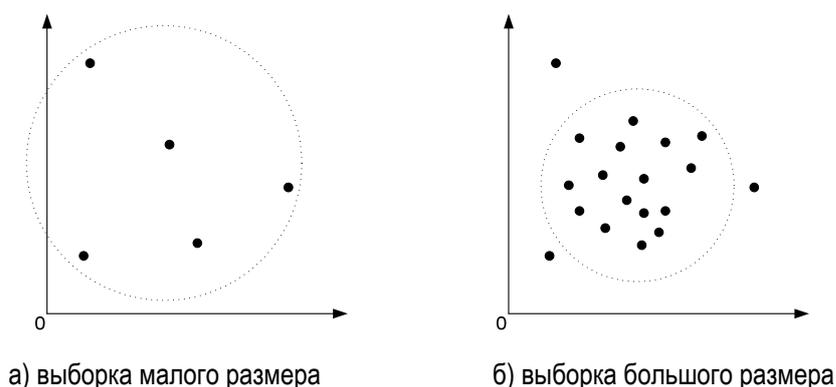


Рисунок 1 – Примеры выборки различного размера

В факторном анализе используются вычислительные процедуры (например, обращение матрицы), и поэтому численная погрешность процедуры факторного анализа определяется численной погрешностью методов вычислительной алгебры. Поскольку факторный анализ использует процедуры обращений матрицы, вся погрешность будет обосновываться числом обусловленности матрицы. Если число обусловленности матрицы существенно отличается от нуля, то результаты достоверны.

Заключение

Таким образом, выполняя классификацию инфокоммуникационных услуг методами непараметрической статистики, необходимо проанализировать погрешность процесса и результата классификации. Иначе, полученные классы услуг могут быть недостоверными, и требования к сетям и сетевому оборудованию, выдвигаемые этими классами, могут быть неверными, что может отразиться на функционировании телекоммуникационной сети в целом при предоставлении услуг.

Благодарности

Настоящая работа была выполнена при поддержке интернационального проекта ITHEA XXI Института информационных теорий и их приложений FOI ITHEA и Ассоциации ADUIS Украина (Ассоциация разработчиков и пользователей интеллектуальных систем).

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Библиография

- [StatSoft, 2011] "Официальный сайт StatSoft", электронный учебник по статистике и пакет программ для работы со статистикой, Internet-ресурс: <http://www.statsoft.ru/>.
- [Горбань, 2007] Горбань И. И. Теория гиперслучайных явлений – К.: Национальная академия наук, 2007. – 184с.
- [Павлов, 2009] Павлов С.В. Пошаговое выделение классов и классификационных формул инфокоммуникационных услуг / С.В. Павлов // Науковий збірник ДонНТУ. – 2009. – №147.– С. 68-71.
- [Коваленко, 1982] Коваленко И.Н., Филипова А.А. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1982. – 260 с.

Информация об авторах



Павлов Семен – ассистент кафедры информационно-коммуникационных технологий Одесской государственной академии холода; почта: 65082, Дворянская 1/3, Одесса, Украина; e-mail: semyen@rambler.ru

Главные области научного исследования: информационно-коммуникационные технологии, ВЭБ-технологии.