

ОСНОВНЫЕ ПОДХОДЫ ПРИ АНАЛИЗЕ СОЦИОЛОГИЧЕСКИХ ДАННЫХ

Е.А. Барановская, А.Б. Иващенко

Донецкий национальный технический университет

Предметом статті є комп'ютерна соціологія. Методи та засоби комп'ютерної соціології можуть ефективно використовуватися для розробки рекомендацій з управління різноманітними соціальними явищами, процесами, соціальними системами з використанням комп'ютерних моделей, які призначені для проведення імітаційного комп'ютерного моделювання.

К настоящему времени имеется значительный опыт формализации классических социологических теорий; хорошо изучены типы формализации, типы измерения; изучены познавательные возможности основных формальных метамоделей (позитивистская, инструментальная, интегративная); с помощью математических методов, разработаны новые подходы и ряд новых математических методов для решения данных проблем.

С помощью компьютерных имитационных экспериментов можно быстрее и проще решить ряд теоретических задач, по сравнению с использованием других математических методов. С использованием компьютерного имитационного моделирования осуществлено подтверждение некоторых положений классических и современных социологических теорий. Компьютерное имитационное моделирование позволило проверить некоторые положения социологических теорий (Г. Зиммель, Р. Мертон, Р. Уильямс) о динамике норм в индустриальных обществах.

В рамках данной тенденции, на основе ранее созданных формальных теорий и моделей социальных сетей, сетевых индексов и методов сетевого анализа, за последние годы было разработано множество компьютерных систем для анализа и моделирования структуры и динамики социальных сетей, например, AGD, Agna, ANTHROPAC, Apache Agora, BLANCHE, Carter's Archive of S Routines, da Vinci, Doug White's software, DyNet, Ecosystem Network Analysis, Egonet, EigTool, FATCAT, gem3Ddraw, GLAD, GRADAP, GraphEd, Graphlet, GraphPlot, Graph Viz, IKNOW, InFlow, JUNG, KeyPlayer, KliquesFinder, KrackPlot, MatView, MDLogix, MultiNet, NEGOPY, NetForm, NetMiner, NETVIZ, NetVis, Noldus, PREPSTAR,

PSPAR, Pajek, PermNet, ReferralWeb, SIENA, SNAPS, SocioMetrica LinkAlyzer, SNAFU, SocNetV, Sparse Matrix Package, StOCNET, UCINET, Visone, yFiles, ZO. Здесь ясно прослеживается связь между разработкой формальных теорий, включающих теоретически обоснованные сетевые индексы и методы анализа, реализацией данных теорий в компьютерных системах и широким использованием разработанных компьютерных систем для анализа эмпирических данных.

Анализ литературных источников и Интернет-ресурсов позволяет выделить следующие основные группы методов, используемых при анализе социологических данных.

1. Системы анализа нечисловой информации

Система анализа нечисловой информации предназначена для обработки данных, измеренных в разнородных шкалах: номинальной, порядковой, интервальной и количественной. В первую очередь она ориентирована на социально-экономические задачи, но может использоваться в биологии, медицине, других областях, где приходится иметь дело с нечисловой информацией.

Система позволяет одновременно обрабатывать до 320 признаков (вопросов); число объектов (респондентов) не должно превосходить 32 000 для категоризованного признака и 8 000 – для числового. Однако приведенные ограничения, имеют чисто технический характер, и допустимый объем данных зависит от используемого метода и типа компьютера.

Методы, используемые САНИ, распадаются на три группы. Первая – реализует возможности базы данных, вторая – объединяет средства предварительного анализа, позволяющие сформировать гипотезы о структуре данных, выявить «выбросы». При этом эффективно применяются графические возможности современной вычислительной техники. Методы, входящие в третью группу, используют вероятностные предположения о природе данных и позволяют проверять различные гипотезы. Особое внимание уделено анализу нечисловых признаков, изменяющихся во времени.

2. Факторный анализ

Факторный анализ это процедура, с помощью которой большое число переменных, относящихся к имеющимся наблюдениям, сводит к меньшему количеству независимых влияющих величин, называемых факторами. При этом в один фактор объединяются переменные, сильно коррелирующие между собой. Переменные из разных факторов слабо коррелируют между собой.

Таким образом, целью факторного анализа является нахождение таких комплексных факторов, которые как можно более полно объясняют наблюдаемые связи между переменными, имеющимися в наличии.

Факторный анализ позволяет решить две важные проблемы исследователя: описать объект измерения всесторонне и, в тоже время, компактно. С помощью факторного анализа возможно выявление скрытых переменных факторов, отвечающих за наличие линейных статистических связей корреляций между наблюдаемыми переменными.

Можно выделить 2 цели факторного анализа:

- определение взаимосвязей между переменными, их классификация;
- сокращение числа переменных.

3. Кластерный анализ для типологизации признаков

Типологический анализ (ТА) – один из распространенных методов решения задач, возникающих при анализе социально-экономической информации. Примером может служить выделение социально значимых, качественно отличных друг от друга, внутренне однородных групп процессов или явлений. Речь идет либо о попытке построения некоторой типологии, либо о проверке гипотезы о существовании предполагаемой типологии.

Как известно, первичная информация, получаемая в результате опросов, чаще всего представляется в виде таблицы типа «объект – признак», где объектами являются респонденты (вернее, заполненные ими анкеты), а признаками – их мнения по различным вопросам анкеты. Каждый признак порождает на множестве анкет разбиение респондентов по числу градаций данного признака.

Поскольку большинство признаков порознь являются малоинформативными, возникает необходимость создания на их основе обобщающего качественного признака. Классификация по его значениям должна выявлять группы респондентов, однородные по большому числу исходных признаков. Построение таких обобщенных признаков, как функций от исходных – важный этап типологического анализа анкет.

4. Методика долгосрочного упреждающего анализа данных

В формах конкретизации абстрактного понятия "предвидение" целесообразно развести два конкретных производных понятия: "предугадывание" и "прогнозирование". Оба заключают в себе третье конкретное понятие – "предсказание" (состояния явления или процесса в будущем). Но в первом случае предсказание, безусловно,

оно характеризуется глаголами "будет" или "станет". А во втором – сугубо условно, инструментально: "может быть или стать при определенных условиях", на которых и сосредоточивается внимание исследователя.

Типичные примеры предугадывания (которое, как правило, выступает под именем прогнозирования): кто победит на выборах или в матче, каков будет курс доллара и т.п. Кстати, именно "прогнозы" подобного типа только и интересуют пока что заказчика, поскольку иные подходы ему просто неведомы. Именно на них ориентирована работа практически всех "центров анализа и прогноза" у нас, почти всех за рубежом. Оценка такого рода прогнозов производится "по степени оправдываемости", которая, в свою очередь, располагается по шкале: сбылось – не сбылось.

5. Метод группового учета аргументов

Метод группового учета аргументов (МГУА) разрабатывается академиком НАН Украины А.Г. Ивахненко и его школой, и является типичным методом индуктивного моделирования и одним из наиболее эффективных методов структурно-параметрической идентификации сложных объектов, процессов и систем по данным наблюдений в условиях неполноты информации.

МГУА обладает определенным разнообразием возможностей на всех этапах процесса моделирования по сравнению с другими методами построения моделей. Это касается, прежде всего, генераторов моделей и применяемых критериев качества структур, а также классов моделей (базисных функций). Метод отличается активным применением принципов автоматической генерации вариантов, последовательной селекции моделей и внешних критериев для построения моделей оптимальной сложности. Он имеет оригинальную многорядную процедуру автоматической генерации структур моделей, имитирующую процесс биологической селекции с попарным учетом последовательных признаков. Такая процедура в современной терминологии называется полиномиальной нейронной сетью, причем ее структура является явной и строится автоматически, в режиме самоорганизации.

Для сравнения и выбора лучших моделей применяются внешние критерии, основанные на разделении выборки на две и более частей, причем оценивание параметров и проверка качества моделей выполняется на разных подвыборках. Это позволяет обойтись без обременительных априорных предположений, поскольку разделение выборки позволяет неявно (автоматически) учесть различные виды априорной неопределенности при построении модели. МГУА

обладает преимуществом при малых выборках данных за счет выбора сложности модели, оптимально учитывающей информативность данных.

Эффективность метода многократно подтверждалась решением множества конкретных задач из областей экологии, экономики, гидрометеорологии и т.д.

Выводы

Представленный обзор дает основание полагать, что компьютерная социология способна принести значительную научную и практическую пользу как социологии, так и самому обществу. И, естественно, дальнейшее исследование, разработка и развитие методов компьютерного социологического анализа представляет интерес.

Библиографический список

1. А.А. Давыдов. Компьютерные технологии для социологии (обзор зарубежного опыта).
2. И.В. Бестужев - Лада. Методика долгосрочного упреждающего анализа данных в технологическом прогнозировании.
3. В.А. Бритвихин, Ф.А. Красина, С.Н. Симонцев. Использование кластерного анализа для типологизации признаков.
4. М.И. Жабский. Идеология выборочных исследований.