

Заключение

В качестве заключения сформулируем рекомендации, основанные на результатах проведенного методического эксперимента по использованию EM-алгоритма и регрессионного моделирования для реконструкции пропусков в данных.

По параметрам изменения средних значений, количества верных подстановок, потерь дисперсии и качества логистических моделей между долей исходных пропусков и качеством импутирования наблюдается обратная взаимосвязь: с ростом количества пропусков точность импутирования снижается. Этот вывод вполне логичен, если принять во внимание, что при увеличении пропусков в данных уменьшается совокупность полных наблюдений, которая используется в качестве данных для прогнозирования отсутствующих значений. Получается, что необходимо предсказывать больше пропусков, используя для этого меньше данных, и сохранить при этом прежний уровень точности оценок не возможно.

В случае количественных переменных, характеризующихся большой дисперсией значений и при увеличении доли пропусков в данных EM-алгоритм является более эффективным методом импутирования, чем регрессионное моделирование, так как дает меньшее отклонение от известных значений, по сравнению с регрессионным моделированием.

В случае однородной по некоторой количественной характеристике совокупности складывается неопределенная ситуация. Здесь изменения мер средней тенденции, рассчитанных на импутированных разными методами данных разнонаправлены: регрессионное заполнение пропусков приводит к занижению среднего значения, а EM-алгоритм — к его переоценке.

Что касается показателей разброса в данных, то оба метода даже при минимальном количестве исходных пропусков занижают дисперсию значений переменных.

Если нужно восстановить значения категориальной переменной (порядковой или номинальной), регрессионное моделирование предпочтительнее для сохранения исходной дисперсии: в среднем теряется на 10 % меньше дисперсии, чем при EM-импутировании, а прирост верных подстановок оказывается даже выше.

Если необходимо оценить степень однородности исследуемой совокупности и пропусков в данных немного, следует ограничиться анализом полных наблюдений, так как импутирование чревато существенным занижением оценок.

Многоуровневый ковариационный анализ эффекта этничности в социальных сетях школьников

Иванюшина Валерия Александровна,
Александров Даниил Александрович,
СПб. филиал НИУ ВШЭ

Настоящая работа является частью долгосрочного проекта Научно-учебной лаборатории социологии образования и науки проекту «Дети-мигранты в российских школах»¹. Этот проект, начатый в 2007 году, сочетает качественные и количественные методы: массовые опросы школьников увязаны с детальным исследованием отдельных школ (по методу «кейс-стади») и глубинными интервью с учениками, учителями, администраторами и родителями. Географически наше исследование охватывает Московскую область и Санкт-Петербург. Эти регионы были выбраны потому, что они являются основными центрами притяжения трудовых мигрантов (как внутренних, так и трансграничных) в Российскую Федерацию. Согласно данным Госкомстата, из всех въезжающих в РФ трансграничных мигрантов в Московскую область прибывает 26 %, в Санкт-Петербург — 12 %. Таким образом, именно в этих городах можно ожидать обнаружить школы с большим количеством детей из семей мигрантов.

В России, в отличие от ряда стран Западной Европы, рост числа детей из семей мигрантов в школах еще только начинается, однако уже сейчас в урбанизированных районах РФ есть школы, где доля иноэтничных мигрантов достигает 40 %. Изучение детей из семей мигран-

¹ Александров Д.А., Баранова В.В., Иванюшина В.А. Дети из семей мигрантов в школах Санкт-Петербурга: предварительные данные. СПб.: Изд-во СПб. Политехнического Университета, 2010.

тов актуально и в прикладном отношении — для принятия решений в системе образования, и в теоретическом — сейчас есть возможность поймать момент становления новых сообществ, который в европейских странах был упущен исследователями несколько десятков лет назад.

Одним из фокусов нашего исследования является изучение адаптации и интеграции иноэтничных детей в российской школе. Традиционно для измерения адаптации и интеграции мигрантов использовались их школьные успехи и образовательные притязания, а также анализ аттитудов — как самих мигрантов, так и принимающего сообщества. В настоящее время, как в исследованиях школы, так и в миграционных исследованиях наряду с классическими статистическими методами широко применяются современные методы анализа социальных сетей¹. Чтобы оценить позицию мигрантов на основе сравнения большого числа сетей, необходим многоуровневый анализ.

Планирование исследований по этой теме затруднено отсутствием базовой статистики о количестве детей из семей мигрантов в российских школах. Официально такая статистика не собирается, а инициативно собираемые данные не отличаются надежностью и сопоставимостью, поскольку, в первую очередь, нет ясности в том, кого следует считать мигрантами и по каким показателям собирать такую статистику.

Предлагаемая статья основана на обработке данных опроса школьников восьмых–десятых классов, проведенного весной 2010 г. в Санкт-Петербурге по стратифицированной выборке. Генеральной совокупностью в нашем исследовании являются государственные общеобразовательные школы Санкт-Петербурга (598 школ). Частные школы, коррекционные школы, начальные школы, интернаты, кадетские корпуса были исключены из анализа.

При конструировании выборки перед нами стояла двоякая задача: получить репрезентативные данные и найти классы с достаточно высокой долей мигрантов, чтобы изучить сетевые эффекты в полиэтничных классах.

Выборка строилась в два этапа. На первом этапе из генеральной совокупности случайным образом были отобраны 30 «обычных школ» и 10 «школ повышенного статуса». На втором этапе была осуществлена довыборка малочисленных школ, т.к. из наших предварительных исследований мы знали, что именно в этих школах концентрируются дети из семей иноэтничных мигрантов. Всего в выборку попали 104 школы из всех 18 районов Санкт-Петербурга. Было опрошено 419 классов (8–10 классы) и собрано 7300 анкет.

Для опросов школьников нами разработан оригинальный инструмент, позволяющий оценивать отношение ученика к школе и образованию, его ощущение включенности в школьную жизнь. В анкету также включены вопросы на сетевые отношения в классе, позволяющие измерять социальную ин/экслюзию непосредственно, а не только путем анализа мнений. Важно подчеркнуть, что опросы проводятся во всем классе целиком, что позволяет осуществлять сравнение между детьми разной этничности и разной миграционной истории.

При разработке анкеты были использованы статьи американских, канадских, британских исследователей, а также опросные инструменты масштабных международных кросс-секционных и лонгитюдных исследований образования и миграции: National Education Longitudinal Study (NELS), США, National Longitudinal Survey of Youth (NLSY), США, Youth in Transition Survey (YITS), Канада, Youth Cohort Study (YCS), Великобритания, Children of Immigrants Longitudinal Study (CILS), США, Programme for International Student Assessment (PISA).

Анкета включала подробный социально-демографический блок, вопросы о материальном положении семьи; вопросы о культурном капитале; вопросы о текущих школьных успехах ученика; об образовательных планах; о выборе будущей профессии. Блок «Миграционная и языковая история» включал вопросы о времени переезда семьи в Петербург; о месте рожде-

¹ Luken V., Tranmer M. Personal support networks of immigrants to Spain: A multilevel analysis // *Social Networks*. 2010. Vol. 32. P. 253–262; Vermeij L. et al. Ethnic segregation in context: Social discrimination among native Dutch pupils and their ethnic minority classmates // *Social Networks*. 2009. Vol. 31. P. 230–239; Baerveldt C. et al. Ethnic boundaries and personal choice: Assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils' networks // *Social Networks*. 2004. Vol. 26. P. 55–74.

ния родителей; о родном языке матери, отца, и самого ученика; о том, на каком языке говорят дома. Вопросы о родном языке помогают устанавливать этничность респондента в тех случаях, когда он не отвечает на вопрос о национальности. В вопросе о времени переезда семьи в Петербург (Подмосковье) предлагались варианты ответа, разбитые на периоды. Это важно, т.к. длительность проживания в принимающей стране является важным фактором, наряду с этнической принадлежностью, влияющим на адаптацию.

Для оценки социально-экономического статуса семьи в анкету были включены открытые вопросы о трудовой занятости отца и матери школьника. Требовалось не только назвать профессию, но и кратко описать, чем они занимаются на своей работе. Этот инструмент позволяет достаточно точно переводить ответы школьников о работе родителей в код ISCO-88 (International Standard Classification of Occupations), который далее перекодировался в шкалу социально-экономического статуса ISEI (International Socio-Economic Index).

Для оценки про/антишкольных настроений оценки вовлеченности учеников в школьную жизнь (sense of belonging) и самооценки популярности в классе в анкету было включено несколько блоков так называемых «аттitudных вопросов», которые оценивались по 4-балльной шкале.

В анкету также входили «сетевые» вопросы для изучения межличностных и межгрупповых отношений в классе. Вопрос звучал следующим образом: «Назови, с кем ты больше всего общаешься в классе»; «Назови, с кем ты меньше всего общаешься в классе». В том и в другом случае можно было выбрать до 10 человек.

Обработка сетевых данных требует специальных методов анализа, отличных от классических статистических процедур, поскольку многие предположения, на которых основана классическая статистика, неприменимы к сетевым данным, в частности, предпосылка о независимости наблюдений (так называемая проблема сетевой автокорреляции). Сетевые данные по определению взаимозависимы. В настоящее время разработано несколько подходов для решения проблемы сетевой автокорреляции. Одним из таких подходов является метод, при котором путем компьютерного моделирования генерируется множество сетей, сохраняющих заданные характеристики исходной сети, и подсчитывается вероятность случайного возникновения того или иного изучаемого эффекта.

Такие алгоритмы стали возможными при вычислительных возможностях современных компьютеров; отметим, что расчет не самой сложной сетевой модели для 150 акторов занимает несколько часов, а с увеличением числа акторов или сетей время вычислений возрастает многократно.

Для изучения сегрегации и социального исключения мы применяли два подхода. Во-первых, использовали социально-психологическую перспективу через изучение мнений и аттitudов: детям задавали вопросы на «чувство принадлежности», на самооценку популярности. В соответствии с ожиданиями, в ситуации успешной интеграции, то есть при отсутствии социальной эксклюзии, самоощущения у детей из семей иноэтничных мигрантов не будут отличаться от таковых этнического большинства.

Во-вторых, мы использовали сетевой анализ, то есть изучали реальные связи и отношения между школьниками в классе. Этот метод позволяет непосредственно измерить интеграцию тех или иных детей в социальную ткань класса и школы.

Для обработки сетевых данных была применена многоуровневая модель p2, реализованная в компьютерном пакете StOCNET. Модель p2 предложили в 2004 г. ван Дуийн, Снайдерс и Зильстрой; в 2006 году модель усовершенствована для многоуровневого анализа, в результате чего стал возможен одновременный анализ нескольких сетей¹. Этот метод позволяет моделировать случайные эффекты на диадных данных, причем в качестве независимых переменных могут выступать как атрибуты акторов, так и свойства диад (диадные ковариаты).

¹ Van Duijn M.A.J. et al. P2: a random effects model with covariates for directed graphs // *Statistica Neerlandica*. 2004. Vol. 58. P. 234–254; Zijlstra B.J.H. et al. The multilevel p2 model: A random effects model for the analysis of multiple social networks // *Methodology*. 2006. Vol. 2. P. 42–47.

Для построения моделей были использованы классы с тремя и более учениками, принадлежащими к этническим меньшинствам. Таких классов в нашей выборке было 80, они распределялись по 53 школам, общее количество школьников в этих классах равнялось 1575. Основные характеристики классов приведены в Таблице.

Подготовка данных для сетевого анализа является отдельным трудоемким этапом работы. Во-первых, необходимо унифицировать все имена, названные респондентами, и сопоставить их с соответствующими атрибутами (характеристиками акторов). Во-вторых, необходимо преобразовать данные из анкет в виде матриц сопряженности в формате, пригодном для ввода в программу StOCNET. Для выполнения обоих этапов мы воспользовались специально написанными скриптами на языке Python.

Таблица

Класс	Min	Max	Mean	Median
Размер	8	29	20,5	20
Число minority	3	10	4,2	4,0
% minority	4,0	43,8	21,3	18,6

В нашем анализе зависимой переменной являлись дружеские связи между учениками. Моделировались вероятности образования связей между школьниками в зависимости от их пола, этничности, социально-экономического статуса семьи, отношения к школе и образованию и т.д. Анализ проводился в несколько этапов: вначале каждая из объясняющих переменных была проверена индивидуально, затем в модель последовательно добавлялись переменные. В конечной модели были сохранены все переменные, для которых был обнаружен статистически значимый эффект.

Интерпретация коэффициентов модели r^2 как коэффициентов логистической регрессии позволила сделать следующие выводы.

- Самый сильный эффект на образование связей оказывает гендерная гомофилия, причем среди мальчиков он выражен сильнее, чем среди девочек. Это известная закономерность, которая неизменно выявляется во всех исследованиях подросткового общения и дружб. Подчеркнем, что все остальные эффекты считались с учетом этого основного влияния.
- Русскоязычные дети (majority) не имеют предпочтений относительно этнической принадлежности своих друзей; этнические меньшинства (minority) проявляют тенденцию образовывать связи с minority.
- С учетом эффектов гендера и этничности была также выявлена гомофилия по анти-школьным аттитюдам и по намерениям относительно получения высшего образования, а именно: дружба более вероятна между учениками с близким уровнем анти-школьной культуры, а также со сходными планами относительно высшего образования.
- Гомофилии по социально-экономическому статусу, чувству принадлежности, намерениям оставить школу после 9 класса не обнаружено.
- Ни одна из групповых характеристик (тип школы, размер школы, число и процент иноэтничных детей в классе) не оказалась значимой в нашей модели, т.е. выявленные закономерности образования дружеских связей не связаны с типом и составом школы.

Изучение отчислений среди студентов бакалавриата/специалитета НИУ ВШЭ

Колотова Елена Васильевна, *НИУ ВШЭ*

Введение

Отчисления студентов, в наиболее широком понимании включающие в себя все случаи выбывания из университета до момента получения диплома, являются популярной темой для исследований как в США, так и в большинстве стран Европы. В этих странах уровень отсева является одним из показателей конкурентоспособности вуза: чем ниже уровень отчислений,