

УДК 303.7:159.942

А.А. Бова,

кандидат социологических наук,
старший научный сотрудник

ИССЛЕДОВАНИЕ ВОСПРИЯТИЯ ЛИЧНОЙ БЕЗОПАСНОСТИ В СТРАНАХ МИРА МЕТОДАМИ ТИПОЛОГИЧЕСКОЙ РЕГРЕССИИ

Для изучения зависимости восприятия личной безопасности от уровня убийств, человеческого развития, доверия к правительству и людям построены регрессионные уравнения на однородных группах. Модели типологической регрессии характеризуются лаконичностью, точностью и вариативностью, а также лучшей содержательной интерпретацией. Эмпирической базой исследования служили обобщенные результаты Всемирного опроса Гэллапа 2011 и материалы Доклада о человеческом развитии 2013 ООН по 115 странам мира.

Ключевые слова: *восприятие личной безопасности, кросс-национальные исследования, типологии, деревья решений, кусочно-линейная регрессия.*

Для вивчення залежності сприйняття особистої безпеки від рівня вбивств, людського розвитку, довіри до уряду і людей побудовані регресійні рівняння на однорідних групах. Моделі типологічної регресії характеризуються лаконічністю, точністю і варіативністю, а також кращою змістовною інтерпретацією. Емпіричною базою дослідження служили узагальнені результати Всесвітнього опитування Геллана 2011 і матеріали Доповіді про людський розвиток 2013 ООН по 115 країнах світу.

Ключові слова: *сприйняття особистої безпеки, крос-національні дослідження, типології, дерева рішень, кусочно-лінійна регресія.*

A regression modeling approach for homogeneous subgroups are structured for studying on perceptions of personal safety from homicide rate, the level of human development, confidence in the government. The models of typological regression are characterized with laconism, exactness and variaty, as well as with the better interpretation. The empirical basis of the research are the generalized results of Gallup World Poll 2011 and materials of the Human Development Report 2013 of the United Nations in 115 countries of the world.

Keywords: *perception of personal safety, cross-national research, typologies, decision trees, piecewise-linear regression.*

Регрессионный анализ является наиболее часто употребляемым методом в эмпирических социальных исследованиях. В социологии при массовых опросах общественного мнения или изучении социально-экономической статистики довольно часто возникает ситуация, когда совокупность объектов является неоднородной, а закономерности носят нелинейный характер. В таких случаях целесообразно использовать регрессионный анализ совместно с преимуществами, предоставляемыми классификационными моделями [1–4]. Теоретические и практические разработки в области типологической регрессии активно проводились в 1970–1990-х годах, что было связано с потребностями общества и прогрессом вычислительной техники. Совершенствование указанных методов интеллек-

туального анализа данных (*Data Mining*) идет с одной стороны от регрессии к её кусочно-линейным реализациям, а с другой, – путем индукции деревьев решений с последующим построением линейных или полиномиальных регрессионных моделей. Заметим, что аппроксимация эмпирических данных может решаться разными методами, в том числе и посредством достаточно сложных вычислительных процедур, результаты которых зачастую затруднительно непосредственно содержательно интерпретировать (например, искусственные нейронные сети) [5]. В то же время система регрессионных уравнений, построенных на различных классификациях (кластерный анализ, логические закономерности), имеет для социолога вполне ясный смысл и позволяют осуществлять сравнения интенсивности тех или иных процессов в отдельных подгруппах. Основные способы создания типологической регрессии такие.

1. Разбиение значений признаков на интервалы, исходя из содержательных предположений или статистических требований (по квартилям или точкам изменения тренда) с последующим построением различных регрессионных моделей на образованных подмножествах. Существенно увеличивает долю объяснённой дисперсии кусочно-линейное решение при разделении значений зависимой переменной по среднему значению. При применении такой модели к новым данным принадлежность наблюдений к образованным подмножествам отклика может быть найдена сначала путем предварительного использования какого-либо алгоритма классификации, например случайного леса (*Random Forest*).

2. Кластеризация данных с дальнейшим построением регрессионных моделей в типологически однородных группах, использование регрессионного кластеринга, одновременно классифицирующего наблюдения с построением зависимостей между переменными кластеризации с максимизацией коэффициента детерминации или регрессия на латентных классах. Характеристика групп осуществляется путем подсчета средних значений переменных кластеризации или внешних переменных, не участвующих в соответствующей процедуре.

3. Применение алгоритмов индукции деревьев решений с вычислением средних значений в подмножествах (кусочно-константные регрессионные деревья – *CHAID*, *CART*), а также специализированных алгоритмов кусочно-линейной и кусочно-полиномиальной регрессии (*MARS*, *M5'*, *M5opt*, *GUIDE*, *Microsoft Decision Trees algorithm*). Деревья регрессии позволяют обрабатывать разнотипные независимые переменные, имеющие пропущенные значения, строить разнообразные линейные, в том числе робастные, и полиномиальные модели, задавать ансамбли моделей с общим регрессионным уравнением для всей выборки и отдельных подвыборок, учитывать значения *k*-ближайших соседей для соответствующей коррекции прогнозных значений, включать в обработку временные ряды и др. выборки, управлять процессом сегментации выборки, задавая переменную, которая первая включается в процесс построения дерева решения. В таком случае для одних и тех же данных можно получить ряд типологий (подробных или лаконичных), отвечающих научным интересам разных наук (например, социологии, криминологии, экономики).

Продемонстрируем возможности типологической регрессии на примере данных относительно восприятия населением личной безопасности от преступных посягательств. Многомерный статистический анализ осуществлялся в пакетах *JMP*, *STATISTICA*, *WEKA*, *Cubist*.

Вопросы, фиксирующие страх перед преступностью, включаются в национальные (например, Институтом Гэллага в США – с 1965 г.) и сравнительные опросы общественного мнения (Международный опрос жертв преступлений,

Евробарометр, Афробарометр, Европейское социальное исследование, Всемирный опрос Гэллапа). Собранные данные служат в качестве оценки качества жизни (в частности, присутствуют в Европейской системе социальных индикаторов) и криминальной ситуации в стране, а в обобщенном виде используются для формирования различных региональных и глобальных рейтингов безопасности. Важность показателя восприятия личной безопасности обусловило включение результатов Всемирного опроса Гэллапа за 2011 г. в аналитические таблицы Доклада о человеческом развитии 2013 Программы развития ООН [6, с. 144–147, 174–177]. Существует небольшая статистически значимая ($\alpha \pm = 0,01$) корреляционная связь между уровнем ощущения личной безопасности и некоторым показателем социально-экономического развития: индексом человеческого развития (отражающим среднюю величину достижений в трех основных измерениях человеческого развития – здоровье и долголетию, знаниях и достойных условиях жизни) – 0,25; валовым национальным доходом на душу населения в долларах США, выраженным по паритету покупательной способности – 0,38; коэффициентом умышленных убийств – 0,44; удовлетворенностью населением свободой выбора, тем, чем они могут заниматься в жизни, – 0,26 (парные коэффициенты корреляции Пирсона рассчитаны для $N=149$ стран) удовлетворенностью работой – 0,22; доверием к национальному правительству – 0,36 ($N=135$); мнением о том, что в целом большинству людей можно доверять, а не проявлять с ними осторожность – 0,4 ($N=128$).

Рабочая схема объяснения восприятия личной безопасности включает в себя информацию относительно социально-экономического развития страны, объективной ситуации с преступностью, эффективностью деятельности государственных институтов, ценности населения, отражающую солидарность. Показатели определялись таким образом:

Восприятие безопасности (personal safety – PS) – процент респондентов, ответивших “да” на вопрос: “Чувствуете ли вы себя в безопасности, когда прогуливаетесь в одиночестве ночью в городе или районе, в котором вы живете?”.

Индекс человеческого развития (Human development index – HDI) – комбинированный индекс, измеряющий среднюю величину достижений в трех основных измерениях человеческого развития: здоровье и долголетие, знания и достойные условия жизни.

Доверие к правительству страны (trust in national government – TNG) – процент респондентов, ответивших “да” на вопрос: “Доверяете ли вы правительству вашей страны?”.

Доверие к людям (trust in other people – TP) – процент респондентов, ответивших “да” на вопрос: “В целом, считаете ли вы, что большинству людей можно доверять, или вы считаете, что следует проявлять осторожность, имея дело с людьми?”.

Коэффициент убийств (homicide rate – HR) – число умышленных убийств, т. е. смертей, незаконно причиненных человеку другим человеком, в пересчете на 100 тыс. человек.

Ниже представлено уравнение линейной регрессионной модели с оценкой параметров по методу наименьших квадратов, рассчитанное для совокупности 115 стран.

$$\hat{PS} = 27,4 - 0,31HR + 0,32TNG + 0,41TP + 14,8HDI$$

Качество модели – скорректированный коэффициент детерминации, отражающий процент дисперсии зависимой переменной, который описывается

комбинацией предикторов, составляет $R^2=0,44$, стандартная ошибка оценки – 12. Все коэффициенты являются статистически значимыми. Исходя из их значений в стандартизованном масштабе, наибольший вклад в модель вносят колебания уровней доверия населения к национальному правительству (0,39) и людям (0,30), коэффициент убийств (-0,29) и, наконец, Индекс человеческого развития (0,17).

Кластерный анализ по методу максимизации ожиданий совокупности 115 стран по пяти переменным находят оптимальным разделением на три кластера. На рис. 1 представлен результат иерархической кластеризации, дающий наглядное представление о похожести профилей стран. Чем длиннее на рисунке линии, тем дальше кластеры расположены в пространстве один от другого. Так, например, ближайшими соседями Украины являются Литва, Румыния, Латвия и Греция.

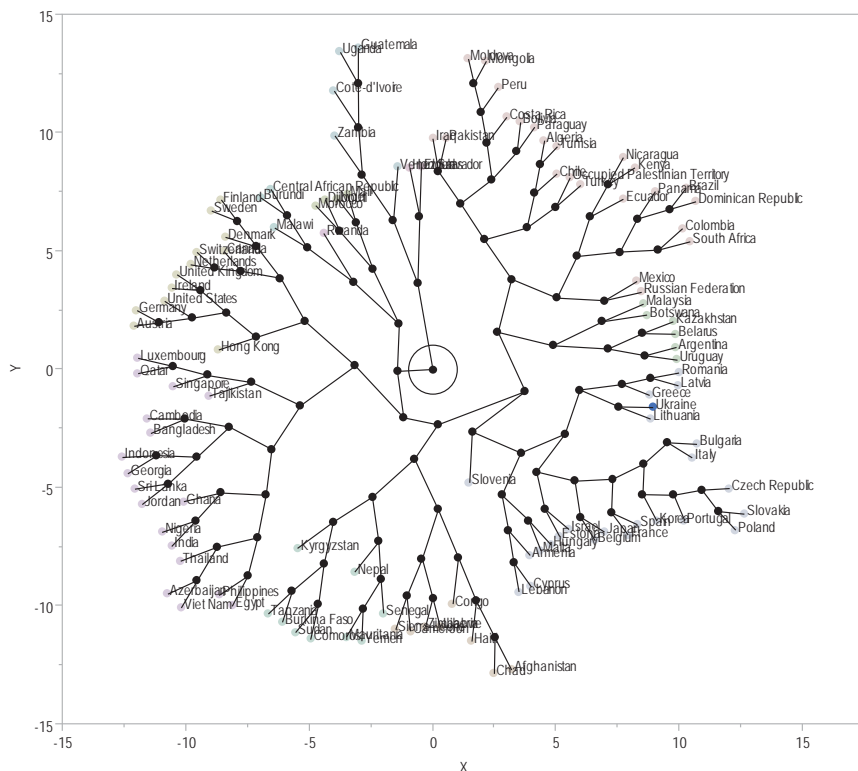


Рис. 1. Созвездия кластеров (метод средней связи)

На восприятие личной безопасности влияют как объективные, так и субъективные факторы. Решение для всей выборки, в случае её неоднородности, не совсем точно отражает характер закономерностей. Для иллюстрации различий между странами две переменные были разбиты на три одинаковые по численности группы (соответственно с низким, средним и высоким уровнем человеческого развития и убийств) и по комбинации факторов вычислено среднее значение доли респондентов, позитивно оценивающих личную безопасность. Из рис. 2 видно, что в странах с низким уровнем человеческого развития, вне зависимости от уровня убийств, восприятие личной безопасности населением является примерно одинаковым. В государствах со средним уровнем человеческого развития и высоким уровнем убийств существенно ниже процент людей, считающих ночное время на улице безопасным (в эту категорию попадает ряд латиноамериканских стран). В странах с высоким уровнем человеческого развития отсутствует высокий уровень убийств. И последний показатель весьма существенно влияет на восприятие персональной безопасности (дифференциацию разных стран).

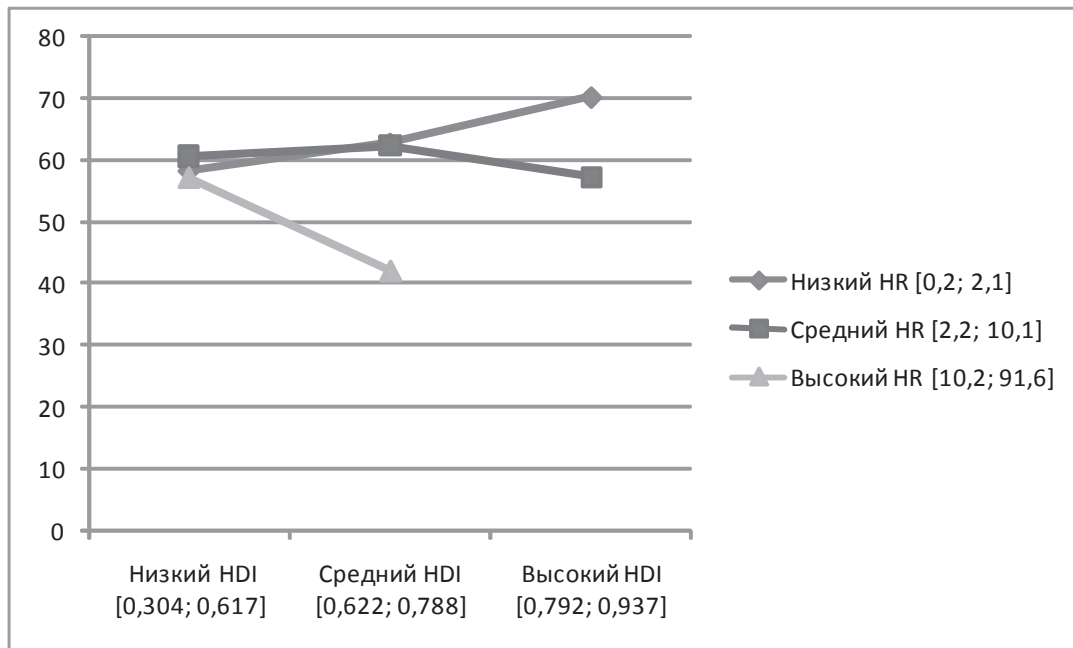


Рис. 2. Среднее значение восприятия личной безопасности в зависимости от комбинации уровней факторов

В странах с высоким уровнем убийств ($N=38$) фиксируется отрицательная статистически значимая ($\alpha \pm = 0,01$) корреляция между Индексом человеческого развития и уровнем личной безопасности ($r = -0,43$), с низким уровнем убийств ($N=38$) – статистически значимая ($\alpha \pm = 0,05$) положительная корреляционная связь ($r = 0,34$). В странах со средним уровнем убийств ($N=39$) корреляционная связь между указанными переменными отсутствует.

На основе предварительного анализа можно предположить, что предикторы могут иметь сложную и немонотонную связь с откликом. С помощью многомерных адаптивных регрессионных сплайнов (*Multivariate adaptive regression splines – MARS*) с редукцией кусочно-линейных базисных функций до четырех построено уравнение зависимости восприятия личной безопасности от трех предикторов (скорректированный $R^2 = 0,49$, стандартная ошибка – 11,2).

$$\hat{PS} = 40,75 + 2,55 \max(0, 7,8 - HR) - 1,45 \max(0, TP - 34) + 0,33 \max(0, TNG - 30) + 1,3 \max(0, TP - 24)$$

Для вычисления прогнозного значения для страны выбирается большее число между нулем и разностью двух чисел, что фактически определяет интервал значений параметров, в наибольшей степени влияющих на повышение процента людей в стране, которые не боятся преступности. В общем случае интерпретация модели множественных адаптивных линейных сплайнов вызывает трудности по сравнению с прозрачной структурой деревьев регрессии.

При выполнении типологизации разбиению могут подвергаться как факторы, так и отклик. В результате деления выборки на подгруппы по среднему значению зависимой переменной существенно увеличилась величина коэффициента детерминации – $R^2 = 0,79$ и уменьшилась стандартная ошибка оценки – 7,1. Для удобства представим единую формулу уравнения регрессии в виде логических правил с одинаковым набором переменных для двух групп.

Если $PS \leq 59\%$, то: $\hat{PS} = 53,3 - 0,23HR + 0,23TNG + 0,03TP + 10,6HDI$
 $(\overline{PS} = 47\%, \overline{HR} = 15, \overline{TNG} = 43\%, \overline{TP} = 19\%, \overline{HDI} = 0,67)$.

Если $PS > 59\%$, то: $\hat{PS} = 33,7 - 0,08HR + 0,004TNG + 0,18TP + 16,2HDI$
 $(\overline{PS} = 73\%, \overline{HR} = 5,7, \overline{TNG} = 58\%, \overline{TP} = 29\%, \overline{HDI} = 0,69)$

Средние значения переменных в первой группе ($N=60$ или 52%) существенно отличаются от средних значений во второй группе ($N=55$ или 48% стран), кроме низкого уровня личной безопасности, более низким уровнем доверия к правительству и людям, более высоким уровнем убийств. Описываемые уравнения тенденции аналогичны уравнению регрессии, построенному на всей выборке. При предсказании новых значений необходимо вычислить по соответствующей классификационной функции отнесения объекта к той или иной группе зависимой переменной. Логистическая регрессия (R^2 Найджелкерка=0,52) правильно предсказывает принадлежность только $77,4\%$ наблюдений. Последовательное комбинированное решение (логистическая и кусочно-линейная регрессия) снижает R^2 до $0,42$, а стандартную ошибку увеличивает до $12,6$.

Для выделения структурных различий между странами по ситуации с безопасностью и отбора наиболее информативных переменных построены линейные регрессии с оценкой параметров по методу наименьших квадратов для двух логических правил.

Если $HR < 5,05$, то: $\hat{PS} = 29,3 - 0,87HR + 0,32TNG + 0,25TP + 21,5HDI$ ($\overline{PS} = 67\%$)
 Иначе: $\hat{PS} = 22,9 + 0,32TNG + 0,58TP$ ($\overline{PS} = 51\%$)

Такая модель имеет преимущества как над деревьями решений, выражающиеся в компактности и возможности получить точечное прогнозное значение по значениям независимых переменных, а не для интервала значений признаков терминальной вершины, так и над кластеризацией с последующим проведением регрессионного анализа, – в точности и простоте предсказания новых значений без предварительного отнесения объекта к соответствующему кластеру.

Коэффициент детерминации общей модели $R^2=0,52$, стандартная ошибка оценки – $10,8$. Уравнения включают разное количество информативных признаков для первой (53% объектов) и второй (47%) группы стран. Отметим, что при более высоком уровне убийств во второй категории стран улучшение восприятия личной безопасности достигается за счет более выраженного социального капитала, что проявляется в большем доверии к социальным институтам и людям. В 53 странах, где коэффициент убийств составляет более 5 случаев на 100 тыс. населения, увеличение Индекса человеческого развития связано с ростом страха перед преступностью ($r=-0,33, \alpha \pm = 0,05$). Полученные результаты позволяют перенести фокус эмпирического исследования на управленческую практику.

С помощью алгоритма *M5 rules* также были сформированы разбиения выборки на две подгруппы не только по уровню личной безопасности и коэффициенту убийств, но и, отдельно, по Индексу человеческого развития. Первая типология может представлять преимущественный интерес, например, для социологов (различия социально-экономических показателей при заданном мнении), вторая – для криминологов (различия мнений при заданном уровне преступности), третья – для экономистов (различия мнений при заданном уровне человеческого развития). В последнем случае отмечается меньшая разница между средними

значениями зависимой переменной в группах. Коэффициент детерминации общей модели $R^2=0,55$, стандартная ошибка оценки – 10,5.

$$\begin{array}{l} \text{Если } HDI \leq 0,745, \text{ то: } \hat{PS} = 31,4 + 0,32TP + 0,42TNG - 0,24HR \quad (\overline{PS} = 58\%) \\ \text{Иначе:} \quad \hat{PS} = -91,7 + 170,2 HDI + 0,25TNG - 0,49HR \quad (\overline{PS} = 62\%) \end{array}$$

Во второй группе стран ($N=47$) по сравнению с первой ($N=68$) отмечается более высокий уровень человеческого развития – соответственно 0,85 против 0,56, валового национального дохода на душу населения – 26 162 доллара против 3 854 доллара, доверия к людям – 27 % против 21 %. В первой группе по сравнению со второй гораздо выше коэффициент убийств – 14,9 против 4,2 и уровень доверия к правительству соответственно – 54 % против 44 %. На основании значений стандартного отклонения, можно сделать вывод про большую однородность второй группы по сравнению с первой по показателю Индекса человеческого развития, валового национального дохода и коэффициенту убийств.

Увеличение показателя соответствия модели данным достигается и путем разделения дерева на большее число терминальных узлов, заключающих в себе линейные уравнения и константы, однако при небольшом объеме эмпирических данных целесообразно ограничить глубину дерева одним-двумя уровнями. Прогноз также может вычислить как среднее значение по комитетам (ансамблям) регрессионных сплайнов, а отдельные модели – иметь самостоятельную содержательную интерпретацию. Значимость переменных определяется по их вхождению в части уравнений ансамблей. Так, первое по значимости место занимает переменная HR , второе TNG и TP , третье – HDI . Таким образом, использование нескольких регрессионных уравнений позволяет существенно дополнить закономерности, выявленные для выборки в целом, и повысить точность прогноза.

С учетом изложенного, можно сделать такие выводы.

Страх перед преступностью частично связан с распространением убийств, отражающих потенциал агрессии населения, социально-экономическим развитием, шансами, которые предоставляет человеку общество, проявлениями неблагополучия последнего, эффективностью функционирования государственных институтов и уровнем доверия между людьми. На основании проведенных аналитических и эмпирических типологий можно предположить, что в общей выборке существует несколько крупных групп стран, которые проявляют большую однородность между собой по исследуемым переменным. Социально-экономическое развитие имеет нелинейную связь с криминальной ситуацией. Оно не гарантирует автоматического повышения защиты личности от криминальных посягательств. Институциональная составляющая общественного прогресса в этом контексте приобретает особую важность.

Типологическая регрессия является гибким инструментом исследования связей в неоднородных совокупностях. В отличие от типологической регрессий, основанной на предварительном содержательном или эмпирическом разбиении совокупности, например, на квантили или кластеры, типологическая регрессия, основанная на классификационных методах с логическими условиями (деревьях решений), позволяет оптимально сегментировать данные и точнее определять значения новых наблюдений. С помощью методов *Data Mining* были обнаружены и количественно описаны неочевидные закономерности. Они хорошо описывают поведение только половины наблюдений обучающей выборки, что может быть связано, в том числе, с различными источниками формирования доверия к социальным институтам в странах и неполнотой учетов преступлений. Полученные

в адаптивной регрессии и регрессионных деревьях значения независимых переменных являются маяками для стран, стремящихся улучшить восприятие населения личной безопасности от преступных посягательств.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Андреев В.Г., Толстова Ю.Н., Мирзоев А.А.* Типология и классификация в социологических исследованиях / отв. ред. В.Г. Андреев, Ю.Н. Толстова. – М. : Наука, 1982. – 296 с.
2. *Пакет прикладных программ ОТЭКС / Загоруйко Н.Г., Елкина В.Н., Лбов Г.С., Емельянов С.В.* – М. : Финансы и статистика, 1986. – 158 с.
3. *Статистическое моделирование и прогнозирование : Учеб. пособие / Г.М. Гамбаров, Г.М. Журавель, Ю.Г. Королев и др. ; Под ред. А.Г. Гранберга.* – М. : Финансы и статистика 1990. – 383 с.
4. *Эконометрика : учебник / Мхитарян В. С., Архипова М. Ю., Балаш В. А. и др. ; Под ред. В.С. Мхитаряна.* – М. : Проспект, 2009. – 384 с.
5. *Бова А.А.* Линейный и нелинейный регрессионный анализ (на примере изучения личной безопасности в странах мира) // Вісник НУТУ “КПІ”. Політологія. Соціологія. Право. – 2014. – Випуск 2 (22). – С. 28–34.
6. *Доклад о человеческом развитии 2013. Возвышение Юга : человеческий прогресс в многообразном мире / [директор и основной автор Х. Малик].* – М. : Издательство “Весь Мир”, 2013. – 203 с.

Отримано 07.04.2016