

Применение альтернативного варианта конфирматорного факторного анализа (на материалах исследований фонематического восприятия речи у американских школьников)

Куравский Л. С. *,

*доктор технических наук, профессор,
декан факультета информационных технологий, МГППУ*

Абрамочкина В. И. **,

студентка IV курса факультета информационных технологий МГППУ

Петрова Е. А. ***,

студентка IV курса факультета информационных технологий МГППУ

Рассмотрены результаты исследований фонематического восприятия речи у американских школьников, на основании которых были выявлены недостатки традиционного конфирматорного факторного анализа. Представлен разработанный авторами альтернативный вариант данного метода, который предполагает исследования взаимосвязи свободных параметров факторной модели. Он позволяет находить параметры прямым методом, обеспечивая однозначное оптимальное решение. Показано, что суть разработанного метода состоит в составлении переопределенной системы уравнений, решение которой осуществляется прямым методом, а также в проверке адекватности полученных систем с помощью статистических критериев согласия. Адекватность полученной модели определяется из статистики χ^2 -квадрат, которая позволяет проверять гипотезу о представимости выбороочных дисперсий и ковариаций, составляющих правую часть переопределенной системы уравнений, компонентами дисперсий и ковариаций латентных переменных, содержащихся в исследуемой модели.

Разработанный альтернативный вариант конфирматорного факторного анализа позволяет также оценить статистическую значимость компонентов модели.

Полученные оценки свободных параметров с помощью альтернативного варианта факторного анализа не обнаружили значимых отличий от результатов, полученных с помощью традиционного факторного анализа, что позволяет говорить о перспективности нового подхода.

Ключевые слова: прямой метод, переопределенная система уравнений, оценка максимального правдоподобия, критический процент.

* rusav@aha.ru

** ron666@bk.ru

*** petrovaeva@mail.ru

В связи с увеличением потока иммигрантов из Мексики и ростом испаноговорящих детей в образовательных учреждениях Соединенных Штатов Америки встал вопрос, как построить образовательную программу, чтобы дети иммигрантов могли успешно обучаться английскому языку [6, с. 170–181], [10, с. 439–449]. Поэтому американскими психологами была поставлена задача по исследованию фонематического восприятия у испаноговорящих детей. Целью работы было выявление степени общности восприятия фонем для различных языков, а также решение вопроса, в какой степени восприятие написанного слова отличается от восприятия фонемы слова. Исследование американских психологов было выполнено с помощью традиционного конфирматорного факторного анализа.

Разработанный альтернативный вариант конфирматорного факторного анализа позволяет находить параметры исследуемой модели прямым методом, позволяющим избежать трудоемкого и неоднозначного процесса итерационного решения задачи многомерной оптимизации [9] и обеспечивая однозначное оптимальное решение.

Традиционный конфирматорный факторный анализ предполагает наличие строго сформулированной модели изучаемого явления. Факторная модель, связывающая наблюдаемые и латентные переменные, составляется, опирается на знание предметной области. Гипотезы о структуре модели должны основываться на анализе природы исследуемых факторов (учитывается как теория, так и наблюдения) [9]. Можно делать количественные предположения о значениях корреляций между латентными переменными и величинах факторных нагрузок. Свободные параметры модели подбираются так, чтобы получить наилучшее, с точки зрения заданного критерия, приближение к матрицам корреля-

ций (или ковариаций) между наблюдаемыми переменными.

Объектом конфирматорного факторного анализа являются ковариационные или корреляционные матрицы наблюдаемых переменных. Цель анализа – выявить значения параметров модели, которая с приемлемыми ошибками объясняет изменчивость наблюдений.

Конфирматорный факторный анализ имеет следующие особенности:

- ненулевые (свободные) факторные нагрузки в уравнениях модели и число исследуемых факторов определяются заранее;
- допускается корреляция между ошибками измерений;
- факторные нагрузки и ковариации между латентными переменными могут быть свободными параметрами модели или приравняться заданным константам;
- допускается анализ нескольких групп моделей;
- можно проверять, согласуются ли заданные ограничения с результатами наблюдений;
- оценки свободных параметров определяются методом максимального правдоподобия.

С последней особенностью связана основная проблема метода. Для определения оценок свободных параметров модели методом максимального правдоподобия в качестве минимизируемого критерия используется функция:

$$F = [\ln |S| - \ln |S| + \text{tr}(S\Sigma^{-1}) - m] (N-1),$$

где S – выборочная ковариационная матрица наблюдаемых переменных, Σ – прогнозируемая ковариационная матрица наблюдаемых переменных, $|S|$ и $|S|$ – определители матриц S и S , $\text{tr}(S\Sigma^{-1})$ – след матрицы $(S\Sigma^{-1})$, N – объем выборки, использованной для вычисления матрицы S , m – число наблюдаемых переменных [3].

Элементы прогнозируемой ковариационной матрицы представляют собой аналитические выражения относительно свободных параметров модели [4]. В случае многомерного нормального распределения наблюдаемых переменных значения критерия F описываются распределением хи-квадрат.

Таким образом, для определения свободных параметров модели необходимо численно решать итерационными методами достаточно трудоемкую задачу локальной многомерной оптимизации. Из этого в общем случае вытекает невозможность определения глобального минимума, так как в результате решения находится один из локальных минимумов, зависящий от начального приближения. Следовательно, решение неоднозначно.

Суть разработанного альтернативного варианта конфирматорного факторного анализа – составление переопределенной системы уравнений, решение которой осуществляется прямым методом, а также проверка адекватности полученных систем с помощью статистических критериев согласия [2, с. 15–31], [4], [7].

Для составления системы уравнений используется модель дисперсионных компонентов, которая предполагает равенство единице путей коэффициентов регрессии. Каждой наблюдаемой дисперсии и ковариации ставится в соответствие уравнение, выражающее ее прогнозируемое значение через свободные параметры модели (дисперсии и ковариации латентных переменных) [5], [8].

В результате составляется система уравнений

$$Ax = b,$$

где A – матрица системы, коэффициенты ко-

торой определяются, опираясь на факторную модель; b – вектор-столбец n выборочных оценок дисперсий и ковариаций, которые определяются по результатам наблюдений; x – вектор-столбец m искоемых свободных параметров системы (компонентов дисперсий и ковариаций латентных переменных).

Рассмотрим вектор $\varepsilon = Ax_* - b$, представляющий *невязку псевдорешения* x переопределенной системы, полученного методом наименьших квадратов. Полагая в общем случае, что компоненты вектора невязок коррелированы, запишем их ковариационную матрицу в виде $\sigma^2 V$. Сделаем замены

$$b = V^{1/2} b_0, \quad A = V^{1/2} A_0,$$

где $V = V^{1/2} V^{1/2*}$, перейдем к системе

$$A_0 x = b_0,$$

ковариационная матрица вектора невязок $\varepsilon_0 = V^{-1/2} \varepsilon$ которой имеет вид $\sigma^2 E$, где E – единичная матрица.

Если система (3) невырождена ($\text{rank } A = m$), вектор невязки ε_0 имеет многомерное нормальное распределение, а

$$x_* = (A_0^T A_0)^{-1} A_0^T b_0 = (A^T V^{-1} A)^{-1} A^T V^{-1} b$$

псевдорешение системы (3), полученное методом наименьших квадратов, то это псевдорешение является *оценкой максимального правдоподобия*, а статистика

$$\chi^2 = (b_0 - A_0 x_*)^T (b_0 - A_0 x_*) / \sigma^2 = (b - Ax_*)^T V^{-1} (b - Ax_*) / \sigma^2$$

имеет распределение хи-квадрат со степенями свободы [1].

При решении поставленной задачи использовались упрощенные путевые диаграммы и ряд допущений (диаграмма). В качестве наблюдаемых переменных выступают тесты на английском и испанском языках, с помощью которых исследовались школьники. Основная путевая диаграмма, взятая из результатов исследований аме-

*Для любой симметричной неотрицательно определенной матрицы V (именно к этому классу относятся ковариационные матрицы) существует единственная симметричная неотрицательно определенная матрица $V^{1/2}$, называемая *квадратным корнем* из V , такая, что $(V^{1/2})^2 = V$.

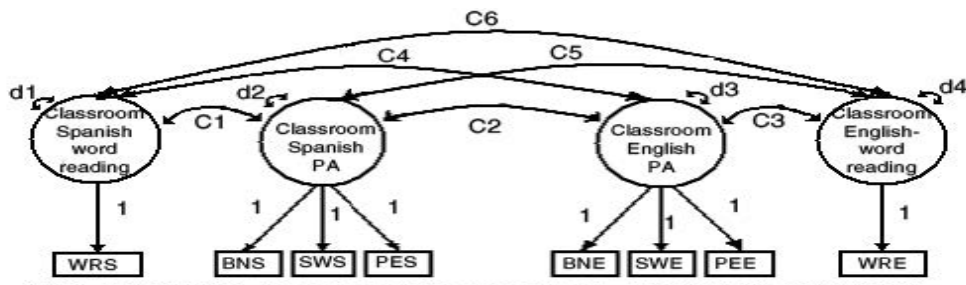


Рис. Упрощенная путевая диаграмма

риканских психологов, была разделена на две независимые части: индивидуальный и групповой уровень. Значения дисперсий и ковариаций наблюдаемых переменных заменялись переменными: $d_{1,2,3,4}$ для дисперсий и $C_{1,2,3,4,5,6}$ для ковариаций.

Предполагалось, что

- компоненты вектора невязок ε не коррелируют между собой;

- значения среднеквадратических отклонений различных компонентов вектора ε приравняются к одному и тому же фиксированному проценту от соответствующих компонентов вектора b ;

- указанный процент подбирается так, чтобы выполнялось равенство $\chi^2 \geq \chi^2_{n-m}$; α при уровне значимости $\alpha = 0.05$ (α – уровень значимости критерия), после чего эта величина оценивается на допустимость. (Уровень данной характеристики легче всего оценивать, установив для него разумное критическое значение, например, 0,1. Таким образом, вместо уровня значимости появляется новый критерий – критический процент.)

Разработанный альтернативный вариант конфирматорного факторного анализа позволяет также оценить статистическую значимость компонентов модели. Для этого необходимо провести сравнение двух моделей:

1) полной модели, в которой исследуемый компонент присутствует;

2) упрощенной модели, в которой этот компонент отсутствует.

Значение статистики хи-квадрат для упрощенной модели сравнивается с аналогичной характеристикой для полной модели. Поскольку разность указанных статистик асимптотически распределена как хи-квадрат с числом степеней свободы, равным разности в числах степеней свободы для полной и упрощенной моделей, эта разность используется для проверки нулевой гипотезы H_f о том, что упрощенная модель согласуется с результатами наблюдений, против альтернативной гипотезы H_f о согласовании с результатами наблюдений полной модели. Если гипотеза H_f не отвергается при заданном уровне значимости, то исследуемый компонент признается статистически незначимым и делается вывод о том, что имеющиеся данные не свидетельствуют о его влиянии на данную характеристику. Если гипотеза H_f отвергается (а гипотеза H_f принимается), можно говорить о влиянии исследуемого компонента на эту характеристику.

В исследованиях фонематического восприятия у американских школьников были выявлены взаимосвязи восприятия фонем на родном испанском языке и английском языке. Хорошее владение фонематическим восприятием на родном языке – испанском – обуславливает лучшее выполнение

заданий на «чужом» английском языке, что наблюдалось как в результатах, полученных в индивидуальном тестировании учеников, так при тестировании всего класса. Таким образом, подтверждено влияние модели преподавания на развитие фонематического восприятия у испаноговорящих американских школьников [6, с.170–181].

Полученные оценки свободных параметров с помощью альтернативного факторного анализа не обнаружили значимых отличий от результатов, полученных с помощью традиционного конфирматорного факторного анализа, что позволяет подтвердить результаты американского исследования и говорит о перспективности нового подхода.

Литература

1. *Королюк В. С., Портенко Н. И., Скороход А. В., Турбин А. Ф.* Справочник по теории вероятностей и математической статистике. М., 1985.
2. *Куравский Л. С., Корниенко П. А.* Применение нейронных сетей для идентификации локусов количественных признаков в психогенетике // *Нейрокомпьютеры: разработка и применение.* 2007. № 4.
3. *Лоули Д., Максвелл А.* Факторный анализ как статистический метод. М., 1967.
4. *Bishop Y. M. M., Fienberg S. E., Holland P. W.* Discrete multivariate analysis: Theory and practice. Cambridge, 1975.
5. *Bollen K. A.* Structural equations with latent variables. N. Y., 1989.
6. *Branum-Martin L., Carlson C. D., Carlo M., Fletcher J. M., Francis D. J., Mehta P. D., Ortiz A.*

- Bilingual phonological awareness: Multilevel construct validation among Spanish-speaking kindergarteners in transitional bilingual education classrooms // Journal of Educational Psychology.* 2006. Vol. 98. № 1.
7. *Goldstein H.* Multilevel statistical models (3rd ed.). L., 2003.
8. *Loehlin J. C.* Latent variable models: An introduction to factor, path, and structural analysis. - Hillsdale, N. J., 1987.
9. *Neale M. C., Cardon L. R.* Methodology for genetic studies of twins and families. Dordrecht, the Netherlands: Kluwer Academic Publishers, 1992.
10. *Schatschneider C., Francis D. J., Foorman B. R., Fletcher J. M., Mehta P.* The dimensionality of phonological awareness: An application of item response theory// *Journal of Educational Psychology.* 1999. Vol. 91.

**APPLICATION OF AN ALTERNATIVE METHOD OF CONFIRMATORY
FACTOR ANALYSIS (BASED ON THE MATERIALS OF PHONEMIC SPEECH
PERCEPTION RESEARCH OF AMERICAN SCHOOL CHILDREN)**

Kuravsky L.S.,

Post Doctorate Degree in Engineering, Professor, Dean, Department of Information Technologies, MSUPE

Abramochkina V.I.,

Student, Department of Information Technologies, MSUPE

Petrova E.A.,

Student, Department of Information Technologies, MSUPE

The research of phonemic speech perception of American school children revealed shortcomings of traditional confirmatory factor analysis. An alternative version of this method was developed.

The new version of confirmatory factor analysis suggests examination of factor model free (spare) parameters interrelation. It allows finding parameters directly providing for single optimal solution. The essence of this method is to build redefined equations set, which is resolved directly, and testing adequacy of obtained set with statistical fit indices. The adequacy of obtained model is determined by chi-square statistic, which allows checking the hypothesis about presentability of sampling variances and co-variances of latent variables that constitute the right-hand part of the overdetermined equations set, and by the components of variances and co-variances of latent variables contained in the analyzing model.

The developed alternative method of confirmatory factor analysis also allows assessing statistical value of model's components.

The results of examination of free parameters by alternative version of factor analysis didn't show significant difference from results of traditional factor analysis. These findings are encouraging and future applications of the new approach are discussed.

Keywords: direct method, overdetermined equations set, estimation of maximal likelihood, critical percentage.

References

1. Korolyuk V. S., Portenko N. I., Skorohod A. V., Turbin A. F. Spravochnik po teorii veroyatnostei i matematicheskoi statistike. M., 1985.
2. Kuravskii L. S., Kornienko P. A. Primenenie neironnyh setei dlya identifikatsii lokusov kolichestvennyh priznakov v psihogenetike // Neurokomp'yutery: razrabotka i primeneniye. 2007. № 4.
3. Louli D., Maksvell A. Faktornyi analiz kak statisticheskii metod. M., 1967.
4. Bishop Y. M. M., Fienberg S. E., Holland P. W. Discrete multivariate analysis: Theory and practice. Cambridge, 1975.
5. Bollen K. A. Structural equations with latent variables. N. Y., 1989.
6. Branum-Martin L., Carlson C. D., Carlo M., Fletcher J. M., Francis D. J., Mehta P. D., Ortiz A.

Bilingual phonological awareness: Multilevel construct validation among Spanish-speaking kindergarteners in transitional bilingual education classrooms // Journal of Educational Psychology. 2006. Vol. 98. № 1.

7. Goldstein H. Multilevel statistical models (3rd ed.). L., 2003.

8. Loehlin J. C. Latent variable models: An introduction to factor, path, and structural analysis. Hillsdale, N. J., 1987.

9. Neale M. C., Cardon L. R. Methodology for genetic studies of twins and families. Dordrecht, the Netherlands: Kluwer Academic Publishers, 1992.

10. Schatschneider C., Francis D. J., Foorman B. R., Fletcher J. M., Mehta P. The dimensionality of phonological awareness: An application of item response theory // Journal of Educational Psychology. 1999. Vol. 91.