

В.Ф. Столярова, Т.В. Тулупьева, М.В. Абрамов, В.Б. Салахова
**ВЫЯВЛЕНИЕ ХАРАКТЕРИСТИК ИНДИВИДУАЛЬНОГО
ЧЕЛОВЕЧЕСКОГО КАПИТАЛА СОТРУДНИКОВ
ОРГАНИЗАЦИИ ПО ДАННЫМ САМООТЧЕТОВ О
ПРОФЕССИОНАЛЬНЫХ НАВЫКАХ И ЛИЧНОСТНЫМ
ОСОБЕННОСТЯМ**

Столярова В.Ф., Тулупьева Т.В., Абрамов М.В., Салахова В.Б. **Выявление характеристик индивидуального человеческого капитала сотрудников организации по данным самоотчетов о профессиональных навыках и личностным особенностям.**

Аннотация. В области рекрутинга и менеджмента персонала существует задача автоматизации процесса оценки характеристик человеческого капитала, учитывающего в том числе особенности личности сотрудника. Статья посвящена вопросу выявления характеристик индивидуального человеческого капитала, имеющих наибольший вклад в некоторые показатели эффективности сотрудника организации, таких как карьерный успех, по данным самоотчетов о профессиональных навыках и ответов на вопросы–утверждения о различных психологических аспектах личности. Предлагается общая структура опросного инструментария, опирающегося на самоотчеты сотрудников, а также формализация предполагаемых методов анализа таких вопросов. Для выявления групп респондентов, обладающих схожими профессиональными навыками, было предложено использовать кластерный анализ, который позволяет сохранить сложную структуру их взаимосвязи. Для выявления личностных особенностей сотрудников из вопросов–утверждений предлагается формировать шкалы и посредством методов современной теории тестирования получить оценки латентной переменной, отражающей личностные особенности. На завершающем этапе исследования предполагается использование аппарата регрессии для оценивания взаимосвязи выявленных кластеров и латентных характеристик личности с тем или иным индикатором успешности сотрудника. Предлагаемый подход представляет собой структуру пилотного исследования, позволяющего выделить характеристики человеческого капитала (профессиональные навыки и особенности личности), обладающие наибольшим вкладом в показатели эффективности сотрудника организации, и направлен на снижение трудозатрат на последующих этапах более подробного и прицельного исследования. Возможности предложенного подхода продемонстрированы на примере данных, собранных среди государственных гражданских служащих различных структур Российской Федерации. В качестве индикатора эффективности сотрудника рассматривается наиболее доступный к наблюдению аспект карьерного успеха, выраженный фактом наличия руководящей должности.

Ключевые слова: оценивание человеческого капитала, кластерный анализ, современная теория тестирования, опросный инструментарий, самоотчет.

1. Введение. Задача оценки характеристик индивида, имеющих наибольший вклад в успех определенной деятельности или функционирования некоторой системы, возникает в различных социоориентированных областях. Например, кредитный скоринг клиентов в сфере банковского обслуживания [1, 2] направлен на выявление благонадежных плательщиков; оценка уровня защищенности

пользователей информационных систем от социоинженерных атак является частью системы кибербезопасности организации [3 – 5]; скоринг водителей используется при оценке риска дорожно–транспортных происшествий [6 – 8].

В области рекрутинга важную роль играют методы оценки человеческого капитала сотрудников при определении потенциала успеха в выбранной сфере деятельности. Корректные и обоснованные решения по найму напрямую влияют на процветание организации [9, 10]. В свою очередь потеря недавно обученных сотрудников или ключевых специалистов всегда сопряжена со значительными издержками со стороны организации [11], и потому управление человеческими ресурсами требует как разработки систем оценки человеческого капитала, так и систем мониторинга производительности сотрудников.

Одной из особенностей подобных систем является многогранность понятия индивидуального человеческого капитала [12], и, соответственно, разнородность характеризующих его факторов. Среди них выделяют как уровень владения профессиональными навыками, так и некоторые психологические особенности личности [9] и когнитивные способности [13]. Кроме того отмечается [13, 14], что в контексте парадигмы Индустрии 4.0 востребованными являются навыки владения техническими инструментами и надпрофессиональные навыки (*soft skills*), которые также относятся к человеческому капиталу сотрудника. В настоящее время существует несколько подходов к измерению человеческого капитала организации [12]: субъективный, приблизительный и прямой. Последний предполагает оценку ненаблюдаемых характеристик человеческого капитала посредством использования уже существующих психометрических шкал и разработки новых. Различают два уровня: микроуровень, который включает оценку факторов *индивидуального* человеческого капитала и их вклада в индикаторы личной успешности сотрудника; и макроуровень, то есть уровень фирмы [15], который заключается в оценке влияния совокупного набора человеческих капиталов сотрудников на эффективность работы организации в целом.

Часто информация о профессиональных успехах сотрудника: пройденных курсах повышения квалификации, завершенных проектах, премировании, а также данные из социальных медиа [11, 16], доступны внутри организации, и могут использоваться для оценки человеческого капитала [11]. Однако опросы и интервью респондентов об их профессиональных навыках также являются ценным источником знаний, который используется, например, в ситуации приема на

работу нового сотрудника или при мониторинге эффективности повышения квалификации. Такие самоотчеты являются составной частью современных подходов к оценке эффективности сотрудников, выявлению ключевых черт их человеческого капитала, как например, в методике 360 градусов (360 degrees feedback) [18], и используются в комбинации с экспертной информацией. Кроме того, сведения о надпрофессиональных навыках сотрудника и его психологических особенностях извлекаются при помощи специализированных тестов, часто основанных на вопросах—утверждениях, результаты которых также важно учитывать при оценивании человеческого капитала. Такие данные используются при исследовании различных аспектов труда, например, его ценностно—мотивационных основ [19].

Таким образом, актуальной является задача разработки методов и подходов к оцениванию человеческого капитала сотрудников, позволяющих использовать информацию из разных источников, а также автоматизации сбора и анализа такой информации.

Для оценивания характеристик человеческого капитала используются методы машинного обучения [11, 16, 17], в том числе и *кластерный анализ* [16, 20], который направлен на выделение схожих по характеристикам групп сотрудников, и *регрессионный анализ* — для построения предсказательных моделей. Кластерный анализ, к примеру, лежит в основе систем ранжирования младшего медицинского персонала [21], оценки эффективности использования цифровых инструментов учителями [22]. На макроуровне взаимосвязь между характеристиками человеческого капитала и экономическими показателями часто имеет сложный характер, который затруднительно уловить посредством классических эконометрических моделей, и потому в этой области нашли свое применение нейронные сети [23, 24]. Однако такие модели опираются на уже выявленные наборы характеристик человеческого капитала, ключевых для каждого контекста деятельности предприятия. В этом плане не существует метода оценивания человеческого капитала, общего для всех областей функционирования организации, всегда требуется учитывать особенности конкретного вида деятельности [25]. Например, в сфере государственного и муниципального управления создаются и функционируют различные ассессмент—центры [26, 27], в которых используются разнородные данные для оценки компетенций сотрудников, большое влияние имеет и экспертная информация.

Таким образом, возникает задача выявления наборов и паттернов характеристик индивидуального человеческого капитала,

которые являются ключевыми в контексте деятельности отдельного предприятия и исследуемого показателя эффективности сотрудника (или организации). Подходы к решению поставленной задачи во многом являются эвристическими, и потому разработка методов, допускающих автоматизацию сбора и обработки информации, является актуальной.

Целью исследования является формализация подхода к оценке индивидуального человеческого капитала по данным из самоотчетов сотрудников о профессиональных навыках и психологических аспектах личности, измеренных с помощью вопросов–утверждений, с целью выявления тех характеристик, которые имеют наибольший вклад в значение показателей эффективности. Под эффективностью сотрудника понимается набор разнородных показателей, отражающих «уровень соответствия результатов какой-либо деятельности поставленным задачам» [28], одним из которых является, например, карьерный успех, индикатором которого служит факт наличия руководящей должности. Предлагается использовать кластерный анализ для выявления паттернов среди профессиональных навыков. Этот метод позволяет создать картину взаимосвязей между навыками, критичными для каждой отдельной области. Для выявления латентных характеристик личности, отраженных в ответах на вопросы–утверждения, предлагается использовать подход современной теории тестирования. На последнем этапе используется модель регрессии для выявления тех факторов человеческого капитала, которые несут наибольший вклад в некоторую характеристику эффективности сотрудника или организации, как карьерный успех. Предлагаемый подход представляет собой структуру пилотного исследования, результаты которого позволяют сконцентрироваться на наиболее важных в контексте конкретного показателя эффективности и области функционирования организации факторах человеческого капитала при разработке дальнейших этапов исследований. Кроме того, предложенный подход позволяет автоматизировать сбор информации о факторах человеческого капитала, и частично автоматизировать процесс обработки полученных данных. Возможности применения этого подхода продемонстрированы на примере пилотного опроса, предложенного государственным гражданским служащим.

Полученный в рамках исследования результат является новым в области оценивания человеческого капитала государственных гражданских служащих и составляет базу для дальнейшего изучения детерминант индивидуального человеческого капитала в этой области деятельности, равно как для разработки методов оценивания взаимосвязи человеческого капитала организации и ее эффективности. В области

информатики и автоматизации новой является формализация задачи выявления характеристик человеческого капитала сотрудников организации по данным самоотчетов, а также сформулированный подход к автоматизации процесса анализа полученных данных с целью определения тех факторов, которые оказывают наибольшее влияние.

2. Используемые методы. Для выявления групп схожих по самооценке профессиональных навыков респондентов предлагается использовать кластерный анализ, для формирования психометрических шкал — анализ ответов в рамках современной теории тестирования (Item Response Theory).

2.1. Кластерный анализ. Кластерный анализ относится к методам машинного обучения без учителя и позволяет выделять паттерны в многомерных таблицах данных. Существуют множество методов кластеризации [29], одни из них основаны на анализе матрицы различий объектов (иерархические), другие — на определении оптимального разбиения пространства данных на заданное число кластеров (k -средних и k -медоидов), третьи — на анализе расположения точек пространства данных (DBScan). При работе с категориальными данными, которые часто являются результатом различных опросов, используются иерархический кластерный анализ на основе подходящей метрики или же метод k -медоидов [29].

Ключевым шагом для проведения кластерного анализа является вычисление матрицы различий объектов [20]. Для признаков, измеренных в ранговой шкале, могут использоваться различные метрики, в частности L_1 -метрика Манхэттен. Для векторов $a = \{a_1, a_2, \dots, a_n\}$ и $b = \{b_1, b_2, \dots, b_n\}$, измеренных в порядковой шкале, расстояние между ними выражается как:

$$r_{Mahnattan} = \sum_{i=1}^n |a_i - b_i|.$$

Для кластеризации категориальных данных используются иерархическая кластеризация и кластеризация при помощи k -медоидов [29, 30], который является методом неиерархического разделения пространства данных. Он опирается на известное заранее число определяемых групп и в ходе реализации алгоритма итеративно выполняет поиск заранее заданного числа точек пространства данных, наиболее удаленных друг от друга, причем точки x_i внутри соответствующих им кластеров C имеют минимальный возможный разброс в смысле выбранной метрики $d(\cdot)$:

$$\text{medoid}(C) := \arg \min_{x_i \in C} \sum_{x_j \in C} d(x_i, x_j).$$

Так как центром кластера является одно из наблюдений выборки, медоид, то этот метод кластеризации является более устойчивым к выбросам и робастным по сравнению с методом k -средних [29]. Алгоритм кластеризации методом k -медоидов РАМ является наиболее распространенным, и состоит из следующих шагов (приведено согласно [29]):

1. Случайно выбрать заранее заданное число k точек — начальных медоидов из выборки объема N .

2. Отнести все остальные наблюдения к ближайшему начальному медоиду согласно матрице расстояний и вычислить штраф, который является суммой расстояний всех точек кластера до его медоида.

3. Для каждого медоида k из множества $1 \dots K$ рассмотреть все остальные $N - K$ точек, и вычислить для каждой такой точки o штраф (как в пункте 2).

4. Выбрать в качестве медоида ту точку, которая дает наименьшее значение штрафа из всех.

5. Повторить пункты 2–4 до тех пор, пока медоид не будет меняться.

Существует несколько способов валидации полученных кластеров, например, средняя ширина силуэтов [29] и оценка устойчивости кластера при репликации исходной выборки (метод бутстреп) [31]. Значение силуэта является внутренней метрикой принадлежности объекта кластеру и может использоваться как для определения оптимального числа кластеров, так и для валидации качества кластеризации. Для отдельной точки x_i , отнесенной к кластеру C_k , значение силуэта определяется как:

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}, \quad (1)$$

где $a_i = \frac{1}{n_k - 1} \sum_{c(j)=k} d(x_i, x_j)$ есть среднее расстояние от x_i до остальных точек из этого же кластера k , а $b_i = \min_{l \neq k} \frac{1}{n_l} \sum_{c(j)=l} d(x_i, x_j)$ есть среднее различие точки x_i с точками из другого ближайшего кластера l . В приведенных формулах n_k и n_l представляют собой число точек в кластерах с номерами n и l соответственно.

Соответственно, средняя ширина силуэтов (average silhouette width) определяется как среднее индивидуальных значений силуэтов.

Метод бутстреп является методом имитационного моделирования, который опирается на создание репликаций исходного набора данных путем формирования новой выборки из исходных элементов, выбранных с повторением. Для валидации результатов кластерного анализа, этот метод используется следующим образом [29]. Пусть D обозначает набор данных, разделенный на K кластеров C_k .

1. Построить бутстреп выборку T^i размера n , определить проверяющее множество $E^i = T^i \cup D$.

2. Вычислить коэффициент соответствия Жаккара s_k^i между оригинальными кластерами C_k и кластерами, выделенными на бутстреп выборке: $C_k^{ET,i}$.

3. Среднее значение коэффициентов $s_k = 1/n \sum s_k^i$ отражает внутреннюю стабильность кластера.

Если значение s_k выше, чем 0.75, то кластер является стабильным, и воспроизводится при различных репликациях исходной выборки.

2.2. Современная теория тестирования (Item Response Theory, теория IRT). Современная теория тестирования [32, 33] возникла для анализа ситуаций, когда при проведении тестирования используются задания разного уровня сложности, при этом требуется учитывать, как респонденты с различными личностными особенностями справляются с каждым вопросом в отдельности. Для этого предполагается что существует некоторая *латентная характеристика* личности, которая отражает паттерн ответов на такие вопросы. Целью методов теории IRT является как оценка этой латентной характеристики, так и анализ вопросов, составляющих психометрическую шкалу. Модели современной теории тестирования используются в самых различных приложениях [34], в том числе при оценке различных характеристик сотрудников организаций [35], измерения состояния здоровья [36], маркетинговых исследованиях [37].

Математическая модель IRT отражает взаимосвязь вопросов теста и латентной характеристики личности. Основными предпосылками использования таких моделей являются:

- существование латентных характеристик респондентов;
- существование доступных для наблюдения факторов, отражающих значения ненаблюдаемых факторов;
- задания, измеряющие такие факторы, независимы для каждого респондента.

В рамках работы используется двухпараметрическая логистическая модель IRT (2PL IRT) [32, 33], увязывающая вероятность положительного ответа на вопрос q_i в дихотомической шкале с латентной характеристикой

θ , имеющей нормальное распределение вероятности, посредством уравнения:

$$P_i(q_i = 1|\theta) = \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}},$$

где a_i – параметр дифференцирующей способности (дискриминативности) и b_i – параметр трудности задания.

В рамках исследования модель 2PL IRT использовалась для формирования шкалы и оценивания латентной характеристики. При формировании шкал использовались те характеристики, для которых показатель дискриминативности выше 0.5, что является эвристическим правилом.

3. Формализация подхода к выявлению характеристик человеческого капитала на основании гетерогенных данных из самоотчетов сотрудников. Математически поставленную задачу выявления характеристик человеческого капитала, которые имеют наибольший вклад в индикаторы эффективности сотрудника в контексте деятельности организации, можно сформулировать следующим образом. Пусть для каждого индивида $i \in 1 \dots n$ измеряется набор взаимосвязанных характеристик человеческого капитала $\{x_i^j, j = 1 \dots K\}$ и некоторый индикатор эффективности сотрудника Y_i . Кроме того, вся ситуация разворачивается в некотором внешнем контексте, влияние которого моделируется с помощью случайных величин $E_t, t = 1 \dots T$. Такие внешние факторы могут отражать как и оценку деятельности сотрудника другими сотрудниками (как при опросе в формате методики 360 градусов), так и иные организационные факторы. Требуется построить статистическую модель взаимосвязи паттернов, т.е. подгрупп среди измеренных характеристик человеческого капитала $\{x_i^j, j = 1 \dots K\}$, и значений индикатора эффективности сотрудника Y_i в контексте $E_t, t = 1 \dots T$, и выявить наиболее значимые для конкретной ситуации факторы. Отметим, что переменные, отражающие контекст, также могут требовать предварительного анализа.

Данные самооценки сотрудниками уровня владения профессиональными навыками являются одним из самых простых и дешевых способов получения информации об этом аспекте человеческого капитала. Такие самооценки могут проводиться как при рутинном опросе сотрудников в системах мониторинга, так и при найме на работу, а также являются составной частью существующих методик оценки человеческого капитала. Среди возможных типов вопросов широко распространены

измерения в пятиступенчатой шкале Ликерта, варьирующейся от уровня «не владею навыком» до «имею профессиональные сертификаты, подтверждающие владений навыком». Отметим, что профессиональные навыки, которые составляют человеческий капитал взаимосвязаны друг с другом и формируют паттерны. Поэтому для анализа такой информации в рамках настоящего подхода предлагается использовать кластерный анализ, результат работы которого обладает высокой интерпретируемостью и позволяют сохранить сложную структуру зависимостей паттерна внутри кластера.

Если говорить о психологической составляющей индивидуального человеческого капитала, то существуют различные шкалы для измерения релевантных характеристик личности, при этом они могут быть специфичными для каждой отдельной области функционирования организации. Поэтому на этапе пилотного исследования бывает важно определить те шкалы, которые релевантны в конкретном контексте: сфере деятельности организации. Для этого предлагается использовать вопросы–утверждения. При этом шкалы могут быть как определены экспертным путем для каждой ситуации, так и быть заранее разработанными. Данные, составленные из ответов на такие вопросы–утверждения, представляют собой набор бинарных взаимозависимых переменных, которые отражают латентную переменную – характеристику личности. Для анализа таких данных предлагается использовать подходы современной теории тестирования.

Итак, предлагаемый подход к сбору и анализу данных, получаемых в результате самоотчетов, для использования в задаче оценки человеческого капитала состоит из следующих этапов.

1. Формирование опросного инструментария, в котором респондентам предлагается оценить профессиональные навыки по пятибалльной шкале.

2. Формирование опросного инструментария о психологических особенностях личностей сотрудников на основе вопросов–утверждений (опирающихся на уже известные психометрические шкалы или же разработанные экспертом).

3. Сбор данных, формирование выборки объема n .

4. Формирование на основании данных самоотчетов о профессиональных навыках k классов респондентов, обладающих сходными паттернами профессиональных навыков: $C_i, i = 1 \dots k$.

5. Формирование на основе данных ответов на вопросы–утверждения p психометрических шкал (возможно, уже существующих),

оценка при помощи методов современной теории тестирования значений латентных переменных $L_j, j = 1 \dots p$.

6. Построение статистической модели, увязывающей вектор выявленных факторов человеческого капитала $\mathbf{X} = (C_1, \dots, C_k, L_1, \dots, L_p)$ с некоторым индикатором эффективности сотрудника, как, например, индикаторы его карьерного успеха, \mathbf{Y} , который может быть измерен как в шкале интервалов, так и в шкале категорий. Для того, чтобы выявить наиболее значимые характеристики, можно использовать модель обобщенной линейной регрессии:

$$E(Y|X) = g(\mathbf{X}\beta).$$

В качестве ключевой характеристики капитала могут выступать различные индикаторы, в том числе индикаторы карьерного успеха [38] или экономические показатели деятельности сотрудника, как например уровень заработной платы.

4. Пилотное исследование подхода на данных опросов государственных гражданских служащих Российской Федерации.

В рамках пилотного исследования был проведен опрос среди государственных гражданских служащих различных структур Российской Федерации, в котором респондентов просили оценить навыки владения различными техническими средствами (как например офисные и аналитические программы, базы данных юридической информации и прочее). Кроме того, респондентам были предложены также 300 вопросов–утверждений, которые легли в основу психометрических шкал для оценки характеристик личности. Были получены данные от 371 респондента.

4.1. Выделение кластеров респондентов со сходными паттернами профессиональных навыков. Для выделения групп использовались следующие данные, релевантные для офисных работников.

1. Самооценка уровня профессиональных навыков: ведения деловой переписки (*переписка*); составления официальных документов (*документы*); сложного форматирования документов (*форматирование*); подготовки презентаций (*презентации*); построения инфографики и сложной визуализации данных (*инфографика*); анализа нормативных правовых актов (*НПА*); систематизации и анализа международного опыта (*международный опыт*); обработки больших объемов данных (*обработка данных*); визуализации и анализа статистической информации (*статанализ*); проведения социологических опросов, интервью

(соцопросы); построение моделей в эконометрических пакетах (эконометрика). Уровень навыка оценивался респондентом по пятибалльной шкале, где 1 – отсутствие навыка, 5 – профессиональный уровень (есть подтверждение о прохождении курсов, действующие сертификаты, соответствующий опыт работы).

2. Самооценка уровня владения программными продуктами: MS Word, MS Excel, MS Access, MS Power Point, MS Visio, MS Project, SPSS, STATA, e-Views, Консультант Плюс, Гарант по шкале, аналогичной п.1.

3. Уровень владения английским языком оценивался по шестибалльной шкале, соответствующей общеевропейской шкале компетенции владения иностранным языком (A1 – начальный уровень; A2 – элементарный уровень; B1 – средний уровень; B2 – средне–продвинутый уровень; C1 – продвинутый уровень; C2 – владение в совершенстве).

Все данные трактовались как порядковые величины. Для простоты интерпретации, вместо пяти уровней владения навыком использовались три: 1 – нет опыта или мало опыта (баллы 1 и 2 по исходной шкале опроса), 2 – владение навыком, не подтвержденное сертификатами (баллы 3 и 4 по исходной шкале), 3 – профессиональное владение, подтвержденное сертификатом (соответствует оценке 5 исходной шкалы). Для определения оптимального числа кластеров были вычислены средние значения коэффициентов силуэтов (формула 1) для различного числа кластеров, анализ позволяет выделить два кластера (рисунок 1).

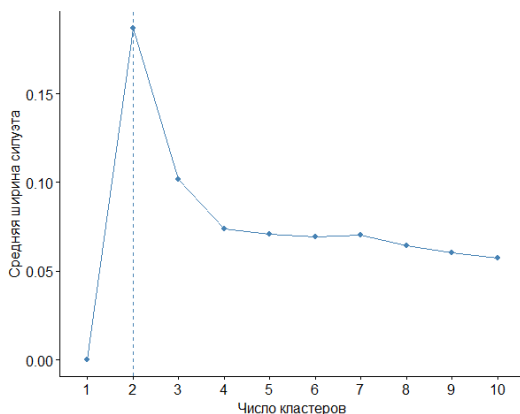


Рис. 1. Средние значения силуэтов для различного числа кластеров при кластеризации алгоритмом k -медоидов, метрика Манхэттен

В таблице 1 представлены медианы и средние значения факторов для двух различных метрик категориальных данных: метрики Манхэттен и корреляционной метрики Кендалла.

Таблица 1. Характеристики выделенных при помощи алгоритма k–медоидов кластеров, метрики Манхэттен и корреляционная метрика. В ячейках указаны медиана (среднее значение) каждой характеристики. Выделены те факторы, которые статистически значимо отличаются между кластерами в отдельной классификации (тест Манна–Уитни, p–значение ниже 0.05)

Характеристики	Классификация I: РАМ, метрика Манхэттен		Классификация II: РАМ, корр. метрика Кендалла	
	кластер I.1	кластер I.2	кластер II.1	кластер II.2
Медоид кластера (номер наблюдения)	196	266	40	57
Переписка	3 (2.74)	2 (2.27)	2 (2.48)	2 (2.34)
Документы	2 (2.36)	2 (1.98)	2 (2.35)	2 (1.8)
Форматирование	2 (2.31)	2 (1.82)	2 (2.1)	2 (1.7)
Презентации	3 (2.87)	2 (2.29)	3 (2.5)	2 (2.46)
Инфографика	2 (2.36)	2 (1.8)	2 (2.1)	2 (1.72)
НПА	2 (2.26)	2 (1.91)	2 (2.17)	2 (1.7)
Международный опыт	2 (2.4)	2 (1.95)	2 (2.25)	2 (1.75)
Обработка данных	3 (2.53)	2 (1.93)	2 (2.25)	2 (1.86)
Статанализ	3 (2.58)	2 (1.99)	2 (2.25)	2 (2.06)
Соопросы	2 (2.4)	2 (2.11)	2 (2.19)	2 (2.26)
Эконометрика	2 (1.76)	1 (1.31)	1 (1.43)	2 (1.52)
MS Word	3 (2.98)	2 (2.34)	3 (2.55)	3 (2.6)
MS Excel	3 (2.65)	2 (2.02)	2 (2.25)	2 (2.19)
MS Access	2 (1.86)	1 (1.21)	1 (1.39)	1 (1.5)
MS Power Point	3 (2.96)	2 (2.28)	3 (2.49)	3 (2.57)
MS Visio	2 (1.69)	1 (1.14)	1 (1.27)	1 (1.43)
MS Project	2 (1.84)	1 (1.20)	1 (1.35)	1 (1.56)
SPSS	2 (1.89)	1 (1.44)	1 (1.53)	2 (1.74)
Stata	1 (1.49)	1 (1.20)	1 (1.24)	1 (1.43)
e-Views	1 (1.2)	1 (1.07)	1 (1.06)	1 (1.21)
Консультант Плюс	3 (2.38)	2 (1.89)	2 (2.2)	1 (1.71)
Гарант	2 (2.19)	2 (1.72)	2 (2.04)	1 (1.48)
Уровень англ.яз.	2 (2.4)	2 (2.22)	2 (2.27)	2 (2.30)

Для валидации выделенных кластеров было проведено имитационное моделирование методом бутстреп, в рамках которого на основе исходного набора данных были сформированы 100 репликаций. На каждом шаге в репликации выборки были выделены два кластера, для которых был рассчитан коэффициент сходства Жаккара с кластерами в исходной классификации [29]. Высокие средние значения коэффициента указывают на стабильность выделения предложенных кластеров. В таблице 2 представлены значения показателей валидации кластеров для

двух рассматриваемых подходов: k-медоидов с метрикой Манхэттен и k-медоидов с корреляционной метрикой Кендалла.

Таблица 2. Значения показателей валидации кластеров для рассматриваемых методов кластеризации

Показатель	Классификация I		Классификация II	
	I.1	I.2	II.1	II.2
Средняя ширина силуэта	0.13	0.19	0.13	0.15
Средний коэффициент Жаккарда, бутстреп, 100 репликаций	0.78	0.89	0.81	0.67

Статистический анализ данных был проведен с помощью системы обработки данных R и пакетов factoextra [20] и fpc [39].

Описание кластеров. Для дальнейшего анализа использовалась классификация I, основанная на метрике Манхэттен, так как различия между двумя группами в этом случае статистически значимы. Классификация II использовалась для подтверждения выводов описания кластеров. Итак, анализ таблицы 1 позволяет выделить следующие особенности двух кластеров.

– Доля респондентов с дипломами, подтверждающими владение заявленными навыками, выше в кластере I.1, чем в кластере I.2, особенно для навыков «переписка», «презентации», «международный опыт», «обработка данных», «статанализ», «соцопросы», «MS Word», «MS Excel», «MS Power Point», «Консультант Плюс».

– Доля респондентов, отмечающих владение навыком без подтверждения сертификатом, выше в кластере I.1, чем в кластере I.2 для навыков: «форматирование», «анализ нормативных правовых актов», «эконометрика», «MS Access», «MS Visio», «MS Project», «SPSS», «Гарант».

– Отметим, что респонденты в кластерах I.1 и I.2 не различаются значительно по уровню владения английским языком (подтверждается классификацией II). Однако не владеют английским языком только респонденты из группы I.2.

– Только респонденты из кластера I.1 обладают профессиональными навыками в области эконометрики (SPSS и e-Views), работы с базами данных (MS Access) и MS Project, а также MS Visio.

– Практически все респонденты из кластера I.1 профессионально владеют продуктами MS Word и MS Excel.

Для иллюстрации таблицы 1 был построен график, представляющий самооценки навыков для каждого индивида (рисунок 2). Чтобы точки не сливались воедино, к значениям 1, 2 и 3 был добавлен небольшой случайный шум, поэтому все точки, расположенные напротив отдельного значения, имеют точно это значение фактора.

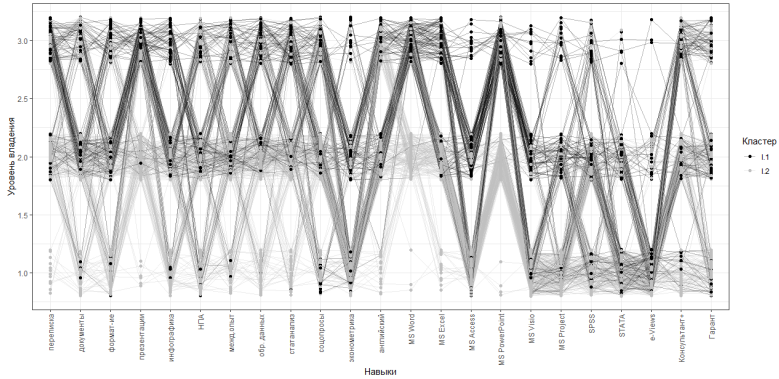


Рис. 2. Значения навыков для групп I.1 и I.2 (зашумленные наблюдения)

В целом можно сказать, что к кластеру I.1 относятся более квалифицированные сотрудники, чем к кластеру I.2, в том числе владеющие узкоспециализированными навыками, как работа в статистических пакетах. В пункте 4.3 построена регрессионная модель, позволяющая определить, какой из двух кластеров может служить фактором карьерного успеха сотрудника.

4.2. Формирование шкал личностных особенностей. В рамках пилотного тестирования респондентам были предложены специально разработанные 300 вопросов–утверждений, предполагающих ответ да или нет. Набор вопросов–утверждений был сформирован на основе следующих психологических шкал: комплексный подход Хоган [40], 16-ти факторный опросник Кеттелла [41], личностный тест Маиерс–Бриггс [42]. Ответы на эти вопросы легли в основу нескольких шкал личностных особенностей, отражающих некоторые характеристики личности, определенные экспертами как важные для государственной гражданской службы. Каждая шкала содержит от 5 до 15 вопросов–утверждений.

Каждая из шкал была проанализирована при помощи подхода IRT, в результате чего была получена оценка латентной (ненаблюдаемой) характеристики личности. Для подгонки регрессионной модели 2PL IRT использовался пакет ltm среды статистической обработки данных R [43].

4.3. Выявление факторов человеческого капитала государственных гражданских служащих в контексте их руководящей позиции. В контексте собранного набора данных в качестве ключевого индикатора карьерного успеха сотрудника использовался факт наличия у него руководящей должности. В выборке присутствуют 156 руководителей и 215 не-руководителей. Для выявления тех факторов человеческого капитала, которые вносят значительный вклад в индикаторы карьерной успешности сотрудника, использовался аппарат логистической регрессии. В таблице 3 представлены исходные значения коэффициентов, их доверительные интервалы и уровень значимости, а также экспонированные коэффициенты для простоты интерпретации в терминах отношения шансов. Отметим, что значимыми являются как паттерны профессиональных навыков, выявленные в рамках кластерного анализа (таблица 1), так и латентные характеристики личностных особенностей. В частности, чем выше у человека значение латентных характеристик «лидерских способностей», «моральных качеств», «амбициозности», тем выше шанс, что этот человек занимает руководящую должность. Кроме того, паттерны профессиональных навыков также могут служить индикатором наличия руководящей должности: если человек хорошо владеет офисными и узкоспециализированными программами, то шанс выше. Однако необходима осторожность при интерпретации последней взаимосвязи, так как она может носить и обратный характер: если сотрудник является руководителем, то он имеет больше возможностей для повышения квалификации.

Так как набор данных невелик для 14 переменных, то для выявления тех коэффициентов, для которых могут быть получены оценки, была использована регрессия по методу лассо. Те коэффициенты, которые отличны от нуля при использовании этого метода, отмечены звездочкой в таблице 3.

Построенная только на значимых переменных модель логистической регрессии позволяет классифицировать поступающие данные с точностью 0.63, 95% доверительный интервал: (0.52, 0.72). Коэффициент корреляции D Соммерса между наблюдаемым значением факта наличия руководящей должности и предсказанными логистической регрессией вероятностями составил 0.362, что говорит об умеренной согласованности этих данных. Вычисленные метрики качества классификации указывают на то, что выбранные факторы индивидуального человеческого капитала находят свое отражение в рассматриваемом индикаторе карьерного успеха, однако необходимо

дополнить модель классификации некоторыми другими более релевантными факторами. Действительно, в контексте карьерного успеха основополагающими могут оказаться пол, возраст или же иные факторы, как например, слабые связи сотрудника [44].

Для получения представленных результатов исходная выборка была разбита на тренировочную, состоящую из 264 наблюдений, на которых была подогнана модель логистической регрессии, и тестовую (107 наблюдений), на которой оценивалось качество классификации.

Таблица 3. Значения коэффициентов логистической регрессии, увязывающей факторы человеческого капитала государственных служащих и их руководящую позицию как индикатор карьерного успеха сотрудника

Переменная	Коэффициент (b)	Уровень значимости	Экспонированный коэффициент (exp(b))
* Принадлежность к кластеру 2	-0.998	< 0.001	0.369
Коммуникативный потенциал	-0.431	0.065	0.650
Когнитивный потенциал	0.015	0.937	1.015
Шкала лжи	-0.302	0.165	0.739
* Лидерские способности	0.724	0.015	2.063
* Моральные качества	0.457	0.089	1.579
Предпочтение стабильности	-0.257	0.279	0.773
* Амбициозность	0.537	0.029	1.711
Целеустремленность	-0.172	0.555	0.842
Эмоциональная устойчивость	-0.323	0.179	
Творческий потенциал	-0.165	0.482	0.724
Уверенность в себе	0.026	0.928	1.026
Внимание к финансовому благополучию	-0.113	0.676	0.893

5. Обсуждение. Предложенный подход к оценке факторов человеческого капитала опирается как на экспертную информацию в плане выявления латентных характеристик личности и интерпретации кластеров, так и на конкретные данные. Эти особенности подхода обусловлены тем, что задача оценки человеческого капитала организации зависит от структуры самой организации, области ее функционирования и ключевых индикаторов эффективности ее деятельности. Поэтому в ряде ситуаций прежде, чем определить, какие шкалы использовать при анкетировании сотрудников, необходимо предварительно выявить те характеристики личности и ключевые навыки, которые играют роль для отдельного индикатора эффективности сотрудника и/или организации. Предлагаемый метод является быстрым и недорогим, так как опирается лишь на опросы сотрудников.

Отметим также, что в зависимости от ситуации различной может быть и опора на данные, полученные из самоотчетов, которые могут обладать свойствами неполноты и неточности. Например, информация о профессиональном владении навыками может быть доступна отделам кадров организации без обращения к самоотчетам.

Личностные особенности являются составной частью человеческого капитала и оказывают влияние как на карьерную успешность сотрудника, так и на эффективность деятельности организации. Выявление характеристик личности, важных для каждой отдельной области функционирования организации, является сложной задачей, требующей пилотных исследований и экспертных знаний. В рамках представленного подхода предлагается экспертным путем сформировать набор шкал, складывающихся из вопросов–утверждений, и выявить те, которые имеют наибольший вклад в ключевые показатели эффективности сотрудника или организации. Дальнейшим этапом исследований следует рассматривать формирование психометрических опросов, направленных на обоснованную оценку латентной характеристики личности, выявленной на этапе пилотного исследования.

Отметим, что предложенный подход может использоваться как для выявления ключевых характеристик индивидуального человеческого капитала, так и для оценки человеческого капитала небольших сходных по выполняемым функциям групп сотрудников. Например, если говорить о компетенциях сотрудников государственных гражданских служащих различных структур Российской Федерации, то наличие руководящей должности связано с самооценкой лидерских способностей, опоры на моральные принципы и амбициозностью: руководители оценивают себя выше по этим пунктам, чем не-руководители. При этом сотрудник, занимающий руководящую должность, в целом обладает большим количеством профессиональных навыков, в паттерне которых определяющим являются проведение социологических опросов, обработка данных и построение инфографики, а также использование узкоспециализированных программ как эконометрические пакеты, MS Access, MS Project.

Предложенный подход имеет ряд ограничений. Во-первых, методы, предлагаемые для выявления факторов человеческого капитала, являются описательными, привязанными к конкретному набору данных. Это ограничение связано с тем, что внутри каждой области функционирования предприятия эти факторы являются уникальными, и потому не существует универсального набора показателей. Предлагаемый подход играет роль

своего рода пилотного, для каждой отдельной области требует экспертного анализа. Во-вторых, в рамках предлагаемого подхода предполагается сбор данных в результате длинного опроса, и потому требуется проведение мероприятий по обеспечению достоверности получаемых ответов.

6. Заключение. В статье был предложен подход к выявлению и моделированию факторов индивидуального человеческого капитала, составляющих наибольший вклад в некоторые показатели эффективности, такие как карьерный успех. Предложенный подход предполагает как формирование пилотного набора данных на основе самоотчетов сотрудников об уровне их профессиональных навыков, так и психометрическое тестирование на основе набора вопросов–утверждений.

На первом этапе предполагается формирование опросного инструментария, охватывающего различные профессиональные навыки, используемые сотрудниками подразделения, и включающего ряд вопросов–утверждений, которые отражают ненаблюдаемые особенности личности. На втором этапе проводится анализ собранных данных. Для анализа первого блока данных предлагается использовать кластерный анализ для выявления интерпретируемых паттернов среди используемых навыков. Ответы на вопросы–утверждения предлагается комбинировать при помощи экспертных знаний в психометрические шкалы с использованием модели 2PL современной теории тестирования. Предложенные методы тесно связаны с конкретным набором данных, и имеют ограниченные возможности распространения на уровень популяции, однако обладают высокой интерпретируемостью и устойчивостью в каждой конкретной задаче.

На финальном этапе при помощи регрессионного анализа (или другого метода статистического вывода) предлагается увязывать оцененные факторы индивидуального человеческого капитала с выбранным показателем эффективности сотрудника, как, например, факт наличия руководящей должности.

Предлагаемый подход может использоваться как при найме на работу, так и при рутинной оценке квалификации сотрудников для выявления наиболее перспективных для выполнения руководящих обязанностей. Среди возможных дальнейших направлений исследования можно отметить включение в подход других показателей человеческого капитала, а также развитие методов комбинации характеристик индивидуального человеческого капитала для оценки человеческого капитала самой организации или ее подразделения.

Литература

1. Dastile X., Celik T., Potsane M. Statistical and machine learning models in credit scoring: A systematic literature survey // *Applied Soft Computing*. 2020. vol. 91. pp. 106263.
 2. Djeundje V.B., Crook J., Calabrese R., Hamid M. Enhancing credit scoring with alternative data // *Expert Systems with Applications*. 2021. vol. 163. pp. 113766.
 3. Абрамов М.В., Тулупьева Т.В., Тулупьев А.Л. Социоинженерные атаки: социальные сети и оценки защищенности пользователей. СПб.: ГУАП, 2018. 266 с.
 4. Олисеенко В.Д., Абрамов М.В., Тулупьев А.Л., Иванов К.А. Прототип программного комплекса для анализа аккаунтов пользователей социальных сетей: веб-фреймворк Django // *Программные продукты и системы*. 2022. Т. 35. № 1. С. 45–53. doi: 10.15827/0236-235X.137.
 5. Khlobystova A., Korepanova A., Maksimov A., Tulupyeva T. An Approach to Quantification of Relationship Types between Users Based on the Frequency of Combinations of Non-numeric Evaluations // *Proceedings of the Fourth International Scientific Conference “Intelligent Information Technologies for Industry” (ITI'19)*. Advances in Intelligent Systems and Computing. 2020. vol. 1156. pp. 206–213.
 6. Kashevnik A., Karelskaya K., Repp M. Dangerous situations determination by smartphone in vehicle cabin: Classification and algorithms // *2019 24th Conference of Open Innovations Association (FRUCT)*. IEEE, 2019. С. 130–139.
 7. Shirmohammadi H., Hadadi F., Saeedian M. Clustering analysis of drivers based on behavioral characteristics regarding road safety // *International Journal of Civil Engineering*. 2019. vol. 17. no. 8. pp. 1327-1340.
 8. Wang X., Xu X. Assessing the relationship between self-reported driving behaviors and driver risk using a naturalistic driving study // *Accident Analysis Prevention*. 2019. vol. 128. pp. 8–16.
 9. Boudreaux M.J., Ferrell B.T., Hundley N.A., Sherman R.A. A personality-based measure of employability // *Journal of Personnel Psychology*. 2022. vol. 21. no. 1. pp. 11–22.
 10. Sharma M., Luthra S., Joshi S., Kumar A. Analysing the impact of sustainable human resource management practices and industry 4.0 technologies adoption on employability skills // *International Journal of Manpower*. 2022. vol. 43. no. 2. pp. 463–485.
 11. Nicolaescu S.S., Florea A., Kifor C.V., Fiore U., Cocan N., Receu I., Zanetti P. Human capital evaluation in knowledge-based organizations based on big data analytics // *Future Generation Computer Systems*. 2020. vol. 111. pp. 654–667.
 12. Wright P.M., McMahan G.C. Exploring human capital: putting 'human' back into strategic human resource management // *Human resource management journal*. 2012. vol. 21. no. 2. pp. 93–104.
 13. Fajaryati N., Akhyar M. The employability skills needed to face the demands of work in the future: Systematic literature reviews // *Open Engineering*. 2020. vol. 10. no. 1. pp. 595–603.
 14. Smaldone F., Ippolito A., Lagger J., Pellicano M. Employability skills: Profiling data scientists in the digital labour market // *European Management Journal*. 2022. vol. 40. no. 5. pp. 671-684.
 15. Ployhart R.E., Moliterno T.P. Emergence of the human capital resource: A multilevel model // *Academy of management review*. 2011. vol. 36. no. 1. pp. 127–150.
 16. Zhang Y., Xu S., Zhang L., Yang M. Big data and human resource management research: An integrative review and new directions for future research // *Journal of Business Research*. 2021. vol. 133. pp. 34–50.
 17. Liu J. Impact of enterprise human capital on technological innovation based on machine learning and SVM algorithm // *Journal of Ambient Intelligence and Humanized Computing*. 2021. pp. 1-13.
- 208 Информатика и автоматизация. 2023. Том 22 № 1. ISSN 2713-3192 (печ.)
ISSN 2713-3206 (онлайн) www.ia.spcras.ru

18. Fleenor J.W., Taylor S., Chappelow C. Leveraging the impact of 360-degree feedback // Berrett-Koehler Publishers, Incorporated. 2020. 184 p.
19. Эфендиев А.Г., Гоголева А.С., Пашкевич А.В., Балабанова Е.С. Ценностно–мотивационные основы и реальность трудовой жизни российских работников: проблемы и противоречия // Мир России. Социология. Этнология. 2020. 29(2). С. 108–133.
20. Kassambara A. Practical guide to cluster analysis in R: Unsupervised machine learning // STHDA. 2017. 187 p.
21. Forsman H., Jansson I., Leksell J., Lepp M., Sundin Andersson C., Engstrom M., Nilsson J. Clusters of competence: Relationship between self-reported professional competence and achievement on a national examination among graduating nursing students // Journal of Advanced Nursing. 2020. vol. 76. no. 1. pp. 199–208. .
22. Schmid M., Brianza E., Petko D. Self-reported technological pedagogical content knowledge (TPACK) of pre-service teachers in relation to digital technology use in lesson plans // Computers in Human Behavior. 2021. vol. 115. pp. 106586.
23. Yang L., Sang-Bing T. Construction of a Hierarchical Neural Network Power Source Model for Human Capital Technology Innovation and Benefit Distribution with Big Data Analysis // Mathematical Problems in Engineering. 2021. vol. 2021. pp. 3939511.
24. Li X., Zhang P. A research on value of individual human capital of high-tech enterprises based on the bp neural network algorithm // The 19th International Conference on Industrial Engineering and Engineering Management. Berlin: Springer, 2013. pp. 71-79.
25. Baron A. Measuring human capital // Strategic HR Review. 2011. vol. 10. no 2. pp. 30–35.
26. Гончарова Е.А., Рукин К.Н. Использование методики ассесмент-центра при оценке государственных гражданских служащих Липецкой области // Государственная служба. 2021. Т. 23. № 3(131). С. 24–32.
27. Литвина С.А., Еварович С.А. Ассесмент-центр как технология оценки компетенций персонала в практике государственного управления: учебное пособие. Томск: Томский государственный университет, 2013. 104 с.
28. Родионова Е.А. Психологические факторы эффективности сотрудников современного предприятия // Общество. Коммуникация. Образование. 2011. Т. 2. № 124. С. 109–114.
29. Hennig C., Meila M., Murtagh F., Rocci R. (Eds.). Handbook of cluster analysis // CRC Press, 2015. 730 p.
30. Schubert E., Rousseeuw P.J. Faster k–medoids clustering: improving the PAM, CLARA, and CLARANS algorithms // International conference on similarity search and applications. Springer, Cham. 2019. pp. 171–187.
31. Hennig C. Dissolution point and isolation robustness: robustness criteria for general cluster analysis methods // Journal of Multivariate Analysis. 2009. vol. 99. pp. 1154–1176.
32. Крокер Л., Алгина Д. Введение в классическую и современную теорию тестов. Учебник. М.: Логос, 2010. 668 с.
33. Van der Linden W.J. Handbook of Item Response Theory, Volume One: Models. Chapman and Hall/CRC, 2016. 624 p.
34. Van der Linden W.J. Handbook of Item Response Theory, Volume Three: Applications. Chapman and Hall/CRC, 2018. 608 p.
35. Lang J.W., Tay L. The science and practice of item response theory in organizations // Annual Review of Organizational Psychology and Organizational Behavior. 2021. vol. 8. pp. 311–338.

36. Cella D., Choi S.W., Condon D.M., Schalet B., Hays R.D., Rothrock N.E., Yount S., Cook K.F., Gershon R.C., Amtmann D., DeWalt D.A. PROMIS® adult health profiles: efficient short-form measures of seven health domains. *Value in health*. 2019. vol. 22. no. 5. pp. 537–544.
37. De Jong M.G., Pieters R. Assessing sensitive consumer behavior using the item count response technique. *Journal of Marketing Research*. 2019. vol. 56. no. 3. pp. 345–360.
38. Abele A.E., Spurk D., Volmer J. The construct of career success: measurement issues and an empirical example // *ZAF*. 2011. vol. 43. pp. 195–206.
39. Hennig C. fpc: Flexible Procedures for Clustering. R package version 2.2-9. 2020. <https://CRAN.R-project.org/package=fpc>.
40. Hogan R. Hogan development survey manual. Tulsa, OK: Hogan Assessment Systems, 2009. 199 p.
41. Cattell R.B., Cattell H.E.P. Personality structure and the new fifth edition of the 16PF // *Educational and Psychological Measurement*. 1995. vol. 55. no. 6. pp. 926–937.
42. Myers I.B. *The Myers-Briggs Type Indicator: Manual*. Consulting Psychologists Press, 1962. 110 p.
43. Rizopoulos D. An R package for Latent Variable Modelling and Item Response Theory Analyses // *Journal of Statistical Software*. 2006. vol. 17. no. 5. pp. 1–25.
44. Granovetter M. The strength of weak ties: A network theory revisited // *Sociological theory*. 1983. pp. 201–233.

Столярова Валерия Фуатовна — младший научный сотрудник, лаборатория теоретических и междисциплинарных проблем информатики, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: анализ данных, байесовские сети доверия, искусственный интеллект. Число научных публикаций — 40. vfs@dscs.pro; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-3337.

Тулупьева Татьяна Валентиновна — канд. психол. наук, доцент, старший научный сотрудник, лаборатория теоретических и междисциплинарных проблем информатики, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН); профессор, Северо-Западный институт управления РАНХиГС. Область научных интересов: психология личности, искусственный интеллект, социальная инженерия, методы обработки данных. Число научных публикаций — 170. tvt@dscs.pro; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-3337.

Абрамов Максим Викторович — канд. техн. наук, руководитель лаборатории, лаборатория теоретических и междисциплинарных проблем информатики, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: информационная безопасность, социоинженерные атаки, социокompьютинг. Число научных публикаций — 170. mva@dscs.pro; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-3337.

Салахова Валентина Борисовна — канд. психол. наук, ведущий научный сотрудник, Центр исследования проблем безопасности РАН; ведущий научный сотрудник, лаборатория гуманной педагогики, Московский городской педагогический университет (МГПУ). Область научных интересов: психология девиантного поведения, психология личности, цифровизация образования. Число научных публикаций — 158. Valentina_naula@mail.ru; улица Гарибальди, 21Б, 117335, Москва, Россия; р.т.: +7(985)221-4051.

V. STOLIAROVA , T. TULUPYEVA , M. ABRAMOV , V. SALAKHOVA
**IDENTIFICATION OF CHARACTERISTICS OF EMPLOYEE'S
INDIVIDUAL HUMAN CAPITAL WITH DATA ON SELF-REPORTS
OF PROFESSIONAL SKILLS AND PERSONAL CHARACTERISTICS**

Stoliarova V., Tulupyeva T., Abramov M., Salakhova V. **Identification of Characteristics of Employee's Individual Human Capital with Data on Self-Reports of Professional Skills and Personal Characteristics.**

Abstract. In the field of recruitment and human resources management, the problem arises of automatization of the assessment process of the characteristics of human capital, taking into account, among other things, the personality characteristics of the employee. The article is devoted to the problem of identification of such characteristics that have the greatest contribution to some indicators of the effectiveness of an employee of an organization with self-reported data on professional skills and answers to questions—statements about various psychological aspects of personality. The general structure of the survey tools based on self-reports of employees is proposed, as well as the formalization of the proposed methods of data analysis. The cluster analysis was used for the identification of groups with similar professional skills. Special psychometric scales based on the questions—statements are selected and analyzed via the item response theory approach, giving the estimates of the latent variable, that reflects personal characteristics. At the final stage of the study, the relationship between the estimated factors (identified clusters and estimated latent variables) and the indicator of employee effectiveness was assessed. As such indicator, the fact of a managerial position was used. The proposed approach is a structure of a pilot study that allows to identify the characteristics of human capital (professional skills and personality traits) that have the greatest contribution to the performance indicators of an employee or organization, and is aimed at reducing labor costs at subsequent stages of a more detailed and targeted study. The possibilities of the proposed approach are demonstrated with data collected among state civil servants in Russia. The fact of having a managerial position is used as an indicator of effectiveness.

Keywords: human capital assessment, cluster analysis, item response theory, questionnaire design, self-report.

References

1. Dastile X., Celik T., Potsane M. Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*. 2020. vol. 91. pp. 106263.
2. Djeundje V.B., Crook J., Calabrese R., Hamid M. Enhancing credit scoring with alternative data. *Expert Systems with Applications*. 2021. vol. 163. pp. 113766.
3. Abramov M.V., Tulupyeva T.V., Tulupye A.L. Socioinzhenernye ataki: social'nye seti i ochenki zashhishhennosti pol'zovatelej [Socialengineering attacks: social media and user security assessment]. SPb.: GUAP, 2018. 266 p. (in Russ.).
4. Oliseenko V.D., Abramov M.V., Tulupye A.L., Ivanov K.A. [A software package prototype for analyzing user accounts insocial networks: Django web framework]. *Programmnye produkty i sistemy – Software Systems*. 2022. vol. 35. no. 1. pp. 45–53. doi: 10.15827/0236-235X.137. (In Russ.).
5. Khlobystova A., Korepanova A., Maksimov A., Tulupyeva T. An Approach to Quantification of Relationship Types between Users Based on the Frequency of

- Combinations of Non-numeric Evaluations. Proceedings of the Fourth International Scientific Conference «Intelligent Information Technologies for Industry» (ITI' 19). Advances in Intelligent Systems and Computing. 2020. vol. 1156. pp. 206–213.
6. Kashevnik A., Karelskaya K., Repp M. Dangerous situations determination by smartphone in vehicle cabin: Classification and algorithms. 2019 24th Conference of Open Innovations Association (FRUCT). IEEE, 2019. pp. 130–139.
 7. Shirmohammadi H., Hadadi F., Saeedian M. Clustering analysis of drivers based on behavioral characteristics regarding road safety. International Journal of Civil Engineering. 2019. vol. 17. no. 8. pp. 1327–1340.
 8. Wang X., Xu X. Assessing the relationship between self-reported driving behaviors and driver risk using a naturalistic driving study. Accident Analysis Prevention. 2019. vol. 128. pp. 8–16.
 9. Boudreaux M.J., Ferrell B.T., Hundley N.A., Sherman R.A. A personality-based measure of employability. Journal of Personnel Psychology. 2022. vol. 21. no. 1. pp. 11–22.
 10. Sharma M., Luthra S., Joshi S., Kumar A. Analysing the impact of sustainable human resource management practices and industry 4.0 technologies adoption on employability skills. International Journal of Manpower. 2022. vol. 43. no. 2. pp. 463–485.
 11. Nicolaescu S.S., Florea A., Kifor C.V., Fiore U., Cocan N., Receu I., Zanetti P. Human capital evaluation in knowledge-based organizations based on big data analytics. Future Generation Computer Systems. 2020. vol. 111. pp. 654–667.
 12. Wright P.M., McMahan G.C. Exploring human capital: putting 'human' back into strategic human resource management. Human resource management journal. 2012. vol. 21. no. 2. pp. 93–104.
 13. Fajaryati N., Akhyar M. The employability skills needed to face the demands of work in the future: Systematic literature reviews. Open Engineering. 2020. vol. 10. no. 1. pp. 595–603.
 14. Smaldone F., Ippolito A., Lagger J., Pellicano M. Employability skills: Profiling data scientists in the digital labour market. European Management Journal. 2022. vol. 40. no. 5, pp. 671-684.
 15. Ployhart R.E., Moliterno T.P. Emergence of the human capital resource: A multilevel model. Academy of management review. 2011. vol. 36. no. 1. pp. 127–150.
 16. Zhang Y., Xu S., Zhang L., Yang M. Big data and human resource management research: An integrative review and new directions for future research Journal of Business Research. 2021. vol. 133. pp. 34–50.
 17. Liu J. Impact of enterprise human capital on technological innovation based on machine learning and SVM algorithm. Journal of Ambient Intelligence and Humanized Computing. 2021. pp. 1–13.
 18. Fleenor J.W., Taylor S., Chappelow C. Leveraging the impact of 360-degree feedback. Berrett-Koehler Publishers, Incorporated, 2020. 184 p.
 19. Efendiev A.G., Gogoleva A.S., Pashkevitch A.V., Balabanova E.S. [The values, Motives and Work life reality of Russian employees: problems and contradictions]. Mir Rossii. Sociologija. Jetnologija – Universe of Russia Sociology. Ethnology. 2020. vol. 29. no. 2. pp. 108–133. (In Russ.).
 20. Kassambara A. Practical guide to cluster analysis in R: Unsupervised machine learning. STHDA, 2017. 187 p.
 21. Forsman H., Jansson I., Leksell J., Lepp M., Sundin Andersson C., Engstrom M., Nilsson J. Clusters of competence: Relationship between self-reported professional competence and achievement on a national examination among graduating nursing students. Journal of Advanced Nursing. 2020. vol. 76. no. 1. pp. 199–208.

22. Schmid M., Brianza E., Petko D. Self-reported technological pedagogical content knowledge (TPACK) of pre-service teachers in relation to digital technology use in lesson plans. *Computers in Human Behavior*. 2021. vol. 115. pp. 106586.
23. Yang L., Sang-Bing T. Construction of a Hierarchical Neural Network Power Source Model for Human Capital Technology Innovation and Benefit Distribution with Big Data Analysis. *Mathematical Problems in Engineering*. 2021. vol. 2021. pp. 3939511.
24. Li X., Zhang P. A research on value of individual human capital of high-tech enterprises based on the bp neural network algorithm. *The 19th International Conference on Industrial Engineering and Engineering Management*. Berlin: Springer, 2013. pp. 71-79.
25. Baron A. Measuring human capital. *Strategic HR Review*. 2011. vol. 10. no 2. pp. 30–35.
26. Goncharova E.A., Rukin K.N. [Using the assessment center methodology while evaluating the civil servants of the Lipetsk region]. *Gosudarstvennaja sluzhba – Public Administration*. 2021. vol. 23. no. 3(131). pp. 24–32. (in Russ).
27. Litvina S.A., Yevarovich S.A. Assessment-centr kak tehnologiya ocenki kompetencii personala v praktike gosudarstvennogo upravleniya [Assessment Center as a technology for assessing the competencies of personnel in the practice of public administration]. *Tomsk: Tomsk State University*, 2013. 104 p. (in Russ).
28. Rodionova E.A. Psychological factors of efficiency of employees of a modern enterprise. *Obshhestvo. Kommunikacija. Obrazovanie – Society. Communication. Education*. 2011. vol. 2. no. 124. pp. 109–114. (in Russ).
29. Hennig C., Meila M., Murtagh F., Rocci R. (Eds.). *Handbook of cluster analysis*. New York: CRC Press, 2015. 773 p.
30. Schubert E., Rousseeuw P.J. Faster k-medoids clustering: improving the PAM, CLARA, and CLARANS algorithms. *International conference on similarity search and applications*. Springer, Cham. 2019. pp. 171–187.
31. Hennig C. Dissolution point and isolation robustness: robustness criteria for general cluster analysis methods. *Journal of Multivariate Analysis*. 2009. vol. 99. pp. 1154–1176.
32. Krockner L., Algina D. Vvedenie v klassicheskuyu i sovremennuyu teoriyu testov [An introduction to the classical and contemporary testing theory]. M.: Logos, 2010. 668 p. (in Russ.).
33. Van der Linden W.J. *Handbook of Item Response Theory, Volume One: Models*. Chapman and Hall/CRC, 2016. 624 p.
34. Van der Linden W.J. *Handbook of Item Response Theory, Volume Three: Applications*. Chapman and Hall/CRC, 2018. 608 p.
35. Lang J.W., Tay L. The science and practice of item response theory in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*. 2021. vol. 8. pp. 311–338.
36. Cella D., Choi S.W., Condon D.M., Schalet B., Hays R.D., Rothrock N.E., Yount S., Cook K.F., Gershon R.C., Amtmann D., DeWalt D.A. PROMIS® adult health profiles: efficient short-form measures of seven health domains. *Value in health*. 2019. vol. 22. no. 5. pp. 537–544.
37. De Jong M.G., Pieters R. Assessing sensitive consumer behavior using the item count response technique. *Journal of Marketing Research*. 2019. vol. 56. no. 3. pp. 345–360.
38. Abele A.E., Spurk D., Volmer J. The construct of career success: measurement issues and an empirical example. *ZAF*. 2011. vol. 43. pp. 195–206.
39. Hennig C. fpc: Flexible Procedures for Clustering. R package version 2.2-9. 2020. Available at: <https://CRAN.R-project.org/package=fpc>. (accessed 03.11.2022).
40. Hogan R. *Hogan development survey manual*. Tulsa, OK: Hogan Assessment Systems, 2009. 199 p.

41. Cattell R.B., Cattell H.E.P. Personality structure and the new fifth edition of the 16PF. Educational and Psychological Measurement. 1995. vol. 55. no. 6. pp. 926–937.
42. Myers I.B. The Myers-Briggs Type Indicator: Manual. Consulting Psychologists Press, 1962. 110 p.
43. Rizopoulos D. An R package for Latent Variable Modelling and Item Response Theory Analyses. Journal of Statistical Software. 2006. vol. 17. no. 5. pp. 1–25.
44. Granovetter M. The strength of weak ties: A network theory revisited. Sociological theory. 1983. pp. 201–233.

Stoliarova Valerie — Junior researcher, Laboratory of theoretical and interdisciplinary problems of computer science, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: data analysis, probabilistic graphical models. The number of publications — 40. vfs@dscs.pro; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-3337.

Tulupyeva Tatiana — Ph.D., Associate Professor, Senior researcher, Laboratory of theoretical and interdisciplinary problems of computer science, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS); Professor, State and Municipal Management at the North-West Institute of management (NWIM), the branch of the Russian Presidential Academy of National Economy and Public Administration. Research interests: personal psychology, artificial intelligence, social engineering, data analysis. The number of publications — 170. tvt@dscs.pro; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-3337.

Abramov Maxim — Ph.D., Head of the laboratory, Laboratory of theoretical and interdisciplinary problems of computer science, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: information security, social engineering attacks, social computing. The number of publications — 170. mva@dscs.pro; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-3337.

Salakhova Valentina — Ph.D., Leading researcher, Center research security problems the RAS; Leading researcher, Laboratory of humanistic approach in education, Moscow City University. Research interests: psychology of deviant behavior, personal psychology, digitalization of education. The number of publications — 158. Valentina_naula@mail.ru; 21Б, Garibaldi St., 117335, Moscow, Russia; office phone: +7(985)221-4051.