МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ

УДК 004.932

Г.А. Кравцов, канд. техн. наук
Ин-т проблем моделирования
в энергетике им. Г.Е. Пухова НАН Украины
(Украина, 03164, Киев, ул. Генерала Наумова, 15,
email: hryhoriy.kravtsov@gmail.com)

Вычисления на классификациях. Подтверждение квалификации в социальных сетях

Операторы крупнейших социальных сетей констатируют наличие проблем при подтверждении квалификации экспертов. Предложен подход, основанный на теории вычислений на классификациях, позволяющий минимизировать субъективизм при самооценке и взаимном оценивании. Дана номинальная оценка квалификации и структуры подтверждения. Совокупность номинальной оценки и структуры подтверждения позволяет применить теорию вариабельности для прагматичного подбора экспертов.

Ключевые слова: классификация, эксперт, квалификация, подтверждение, субъективизм, объективизм, актуализация, оценка.

Оператори найбільших соціальних мереж констатують наявність проблем при підтвердженні кваліфікації експертів. Запроповано підхід, базований на теорії обчислень на класифікаціях, що дозволяє мінімізувати суб'єктивізм при самооцінюванні та взаємному оцінюванні. Дано номінальну оцінку кваліфікації та структури підтвердження. Сукупність номінальної оцінки та структури підтвердження дозволяє застосувати теорію варіабельності для прагматичного підбору експертів.

Ключові слова: класифікація, експерт, кваліфікація, підтвердження, суб'єктивізм, об'єктивізм, актуалізація, оцінка.

Постановка задачи. Во многих публикациях, посвященных проблеме подтверждения квалификации в социальных сетях [1, 2], раскрывается значимость подтверждений эксперта в процессе выбора [3]. В результате анализа открытых публикаций установлено, что несмотря на важность указанной проблемы она до сих пор остается нерешенной. Сформулируем общую постановку задачи.

Пусть существует экспертное сообщество E из N экспертов (мощность множества экспертов). Пусть задана некоторая классификация квалификаций A, в соответствии с которой эксперты могут оценивать уровень квалификации друг друга, а также давать оценку своей собственной квалификации (самооценка). Будем полагать, что все оценки $R_{k,m}$ субъективно находятся в интервале $(0,1]$, где $R_{k,m}$ — оценка экспертом E_k квалификации эксперта E_m (следовательно, $R_{k,k}$ — самооценка эксперта E_k). Необходимо найти подход (эмпирический алгоритм), позволяющий оценить уровень квалификации эксперта. При этом оценка должна быть максимально объективной, интерпретируемой и используемой на практике [4].

Корректное решение данной задачи может быть получено на основе модели вычислений на классификациях, предложенной в работе [4]. В терминах разрабатываемой теории сформулированная задача имеет следующий вид.

Требуется определить математически обоснованную «интегральную» оценку квалификации $\overline{S}\,(E_j, A_I^i, t)$ произвольно выбранного j-го эксперта E_j в произвольно выбранном классе квалификации A_I^i на момент времени t, где согласно [4] A_I^i означает некоторый класс в плоскости деления i классификации квалификаций A. Путь уточнения I однозначно определяет путь в графе от самого общего класса классификации до некоторого уточнения A_I^i с учетом того, что эксперты оценивают друг друга в заранее неизвестном порядке и оценка любого эксперта-оценщика носит исключительно субъективный характер.

Анализ и решение задачи. Первая сложность в решении поставленной задачи — размерность. Интуитивно понятно, что возможность оценивать любым экспертом квалификацию любого другого эксперта на множестве классов классификации квалификации приводит к гипер-кубам. Однако, если ограничиться исключительно одним классом классификации, то все оценки экспертов $R_{k,m}$ могут быть представлены квадратной матрицей размера $N \times N$. О наполнении матрицы судить сложно, так как невозможно заведомо сказать, является матрица разряженной или плотной. Иными словами, можно утверждать, что наилучшая оценка вычислительной сложности решения поставленной задачи для одного класса классификации есть $\Omega(N^2)$, т.е. асимптотическая нижняя граница [5]. Поэтому задача уменьшения размерности матрицы взаимных оценок для одного класса классификации имеет принципиальное значение.

В настоящее время получила широкую известность теория шести рукопожатий, согласно которой два любых человека на земле разделены не более чем пятью уровнями общих знакомых. Исследования ArtenMilner [6] показали, что профессиональные сообщества имеют диаметр социального графа, равный трем. Диаметр социального графа в социологии и математике имеет синонимы: коллаборативное расстояние и число Эрдеша (Erdos number). Таким образом, чтобы расчет уровня компетенции некоторого эксперта был научно обоснован, необходимо, как следует из [6], учитывать взаимные оценки всех экспертов из одного профессионального сообщества, которых разделяет не более чем три уровня оценок.

В то же время, число Эрдеша не дает ответ на вопрос о размерности матрицы взаимооценок $R_{k,m}$. Казалось, на помощь может прийти число Данбара [7] — максимальное число людей, с которыми человек может поддерживать стабильные социальные отношения. Оно равно приблизительно 150. Однако, число Данбара позволяет ответить на вопрос, скольким экспертам произвольно выбранный эксперт может подтвердить уровень квалификации, но не дает ответа на вопрос, сколько экспертов могут подтвердить уровень квалификации одного и того же эксперта. Следовательно, число Данбара можно рассматривать лишь как фактор выявления нетипичного поведения пользователя относительно компетенции других пользователей. Однако непонятно, как с помощью числа Данбара преподаватель высшей школы знаний может оценить своих учеников, когда в год на потоке учится несколько сотен студентов.

Взвешенная оценка эксперта. Если интегральная оценка $\overline{S}(E_j, A_I^i, t)$ квалификации A_I^i эксперта E_j существует и эксперт E_j подтверждает квалификацию эксперта E_k в области A_Y^i с оценкой $R_{j,k}$, то вычислить субъективную среднегеометрическую взвешенную оценку можно по формуле

$$S_j(E_k, A_Y^i, t) =$$
$$= \sqrt[3]{\min(\overline{S}(E_j, A_I^i, t), R_{j,k})\, \overline{S}(E_j, A_I^i, t)(1-\overline{O}(A_I^i, A_Y^i))\, R_{j,k}}. \qquad (1)$$

Формула (1) является субъективной среднегеометрической взвешенной оценкой квалификации A_I^i эксперта E_k, которая рассчитана на основании оценки экспертом E_j с учетом интегральной оценки эксперта-оценщика в классе квалификации A_I^i и скорректирована с помощью коэффициента подобия классов, равного $1-\overline{O}(A_I^i, A_Y^i)$, где $\overline{O}(A_I^i, A_Y^i)$ — строгая математическая мера отличия двух классов, A_I^i и A_Y^i, в рамках классификации A^i [4],

$$\overline{O}(A_I^i, A_Y^i) = 1 - \frac{R(A, A_I^i, A_Y^i)+1}{R(A, A_I^i, A_Y^i)+R(A_I^i, A_I^i A_Y^i)+R(A_Y^i, A_I^i A_Y^i)+1}.$$

Здесь $R(A_I^i, A_Y^i)$ — относительное расстояние, где I, Y — произвольные пути уточнения (деления) классификации A^i. Предлагаемая формула (1) определяет зависимость частной оценки от оценки в ближайшей области квалификации эксперта-оценщика и показывает, какую оценку он выставляет с учетом коэффициента подобия областей квалификации.

Покажем корректность (1) при изолированной самооценке (т.е. эксперт оценивает сам себя и больше никто его не оценивает). Поскольку $\overline{S}\,(E_j, A_I^i, t)$ является оценкой эксперта E_j другими, а оценивал он самого себя, то

$$\overline{S}(E_j, A_I^i, t) = R_{j,j}\ , \tag{2}$$

а следовательно, $S_j(E_k, A_Y^i, t) = \sqrt{R_{j,j} R_{j,j}} = R_{j,j}$, так как $1-\overline{O}(A_I^i, A_I^i) = 1$ в силу $\overline{O}(A_I^i, A_I^i) = 0$ [4]. Полученное равенство (2) показывает, что субъективная среднегеометрическая взвешенная оценка квалификации при изолированной самооценке равна исходной субъективной оценке. Однако согласно формуле (1) требуется, чтобы на момент времени *t* эксперт-оценщик либо был оценен другим оценщиком, либо провел самооценку. Пусть некоторый эксперт E_j с квалификацией A_I^i и $\overline{S}\,(E_j, A_I^i, t_0) = 1{,}0$ дал субъективную оценку $R_{j,k} = 0{,}3$ квалификации A_I^i эксперта E_k, для которого $\overline{S}(E_k, A_I^i, t_0) = 0{,}5$ (например, самооценка). Выполним расчет $S_j(E_k, A_Y^i, t_1)$:

1) поскольку происходит оценка между экспертами с одинаковой квалификацией A_I^i, то $1-\overline{O}\,(A_I^i, A_I^i) = 1$, так как $\overline{O}\,(A_I^i, A_I^i) = 0$ [4];

2) $\min(\overline{S}\,(E_j, A_I^i, t), R_{j,k}) = \min(1{,}0; 0{,}3) = 0{,}3$.

Следовательно, для рассматриваемого случая $S_j(E_k, A_Y^i, t) = \sqrt[3]{0{,}3 \cdot 1{,}0 \cdot 1{,}0 \cdot 0{,}3} = \sqrt[3]{0{,}09} = 0{,}4481$, т.е. субъективная среднегеометрическая взвешенная оценка по предложенной формуле (1) при оценке более квалифицированным экспертом скорректирована в большую сторону, что указывает на завышение требований со стороны более квалифицированного эксперта. Несложно заметить, что возможна обратная ситуация: при попытке оценить более высококвалифицированного эксперта менее квалифицированным субъективная среднегеометрическая взвешенная оценка будет близка к оценке, полученной менее квалифицированным экспертом.

Объективизм оценки. Пусть дано множество субъективных среднегеометрических взвешенных оценок $S_j(E_k, A_Y^i, t)$ квалификации A_I^i эксперта E_k, где *j* — множество экспертов-оценщиков. Интегральная оценка квалификации $\overline{S}(E_k, A_I^i, t)$ произвольно выбранного эксперта E_k в произвольно выбранном классе квалификации A_I^i на момент времени *t* определяется как медиана случайной величины, состоящей из частных среднегеометрических взвешенных оценок $S_j(E_k, A_Y^i, t)$:

$$\overline{S}\,(E_k, A_I^i, t) = \operatorname{median}\{S_l(E_k, A_Y^i, t), S_m(E_k, A_Y^i, t), \ldots, S_j(E_k, A_Y^i, t)\}, \tag{3}$$

где *l*, *m*, *j* — идентификаторы экспертов-оценщиков. Иными словами, полагая субъективные оценки случайными величинами, выбираем такую оценку, которая соответствует медиане случайной величины. Предлагаемый подход к объективизации является развитием подхода В. Тоценко [8] к использованию спектральной характеристики, но более экономичен относительно вычислительных ресурсов.

Актуализация оценки. Для каждого класса квалификации необходимо определить период актуализации, т.е. период времени от настоящего момента назад, такой, что оценки $S_l(E_k, A_Y^i, t), S_m(E_k, A_Y^i, t), ..., S_j(E_k, A_Y^i, t)$ в (3) будут актуальны. Согласно [7] при расчете интегральной оценки не имеет смысла учитывать более чем определяемое числом Данбара количество последних (актуальных) оценок (150), если они не выходят за пределы периода актуализации, что позволяет минимизировать требования по вычислительной мощности.

Под актуализацией будем понимать возможность дополнительной переоценки квалификации A_I^i эксперта E_j экспертом-оценщиком E_i. При повторной оценке сохранять предыдущее ее значение целесообразно только для истории оценивания.

Свойства субъективных среднегеометрических взвешенных оценок. Известно, что геометрическое среднее меньше или равно арифметическому среднему, если величины выбираются из некоторого интервала положительных чисел, что пригодно для оценок и подразумевается далее. Следует заметить, что среднее геометрическое и среднее арифметическое совпадают, если выбраны два одинаковых числа, т.е. для двух минимальных или двух максимальных оценок среднее геометрическое равно среднему арифметическому при соответственно минимальном или максимальном значении. Следовательно, если интегральная оценка эксперта-оценщика равна поставленной им оценке, то частная среднегеометрическая взвешенная оценка квалификации другого эксперта будет равна выставленной оценке.

Валидация. Очевидно, что эксперт может выставить себе наибольшую оценку при самооценке и его интегральная оценка будет максимальной. Можно ли доверять такой оценке? Очевидно, нет. Необходима модель, которая бы учитывала коллаборативное расстояние, подтвержденное измерениями. Пусть $D(E_i)$ — диаметр социального графа эксперта E_i. При этом $D(E_i) \geq 0$. Будем полагать, что при расчете коллаборативного расстояния необходимо знать число путей подтверждения квалификации длиной, не превышающей числа Эрдеша для профессиональных сообществ [6].

Примеры. До тех пор, пока в системе нет ни одного подтверждения, все диаметры социальных графов равны нулю. Поэтому начнем рассматривать ситуацию с первого подтверждения. В табл. 1 представлена динамика графа подтверждений, где цифрами в кружках обозначены номера экспертов.

Предлагаем следующую структуру хранения подтверждений: начало пути (Start), конец пути (End), промежуточные точки (Across), длина пути (L). Будем следовать таким правилам.

Алгоритм.

1. Подтверждение от эксперта-оценщика E_k эксперту E_i регистрируется только один раз.

2. Если добавляется подтверждение от эксперта E_k эксперту E_i , то из табл. 1 выбираются записи, для которых справедливы следующие утверждения:

2.1. Start указывает на эксперта E_i .

2.2. $L < 3$.

2.3. End отличен от E_k.

2.4. В промежуточных точках Across не упоминается E_k.

2.5. На каждую выбранную запись создается новая запись, в которую относительно выбранной вносятся следующие изменения:

2.5.1. Длина пути L увеличивается на единицу.

2.5.2. Start выбранной записи добавляется в список Across в новой записи.

2.5.3. Start указывает на эксперта-оценщика E_k.

2.6. Для каждой новой создаваемой записи (п. 2.5.) выбираются уже существующие записи, у которых:

2.6.1. End указывает на эксперта E_i .

2.6.2. Start отличен от Start создаваемой новой записи (п. 2.5).

2.6.3. Значение L меньше числа Эрдеша, уменьшенного на L создаваемой записи (п. 2.5) и увеличенного на 1.

2.6.4. Start записи не значится в промежуточных точках Across новой записи (п. 2.5).

2.6.5. Across искомой записи не содержит End записи (п. 2.5).

2.7. Каждая извлекаемая запись (п. 2.6) порождает дополнительную запись, в которой:

2.7.1. Start указывает на Start согласно п. 2.6.

2.7.2. End указывает на End записи согласно п. 2.5.

2.7.3. Промежуточные точки получаются объединением промежуточных точек п. 2.6 плюс End п. 2.6. и плюс промежуточные точки согласно п. 2.5.

2.7.4. $L = L$ п. 2.5 плюс L п. 2.6.

Работа предложенного алгоритма представлена в табл. 1

В приведенном примере показано последовательное включение экспертов в процесс оценки в пределах одной условной группы.

Рассмотрим следующий пример: существуют две группы оценивания и происходит оценка эксперта из одной группы экспертом из другой группы. В этом случае работа алгоритма представлена в табл. 2.

Таблица 1

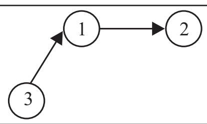
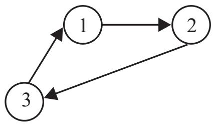
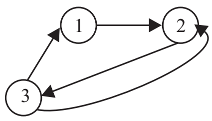
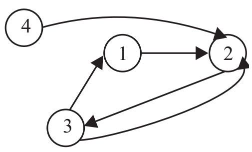
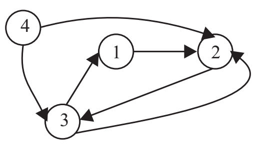
Граф	Номер записи	Start	End	Across	L
	1	1	2	—	1
	1	1	2	—	1
	2	3	2	1	2
	3	3	1	—	1
	1	1	2	—	1
	2	3	2	1	2
	3	3	1	—	1
	4	2	1	3	2
	5	2	3	—	1
	1	1	2	—	1
	2	3	2	1	2
	3	3	1	—	1
	4	2	1	3	2
	5	2	3	—	1
	6	3	2	—	1
	1	1	2	—	1
	2	3	2	1	2
	3	3	1	—	1
	4	2	1	3	2
	5	2	3	—	1
	6	3	2	—	1
	7	4	1	2,3	3
	8	4	3	2	2
	9	4	2	—	1
	1	1	2	—	1
	2	3	2	1	2
	3	3	1	—	1
	4	2	1	3	2
	5	2	3	—	1
	6	3	2	—	1
	7	4	1	2,3	3
	8	4	3	2	2
	9	4	2	—	1
	10	4	1	3	2
	11	4	2	3	2
	12	4	2	3,1	3
	13	4	3	—	1

Таблица 2

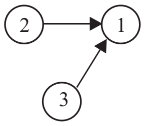
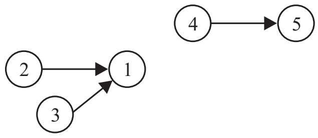
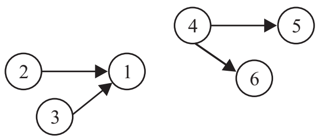
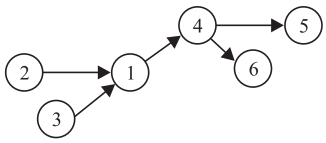
Граф	Номер записи	Start	End	Across	*L*
	1	2	1	—	1
	1	2	1	—	1
	2	3	1	—	1
	1	2	1	—	1
	2	3	1	—	1
	3	4	5	—	1
	1	2	1	—	1
	2	3	1	—	1
	3	4	5	—	1
	4	4	6	—	1
	1	2	1	—	1
	2	3	1	—	1
	3	4	5	—	1
	4	4	6	—	1
	5	1	5	4	2
	6	1	6	4	2
	7	1	4	—	1
	8	2	5	1,4	3
	9	3	5	1,4	3
	10	2	6	1,4	3
	11	3	6	1,4	3
	12	2	4	1	2
	13	3	4	1	2

Таблица 3

Эксперт	Число входящих подтверждений длиной		
	1 (L_1)	2 (L_2)	3 (L_3)
1	1	2	1
2	3	2	1
5 (до объединения)	1	0	0
5 (после объединения)	1	1	2

Существенным достоинством предложенного алгоритма является факт, что записи не модифицируются, а только добавляются новые. Алгоритм может быть легко модифицирован для более длинных путей подтверждения внесением изменений в условие $L < 3$ (см. табл. 2).

Теперь можно легко подсчитать число путей подтверждения с длинами 1, 2 и 3 для экспертов 1 и 2 (см. табл.1) и для эксперта 5 (см. табл. 2). Табл. 3 дает четкое представление о структуре подтверждения квалификации до объединения групп и после объединения. Так, например, очевидно, что все подтверждения эксперта 1 прошли через одно непосредственное подтверждение.

Выводы

Предлагаемый подход позволяет пересчитать интегральную оценку $\overline{S}(E_j, A_I^i, t)$ квалификации A_I^i эксперта E_j в момент подтверждения экспертом-оценщиком, что значительно сокращает потребность в вычислительных ресурсах. К очевидным недостаткам предлагаемой модели следует отнести требование, чтобы на момент оценивания другого эксперта эксперт-оценщик был оценен другим экспертом, либо выполнил самооценку. Однако указанный недостаток не является критичным.

Предложенный алгоритм построения статистики подтверждений дает наглядное представление о структуре подтверждения квалификации. Более того, появляется возможность ответить на вопросы, кто и как принимал участие в подтверждении квалификации выбранного эксперта, что соответствует требованиям интерпретируемости. Приведенные примеры были расчитаны с помощью несложной программы, поэтому можно утверждать, что предложенный подход легко реализуем.

Предложенный подход основан на результатах известных исследований, таких как число Эрдеша для ограничения длины цепочки подтверждений в профессиональной среде и число Данбара для ограничения размера выборки субъективных среднегеометрических взвешенных оценок при актуализации интегральной оценки эксперта.

Полученные метрики для оценки квалификации экспертов позволяют применить теорию вариабельности [9] для прагматичного подбора экспертов, а так же для их мониторинга.

СПИСОК ЛИТЕРАТУРЫ

1. *Morrison M.* Can you trust a LinkedIn profile? Will recruiters trust yours? Available at: https://rapidbi.com/can-you-trust-a-linkedin-profile-will-recruiters-trust-yours/ (Accessed: 14 December 2016).

2. *Adams S.* Everything you need to know about LinkedIn endorsements. Available at: http://www.forbes.com/sites/susanadams/2013/12/24/everything-you-need-to-know-about-linkedin-endorsements-2/#23a19f1e1e4d (Accessed: 14 December 2016).
3. *Maybury M., D'Amore R., House D.* Expert Finding for Collaborative Virtual Environments. Communications of the ACM 14(12): 55-56. In Ragusa, J. and Bochenek, G. (eds). Special Section on Collaboration Virtual Design Environments. December (2001).
4. *Кравцов Г.А.* Модель вычислений на классификациях// Электрон. моделирование. — 2016. — **38**, № 1. — С. 73—87.
5. *Кормен Т.Х., Лейзерсон Ч.И., Ривест Р.Л., Штайн К.* Алгоритмы: построение и анализ. 3-е изд. — М. : «Вильямс», 2013. — 1328 с.
6. *Zhang L., Tu W.* Six Degrees of Separation in Online Society. Available at: http://journal.webscience.org/147/2/websci09_submission_49.pdf. (Accessed: 14 December 2016).
7. *Bennett D.* The Dunbar number, from the guru of social networks. Available at: https://www.bloomberg.com/news/articles/2013-01-10/the-dunbar-number-from-the-guru-of-social-networks (Accessed: 15 December 2016).
8. *Тоценко В.Г.* Методы и системы поддержки принятия решений. Алгоритмический аспект. — Киев: Наук. думка. — 2002. — 382 с.
9. *Clow K.E., Baack D.* Variability theory. Concise Encyclopedia of Advertising, 2005. — P. 176.

Поступила 30.12.16

REFERENCES

1. Morrison, M. (2016), "Can you trust a LinkedIn profile? Will recruiters trust yours?", available at: https://rapidbi.com/can-you-trust-a-linkedin-profile-will-recruiters-trust-yours/ (accessed December 14, 2016).
2. Adams, S. (2013), "Everything you need to know about LinkedIn endorsements", available at: http://www.forbes.com/sites/susanadams/2013/12/24/everything-you-need-to-know-about-linkedin-endorsements-2/#23a19f1e1e4d (accessed December 14, 2016).
3. Maybury, M., D'Amore, R. and House, D. Expert finding for collaborative virtual environments. Communications of the ACM 14(12): 55-56. In Ragusa, J. and Bochenek, G. (eds). Special Section on Collaboration Virtual Design Environments. December (2001).
4. Kravtsov, H.A. (2016), "Model of computations over classifications", *Elektronnoe modelirovanie*, Vol. 38, no. 1, pp. 73-87.
5. Cormen, T.Kh., Leizerson, Ch.I., Rivest, R.L. and Shtain, K. (1990), *Algoritmy: postroyeniye i analiz* [Algorithms: structure and analysis], Viliyams, Moscow, Russia.
6. Zhang, L. and Tu, W. "Six degrees of separation in online society", available at: http://journal.webscience.org/147/2/websci09_submission_49.pdf (accessed December 14, 2016).
7. Bennett, D. (2013), "The Dunbar number, from the guru of social networks", available at: https://www.bloomberg.com/news/articles/2013-01-10/the-dunbar-number-from-the-guru-of-social-networks (accessed December 15, 2016).
8. Totsenko, V.G. (2002), *Metody i sistemy podderzhki prinyatiya resheniy. Algoritmicheskiy aspect* [Models and systems for decision making support], Naukova dumka, Kiev, Ukraine.
9. Clow, K.E. and Baack, D. (2005), Variability theory. Concise Encyclopedia of Advertising.

Received 30.12.16

H.A. Kravtsov

CLASSIFICATION CALCULUS. VALIDATION
OF QUALIFICATION IN SOCIAL NETWORKS

Operators of the largest social media state the problem of experts' qualification validation. The suggested approach is based on the classification calculus theory allowing minimizing subjectivity during self-evaluation and mutual evaluation of skills and is focused on nominal qualification value and endorsement validation structure. The aggregation of nominal qualification value and endorsement validation structure permits the application of variability theory for the purpose of pragmatic expert recruitment.

K e y w o r d s: classification, expert, qualification, endorsement, subjectivity, objectivity, actualization, assessment.

КРАВЦОВ Григорий Алексеевич, канд. техн. наук, докторант Ин-та проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины. В 2000 г. окончил Севастопольский военно-морской ин-т им. П.С. Нахимова. Область научных исследований — кибербезопасность смарт-грид, криптография, программирование, разработка распределенных гетерогенных вычислительных систем.