



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,
e-mail: hryhoriy.kravtsov@gmail.com)

Вычисления на классификациях. Подбор сотрудников как интерпретация проблемы подбора экспертов

Подбор сотрудников — одна из интерпретаций проблемы подбора экспертов. В основу предлагаемого подхода подбора сотрудников положена модель вычислений на классификациях. Важным аспектом подбора является соответствие уровня знаний и навыков. Подход позволяет уменьшить субъективизм представителя работодателя и дать оценку слабым и сильным сторонам кандидата. К нефункциональным достоинствам подхода можно отнести легкость реализации на любом языке программирования высокого уровня.

К л ю ч е в ы е с л о в а: классификация, эксперт, квалификация, навык, знание, оценка, вакансия, соответствие.

Підбір співробітників є однією з інтерпретацій проблеми підбору експертів. В основу підходу, що пропонується, закладено модель обчислень на класифікаціях. Важливим аспектом підбору є відповідність за рівнем знань та навиків. Підхід дозволяє зменшити суб'єктивізм представника роботодавця і дати оцінку слабким та сильним сторонам кандидата. Нефункціональними перевагами підходу можна вважати легкість реалізації будь-якою мовою програмування високого рівня.

К л ю ч о в і с л о в а: класифікація, експерт, кваліфікація, навик, знання, оцінка, вакансія, відповідність.

Постановка задачи. Суть задачи подбора эксперта [1—3] следующая: есть требования к кандидату и есть некоторое множество кандидатов, среди которых необходимо выбрать такую кандидатуру, которая наилучшим образом удовлетворяет заданным требованиям.

Дадим более строгую формулировку. Пусть задана некоторая классификация знаний и навыков A^i [4]. Любой из навыков или любое знание A_j^i могут быть оценены в соответствии с некоторой упорядоченной по возрастанию шкалой R .

© Г.А. Кравцов, 2017

Пусть профиль искомого кандидата задан совокупностью троек $\langle A_I^i, W_I, R_I \rangle$, где A_I^i — класс знаний, который не может быть уточнен (лист дерева); I — путь уточнения класса знаний (навыков) в классификации A^i ; W_I — весовой коэффициент важности наличия соответствующего знания (навыка) у кандидата с уровнем знаний (навыков) R_I . Будем полагать, что совокупность троек $V = \langle A_I^i, W_I, R_I \rangle$ (vacancy) обладает мощностью, равной числу уникальных классов A_I^i , т.е. совокупность троек не содержит записей, у которых один и тот же класс знаний (навыков). Потребуем, чтобы для весовых коэффициентов выполнялось условие $\sum_I W_I = 1, 0 \leq W_I \leq 1$.

Пусть профиль кандидата задан двойкой $C = \langle A_I^i, R_I \rangle$ (candidate), где A_I^i — класс знаний, который не может быть уточнен (лист дерева); I — путь уточнения класса знаний в классификации A^i ; R_I — уровень знаний (навыков) по классу A_I^i . Примем допущение, что уровень знаний (навыков) R_I по классу A_I^i есть величина объективная (отражающая реальный уровень квалификации) без необходимости доказательства. Заметим, что при задании профиля вакансии и профиля кандидата используется одна и та же классификация A^i .

Требуется определить показатель соответствия $S(C, V)$ профиля кандидата $C = \langle A_I^i, R_I \rangle$ профилю вакансии $V = \langle A_I^i, W_I, R_I \rangle$ так, чтобы выполнялось условие $0 \leq S(C, V) \leq 1$, где $S(C, V) = 1$ означает полное (наилучшее) соответствие профиля кандидата профилю вакансии.

Для некоторой вакансии $V = \langle A_I^i, W_I, R_I \rangle$ и множества кандидатов $C_k = \langle A_I^i, R_I \rangle$ необходимо найти решение задачи $S(C_k, V) \rightarrow \max$.

Соответствие по уровню знания или навыка. Согласно строгой постановке задачи профиль вакансии $V = \langle A_I^i, W_I, R_I \rangle$ содержит перечень знаний (навыков) с ожидаемым уровнем R_I в соответствии с некоторой шкалой R . При этом выполняется условие $\min(R) \leq R_I \leq \max(R)$.

Рассмотрим возможные варианты близости двух уровней, R_I и R_Y , принадлежащих одной и той же шкале R , где $\min(R) \leq R_I \leq \max(R)$ и $\min(R) \leq R_Y \leq \max(R)$, $\min(R) \geq 0$.

Наипростейшим вариантом может быть отношение расстояния между R_I и R_Y к длине всей шкалы:

$$S_R = 1 - \frac{|R_I - R_Y|}{\max(R) - \min(R)}. \quad (1)$$

Очевидно, что при $R_I = R_Y$ выражение (1) равно единице. Минимальное значение (1) принимает, если отношение $|R_I - R_Y|$ к $\max(R) - \min(R)$ равно единице. В данном случае R_I и R_Y находятся на концах шкалы R . Несмотря

на то, что выражение (1) отражает близость двух оценок в пределах одной и той же шкалы, оно малоприменимо на практике. Выражение $|R_I - R_Y|$ в числителе обеспечивает одинаковую степень соответствия, несмотря на знак под модулем, т.е. с помощью (1) невозможно определить недостаток (избыток) квалификации. Рассмотрим пример.

Пусть для некоторого знания A_I^i в соответствии со шкалой $R = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ требуется уровень знаний 6. На данную вакансию претендуют два кандидата: у одного уровень компетенции — 4, у другого — 8. Воспользовавшись формулой (1), получим для первого кандидата

$$S_R = 1 - \frac{|R_I - R_Y|}{\max(R) - \min(R)} = 1 - \frac{|6-4|}{9-1} = 1 - \frac{2}{8} = 0,75,$$

для второго кандидата

$$S_R = 1 - \frac{|R_I - R_Y|}{\max(R) - \min(R)} = 1 - \frac{|6-8|}{9-1} = 1 - \frac{2}{8} = 0,75.$$

Предложенная оценка (1) близости уровней знаний (навыков) является неприемлемой. По этой же причине является неприемлемым и метод парзеновского окна [5]. Таким образом, очевидно, что любые распределения величины близости оценок, симметричные относительно ожидаемой оценки (по вакансии), будут также неприемлемы.

Кадровой политикой работодателя может быть предусмотрена готовность инвестировать в будущее развитие нанимаемых кандидатов наряду с неготовностью нанимать кандидатов, имеющих значительно завышенные показатели по некоторому навыку (знанию).

Такая политика может быть единой для всех вакансий работодателя, но не стоит исключать возможность определения политики для каждой вакансии и даже для каждого знания (навыка) в профиле вакансии. На основании этого представление политики сравнения уровня знаний (навыков) должно быть простым и интуитивно понятным.

Сформируем требования, которым должна удовлетворять политика сопоставления уровней знаний.

Требование 1. Если уровень некоторого знания (навыка) согласно требованиям вакансии $V = \langle A_I^i, W_I, R_I \rangle$ совпадает с уровнем соответствующего знания (навыка) кандидата $C = \langle A_I^i, R_I \rangle$, т.е. $R_I = R_Y$, то показатель соответствия $S(C, V)$ должен быть равен единице, т.е. $S(C, V) = 1$.

Требование 2. Если уровень некоторого знания (навыка) согласно требованиям вакансии $V = \langle A_I^i, W_I, R_I \rangle$ ниже уровня соответствующего знания (навыка) кандидата $C = \langle A_I^i, R_I \rangle$, т.е. $R_I < R_Y$, то показатель соот-

ответствия $S(C, V)$ должен быть нестрогим меньше единицы, т.е. $S(C, V) \leq 1$, при условии, что приращение $S(C, V)$ является величиной неположительной (отрицательной или равной нулю), если $R_I < R_Y$.

Требование 3. Если уровень некоторого знания (навыка) согласно требованиям вакансии $V = \langle A_I^i, W_I, R_I \rangle$ выше уровня соответствующего знания (навыка) кандидата $C = \langle A_I^i, R_I \rangle$, т.е. $R_I > R_Y$, то показатель соответствия $S(C, V)$ должен быть нестрогим меньше единицы, т.е. $S(C, V) \leq 1$, при условии, что приращение $S(C, V)$ является величиной неотрицательной (положительной или равной нулю), если $R_I > R_Y$.

Требование 4. Если шкала R представима строго упорядоченным набором из n оценок, функция соответствия должна быть представима в табличном виде, в котором ставятся в соответствие значения $S(C, V)$ на интервале от наибольшей по модулю, но отрицательной по знаку, ошибки до наибольшей положительной ошибки между ожидаемым согласно вакансии уровнем знаний и уровнем знаний кандидата по выбранному знанию (навыку).

Рассмотрим пример. Пусть задана шкала из шести оценок: Basic, Average, Intermediate, Advanced, Proficient и Expert. Ожидается, что кандидаты должны обладать уровнем Advanced. Согласно требованиям 1—4 табличное задание функции соответствия приведено в табл. 1. Функция представлена в виде графика на рис. 1, где ось абсцисс отражает расстояние от уровня знания (навыка), предписанного вакансией, до уровня соответст-

Таблица 1. Табличное задание функции соответствия уровня знания

Уровень знания (навыка)	Отклонение от ожидаемого уровня знаний	Функция соответствия уровня знаний (навыков)	Соответствие заданной шкале
Низкий	-5	0,0	Basic Average Intermediate
	-4	0,2	
	-3	0,3	
	-2	0,8	
	-1	0,9	
Ожидаемый	0	1,0	Advanced
Завышенный	1	1,0	Proficient Expert
	2	0,9	
	3	0,85	
	4	0,8	
	5	0,75	

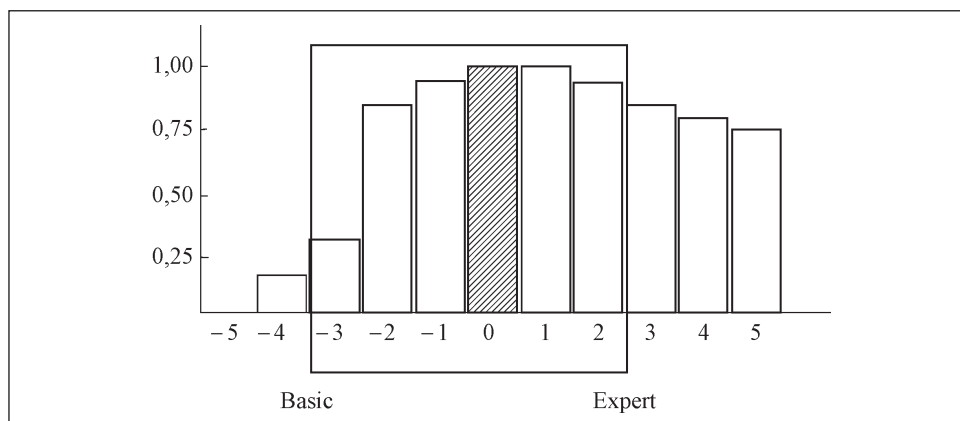


Рис. 1. Графическое представление функции соответствия знания (навыка) согласно табл. 1 и соотнесение с выбранной шкалой

вующего знания (навыка) из профайла кандидата, т.е. $R_I - R_Y$, где R_I — уровень знания (навыка), предписанного вакансией, R_Y — уровень знания (навыка), указанный в профайле кандидата; ось ординат отражает степень близости оценок уровня знаний вакансии и профайла кандидата; прямоугольником выделена зона, покрываемая шкалой; заштрихованный столбик — ожидаемый уровень знания (навыка) согласно вакансии.

Возникает вопрос: как трактовать степень соответствия уровней знаний, чтобы представителю работодателя было легче формировать политику найма? Прежде всего, это вероятностная интерпретация: если кандидат выбирается только по одному знанию (навыку), то функция соответствия уровней указывает на вероятность найма, т.е. если у кандидата такой уровень выбранного знания, как требует вакансия, то это значит, что он подходит для найма.

На рис. 2 согласно требованиям 1—4 показано несколько некорректно определенных функций соответствия по уровню знания (навыка) (без соотнесения со шкалой).

Рассмотрим на примере, как предложенная функция соответствия уровня знания (навыка) применяется для нахождения степени соответствия профиля кандидата профилю вакансии.

Соответствие профиля кандидата профилю вакансии. Как указано выше, профиль вакансии задан тройками $V = \langle A_I^i, W_I, R_I \rangle$, где A_I^i — знание (навык) в соответствии с классификацией A^i ; R_I — уровень знания (навыка) A_I^i , предписанный вакансией (в соответствии со шкалой R : Basic, Average, Intermediate, Advanced, Proficient и Expert). Представим профиль вакансии в виде табл. 2.

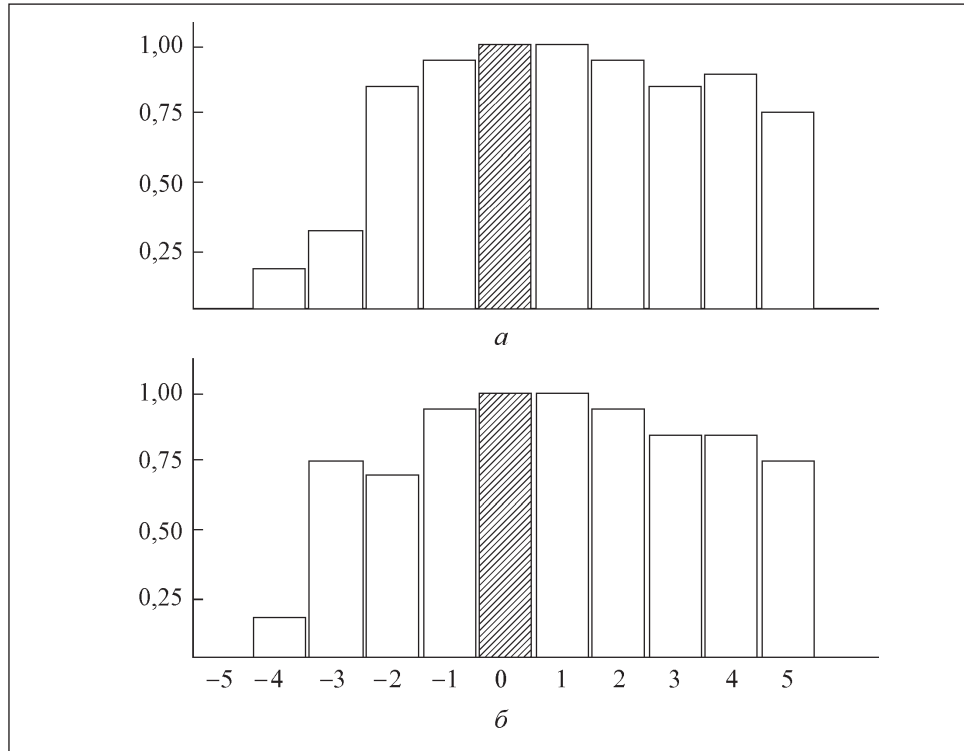


Рис. 2. Пример некорректно определенной функции соответствия уровней знания: *a* — противоречит требованию 2 (столбец 4 выше, чем столбец 3); *б* — противоречит требованию 3 (столбец -3 выше, чем столбец -2)

Зададим профиль некоторого кандидата. Как указано выше, профиль кандидата — это двойка $C = \langle A_I^i, R_I \rangle$, где A_I^i — знание (навык) в соответствии с классификацией A^i ; R_I — уровень знания (навыка) A_I^i в соответствии со шкалой R : Basic, Average, Intermediate, Advanced, Proficient и Expert. Представим профиль кандидата в виде табл. 3.

Таблица 2. Пример табличного задания профиля вакансии

A_I^i	W_I	R_I
1	0,3	Proficient
2	0,3	Proficient
3	0,25	Intermediate
4	0,15	Average

Таблица 3. Пример табличного задания профиля кандидата

A_I^i	R_I
1	Expert
2	Proficient
3	Average
5	Average

Для последующих вычислений необходимо построить стратегию сравнения уровней знаний (навыков), которая должна учитывать меру различия $\bar{Q}(A_I^i, A_Y^i)$ между знаниями A_I^i в рамках классификации A^i [4]. В табл. 4 приведен пример табличного представления меры отличия между знаниями и навыками, которые требуются вакансией, и имеющимися у кандидата. Усеченным вариантом табл. 4 является табл. 5, где в столбцах отражены знания, которыми обладает кандидат, а в строках — знания, необходимые для вакансии.

Сформируем стратегию сравнения согласно следующему алгоритму.
А л г о р и т м.

Шаг 1. Изначально в табл. 6 выделим строки и столбцы, на пересечении которых стоит мера отличия, равная нулю. Как видно из табл. 6, только знание 4 из профиля вакансии не связано ни с одним знанием (навыком) из профиля кандидата. Знания (навыки), присутствующие и в профиле вакансии, и в профиле кандидата, являются безусловно сравнимыми. В этом случае шаги 2 и 3 следует пропустить.

Для знаний (навыков), которые не являются безусловно сравнимыми, необходимо выполнить поиск «ближайшего» знания (навыка), отражающего наиболее сильную сторону соискателя. Это значит, что необходимо выбрать для сравнения такое знание, которое имеет наименьшую меру

Таблица 4. Пример табличного представления матрицы меры отличия между знаниями (навыками)

Знание	Знание 1	Знание 2	Знание 3	Знание 4	Знание 5
1	0	0,25	0,3	0,1	0,2
2	0,25	0	0,4	0,15	0,1
3	0,3	0,4	0	0,1	0,2
4	0,1	0,15	0,1	0	0,1
5	0,2	0,1	0,2	0,1	0

Таблица 5. Таблица мер отличия, усеченная в соответствии с профилями знаний (навыков) вакансии и кандидата

Профиль вакансии	Профиль кандидата			
	Знание 1	Знание 2	Знание 3	Знание 5
Знание 1	0	0,25	0,3	0,2
Знание 2	0,25	0	0,4	0,1
Знание 3	0,3	0,4	0	0,2
Знание 4	0,1	0,15	0,1	0,1

отличия. Если несколько знаний (навыков) имеют одинаковую наименьшую меру отличия от искомого знания (навыка) (как в табл. 6: знание 1, знание 3, знание 5 по отношению к знанию 4 имеют меру отличия, равную 0,1), то переходим к следующему шагу.

Шаг 2. На основании табл. 1 построим таблицу соответствия уровней знаний (навыков) (табл. 7). Учтем при этом, что вакансией предусмотрен для знания 4 уровень Average (см. табл. 2).

Шаг 3. Построим сравнительную таблицу, где в качестве правила сравнения уровней знаний (навыков) воспользуемся табл. 7, как возможным вариантом формализации политики найма (табл. 8).

Как видно из табл. 8, не имеет значения, будет выбрано знание 3 или знание 5 для последующего расчета соответствия профайла кандидата профайлу вакансии, так как скорректированное на меру отличия соответствие для указанных знаний по отношению к знанию 4 из вакансии для

Таблица 6. Шаг 1 на усеченной таблице мер отличия

Профиль вакансии	Профиль кандидата			
	Знание 1	Знание 2	Знание 3	Знание 5
Знание 1	0	0,25	0,3	0,2
Знание 2	0,25	0	0,4	0,1
Знание 3	0,3	0,4	0	0,2
Знание 4	0,1	0,15	0,1	0,1

Таблица 7. Пример табличного задания функции соответствия уровня знания с ожидаемым уровнем Average

Уровень знания (навыка)	Отклонение от ожидаемого уровня знаний	Функция соответствия уровня знаний	Соответствие заданной шкале
Низкий	-5	0	
	-4	0,2	
	-3	0,3	
	-2	0,8	
	-1	0,9	
Ожидаемый	0	1	Basic
Завышенный	1	1	Average
	2	0,9	Intermediate
	3	0,85	Advanced
	4	0,8	Proficient
	5	0,75	Expert

обоих равно 0,9 (максимальное скорректированное значение). Шаги 2 и 3 следует повторить для всех знаний (навыков), которые являются условно сравнимыми по результатам выполнения шага 1.

Шаг 4. Составим следующую таблицу расчета частного соответствия P каждому знанию (навыку) из профиля кандидата (табл. 9). Частное соответствие знанию (навыку) определяется по формуле

$$P_I = (1 - \bar{Q}(A_I^i, A_Y^i)) \bar{S}(R_I, R_Y) W_I. \quad (2)$$

Шаг 5. Интегральная оценка соответствия $S(C, V)$ определяется как сумма частных соответствий (2) на всем множестве A_j^i :

$$S(C, V) = \sum_I P_I. \quad (3)$$

Очевидно, что для (3) выполняется условие $S(C, V) \leq \sum_I W_I, S(C, V) \leq 1$.

Таким образом, интегральная оценка соответствия профиля кандидата (см. табл. 3) профилю вакансии (см. табл. 2) при стратегии найма (см. табл. 1) будет $0,3 + 0,3 + 0,25 + 0,135 = 0,985$. При наличии множества кандидатов каждому из них может быть поставлен показатель соответствия $S(C_i, V)$, где C_i — некоторый кандидат из множества соискателей. Это позволяет

Таблица 8. Выбор соответствия знаний (навыков)

A_j^i кандидата	R_Y кандидата	Профиль вакансии $A_Y^i = \text{знание 4}, R_I = \text{Average}$		
		$\bar{Q}(A_I^i, A_Y^i)$	$\bar{S}(R_I, R_Y)$	Скорректированное соответствие, $(1 - \bar{Q}(A_I^i, A_Y^i)) \bar{S}(R_I, R_Y)$
Знание 1	Expert	0,1	0,8	$(1 - 0,1) 0,8 = 0,72$
Знание 3	Average	0,1	1	$(1 - 0,1) 1,0 = 0,9$
Знание 5	Average	0,1	1	$(1 - 0,1) 1,0 = 0,9$

Таблица 9. Расчет частного соответствия P_I знаний (навыков)

A_j^i по вакансии	R_I по вакансии	A_Y^i кандидата	R_Y кандидата	$\bar{Q}(A_I^i, A_Y^i)$	$\bar{S}(R_I, R_Y)$	W_I по вакансии	P_I по вакансии
Знание 1	Expert	Знание 1	Expert	0	1	0,3	0,3
Знание 2	Proficient	Знание 2	Proficient	0	1	0,3	0,3
Знание 3	Average	Знание 3	Average	0	1	0,25	0,25
Знание 4	Average	Знание 3	Average	0,1	1	0,15	0,135

упорядочить кандидатов согласно соответствию профилей кандидатов C_i профилю вакансии V . В ситуации, когда два и более кандидатов (множество наилучшего выбора) набирают одинаковую максимальную оценку $S(C_i, V)$, следует воспользоваться подходом, разрешающим проблему проклятия размерности [6], а именно из табл. 9 удалять строки, начиная со строк с максимальной мерой отличия классов, пересчитывая только кандидатов из множества наилучшего выбора.

Выводы

Предложенный подход к поиску сотрудника позволяет свести указанную задачу к аддитивной функции взвешенных сопоставлений знаний (навыков). Подход легко реализуется в виде программы.

СПИСОК ЛИТЕРАТУРЫ

1. *Gordon Welty*. Some Problems of Selecting Delphi Experts for Educational Planning and Forecasting Exercises // *California Journal of Educational Research*. — 1973. — Vol. 24, No. 3. — P. 129—134.
2. *Fang H., Zhai C.* Probabilistic Models for Expert Finding. In: Amati G., Carpineto C., Romano G. (eds). *Advances in Information Retrieval. ECIR 2007. Lecture Notes in Computer Science*. Vol 4425. Springer, Berlin, Heidelberg.
3. *Тоценко В.Г.* Методы и системы поддержки принятия решений. Алгоритмический аспект. — Киев: Наук. думка. — 2002. — 382 с.
4. *Кравцов Г.* Модель вычислений на классификациях // *Электрон. моделирование*. — 2016. — **38**, № 1. — С. 73—87.
5. *Флах П.* Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. — М.: ДМК Пресс. — 2015. — 400 с.
6. *Warren B. Powell*. Approximate Dynamic Programming: Solving the Curses of Dimensionality // *Wiley*. — 2011. — 656 с.

Поступила 06.02.17

REFERENCES

1. Gordon, Welty (1973), "Some problems of selecting Delphi experts for educational planning and forecasting exercises", *California Journal of Educational Research*, Vol. 24, no. 3, pp. 129-134.
2. Fang, H. and Zhai, C. (2007), Probabilistic models for expert finding, eds G. Amati, C. Carpineto, G. Romano, *Advances in Information Retrieval. ECIR 2007. Lecture Notes in Computer Science*, Vol 4425, Springer, Berlin, Heidelberg, Germany.
3. Totsenko, V. (2002), *Metody i sistemy podderzhki prinyatiya resheniy. Algoritmicheskii aspekt* [Methods and systems of decision making support. Algorithmic aspect], Naukova dumka, Kiev, Ukraine.
4. Kravtsov, H.A. (2016), "Model of computations on classifications", *Elektronnoe modelirovanie*, Vol. 38, no. 1, pp. 73-87.

5. Flakh, P. (2015), *Mashinnoe obuchenie. Nauka i iskusstvo postroyeniya algoritmov, kotorye izvlekayut znaniya iz dannykh* [Machine learning: The art and science of algorithms that derive knowledge of data], DMK Press, Moscow, Russia.
6. Warren, B. Powell (2011), *Approximate dynamic programming: Solving the curses of dimensionality*, Wiley.

Received 06.02.17

H.A. Kravtsov

THE CALCULUS OVER CLASSIFICATIONS. SELECTION OF PERSONNEL
AS INTERPRETATION OF THE PROBLEM OF EXPERT SELECTION

The recruitment is possible interpretation of the expert selection problem. A method of calculus over classification has been used as the mathematical basis. The relevance by knowledge/skill level is very important aspect in recruitment process. The approach permits reducing the subjectivity of recruiter and giving the complete picture of the weak and strong sides of a candidate. Implementation simplicity in any high level programming language belongs to non-functional advantages.

Key words: classification, expert, qualification, habit, knowledge, estimation, vacancy, accordance.

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.

