# Analysis and Evaluation of Competence of Information Sources in Problems of Intellectual Data Processing

**Krisilov V. A., Komleva N. O.**
Odessa National Polytechnic University
Odessa, Ukraine

***Abstract.*** The issues of work related to solving problems of intellectual data processing is to increase the efficiency of qualimetry, classification, diagnostics, choice, decision making, forecasting, taxonomy, etc., using data mining algorithms and a statistical approach. The aim of the work is to formalize the description and analysis of the source data as a means of improving the quality of solving intellectual problems. Compared with the well-known works, in which the emphasis was placed on certain informational criteria, the approach proposed in the work allows to form a set of quantitative and qualitative criteria for the formalization of information sources. To achieve the goal, taking into account the knowledge of the structure and context of the problem, objective requirements imposed on the input information have been formulated. To ensure the specified quality of solving the problems of intellectual data processing, it is necessary that the input information be objective, reliable, relevant, complete, timely, understandable, etc., and have a certain degree of accuracy, which is possible only with competent experts. The paper considers the concept of single and group expertise. The dependence of the quality of the solution of the data processing problem on the requirements for the source data is considered by the example of the forecasting problem. The accuracy of forecasting the results of external independent estimation in the selected section on the degree of completeness of the initial data and the selected methods of data processing and analysis is analyzed.
***Keywords*:** building models, intellectual tasks, competence, quality of the solution, the properties of the input data, forecasting.
**DOI**: 10.5281/zenodo.3239185

**Analiza și evaluarea competenței surselor de informație în sarcinile procesării inteligente a datelor**
**Krisilov V.A., Komlevaya N.O.**
Universitatea Politehnică Națională din Odessa, Odessa, Ucraina
***Rezumat.*** Rezolvarea problemelor complexe sociale, tehnice, economice și științifice necesită utilizarea unor sisteme ierarhizate pe mai multe niveluri bazate pe prelucrarea inteligentă a datelor. Probleme legate de acest domeniu de cercetare sunt creșterea eficienței rezolvării problemelor de calimetrie, clasificare, diagnosticare, selecție, luare a deciziilor, prognoză, taxonomie etc., utilizând algoritmi de extragere a datelor și o abordare statistică. Scopul lucrării este de a formaliza descrierea și analiza datelor sursă ca mijloc de îmbunătățire a calității rezolvării problemelor intelectuale. În comparație cu lucrările bine-cunoscute, în care s-a pus accentul pe anumite criterii informaționale, abordarea propusă în lucrare permite formarea unui set de criterii cantitative și calitative pentru formalizarea surselor de informare. Pentru atingerea scopului, ținând seama de cunoașterea structurii și a contextului problemei, au fost formulate cerințe obiective impuse informațiilor de intrare, nerespectarea cărora limitează utilizarea rezultatelor rezolvării problemei. Pentru a asigura calitatea specificată de rezolvare a problemelor de prelucrare a datelor intelectuale, este necesar ca informațiile de intrare să fie obiective, fiabile, relevante, complete, la timp, ușor de înțeles etc. și să aibă, de asemenea, un anumit grad de exactitate, posibil numai cu experți competenți. Dependența calității soluției problemei de prelucrarea datelor de cerințele surselor de date sursă este luată în considerare prin exemplul problemei de prognoză. Se analizează precizia prognozării rezultatelor estimării externe independente în secțiunea selectată privind gradul de completare a datelor inițiale și a metodelor selectate de prelucrare și analiză a datelor. Studiul a arătat fezabilitatea utilizării metodei medii mobile și a metodei celor mai mici pătrate pentru a obține o evaluare calitativă a prognosticului.
***Cuvonte-cheie***: modele de construcție, sarcini intelectuale, competență, acuratețe, precizie, exhaustivitate, calitatea soluției, proprietăți ale informațiilor de intrare, prognoză.

**Анализ и оценка компетентности источников информации в задачах интеллектуальной обработки данных**
**Крисилов В.А., Комлевая Н.О.**
Одесский национальный политехнический университет, Одесса, Украина
***Аннотация.*** Решение сложных социальных, технических, экономических и научных задач требует использования иерархических многоуровневых систем, в основе работы которых лежит интеллектуальная

обработка данных. Проблематикой работ, связанных с этим направлением исследований, является повышение эффективности решения задач квалиметрии, классификации, диагностики, выбора, принятия решений, прогнозирования, таксономии и др. с использованием алгоритмов интеллектуального анализа данных и статистического подхода. Целью работы является формализация описания и анализа исходных данных как средство повышения качества решения интеллектуальных задач. По сравнению с известными работами, в которых акценты делались на определенные информационные критерии, предлагаемый в работе подход позволяет сформировать множество количественных и качественных критериев для формализации источников информации. Для достижения цели с учетом знания структуры и контекста разрешаемой проблемы были сформулированы предъявляемые ко входной информации объективные требования, несоблюдение которых ограничивает область использования результатов решения задачи. Для обеспечения заданного качества решения задач интеллектуальной обработки данных необходимо, чтобы входная информация была объективной, достоверной, релевантной, полной, своевременной, понятной и др., а также обладала определенной степенью точности, что возможно лишь при наличии компетентных экспертов. Рассмотрена зависимость качества решения задачи обработки данных от требований, предъявляемых к источникам исходных данных на примере задачи прогнозирования. Проанализирована точность прогноза результатов внешнего независимого оценивания в выбранном разрезе от степени полноты исходных данных и выбранных методов обработки и анализа данных. Исследование показало целесообразность использования метода скользящей средней и метода наименьших квадратов для получения качественной прогностической оценки. При этом, несмотря на отсутствие почти трети исходных данных от компетентных источников, средняя относительная ошибка прогноза не превысила 10%.
***Ключевые слова***: построение моделей, интеллектуальные задачи, компетентность, достоверность, точность, полнота, качество решения, свойства входной информации, прогнозирование.

## ВВЕДЕНИЕ

В настоящее время стремительно растут объемы данных, сопровождающие различные социальные, технические, экономические, научные и другие процессы. В работах [1, 2] показаны пути выявления в огромных информационных массивах причинно-следственных связей и закономерностей. В работах [3, 4] описано построение моделей, которые помогают принимать решения в навигационных и эксплуатационных задачах, оперирующих большими объемами информации. Статьи [5, 6] описывают проблемы приема-передачи информационных блоков большой размерности с помощью сети Internet. Статья [7] рассматривает проблемы автоматизации процесса обучения студентов при большом количестве возможных режимов обучения. Сложность и многообразие описанных ситуаций и проблемы, возникающие при их решении, подчеркивают важность предварительного анализа исходных данных как средства повышения качества результата.

## I. ОБЗОР ИНФОРМАЦИОННЫХ ПОДХОДОВ К ОБЕСПЕЧЕНИЮ КАЧЕСТВА РЕШЕНИЯ ЗАДАЧ

Во всем множестве задач интеллектуальной обработки данных принято выделять следующие типы: квалиметрия, классификация, диагностика (распознавание образов), выбор, принятие решений, формирование заключений, построение прогноза, кластерный анализ (таксономия). Каждая из таких задач может носить как чисто теоретический характер, так и служить основанием для разработки специализированных и комплексных технических, социальных, организационных и экономических систем.

В [8] описаны общие информационные подходы к решению задач и выделены основные этапы. В работе [9] внимание уделяется вопросу прогнозирования и связанному с ним исследованию точности исходной информации. В статье [10] описаны технологии, которые являются вспомогательными инструментами для решения задачи квалиметрии с учетом релевантности исходных данных. Труды [11, 12] показывают влияние степени достоверности информации на качество решения задач классификации и диагностики. Статья [13] рассматривает связь между вероятностными характеристиками атрибутов исходных данных и качеством полученного решения. Проблемы потери качества при принятии решения в задачах систематизации и кластеризации ввиду несвоевременного поступления данных описаны в [14]. Вопросы влияния отдельных выбросов на степень однородности исходных данных и механизмы ее поддержания на требуемом уровне рассмотрены в [15, 16].

Зачастую существуют объективные требования, предъявляемые ко входной информации, несоблюдение которых ограничивает область использования результатов решения задачи или же делает их вовсе непригодными. Так, в работе [17] показано, что качество входной информации и, соответственно, качество решения снижается при наличии пропусков в исходных данных, в работах [18, 19] – при наличии противоречивости этих данных, в [20 – 22] – при наличии аномальных значений и неустранимой шумовой составляющей, в [23] – из-за несоответствия форматов данных, ошибок ввода данных или опечаток, дублирования и т.д. Работа [24] посвящена исследованию функций распределения энтропии исходных данных и предельных допустимых случаев.

Таким образом, анализ и систематизация информационных подходов к обеспечению заданного качества решения описанных выше задач показали, что входная информация должна обладать определенным набором свойств: быть объективной, достоверной, релевантной, полной, своевременной, понятной и др., а также обладать определенной степенью точности [25]. В [26] описана различная природа, представление и принципы оценивания этих свойств.

В общем виде под качеством некоторого решения понимается степень его соответствия поставленной цели, при этом каждое решение получается на основании входной информации, имеющей определенную совокупность свойств. Заметим, что именно точность, достоверность, релевантность и полнота входной информации, представленные в числовом виде, позволяют рассматривать ситуацию снижения неопределенности при решении некоторой выбранной задачи как параметрическое представление функции от нескольких переменных. Знание структуры и контекста разрешаемой проблемы позволяет определить ее целевую функцию, в которой переменными и ограничениями выступают свойства входной информации. Это в очередной раз подчеркивает важность наличия компетентных источников информации, данным от которых можно доверять.

Целью работы является формализация описания и анализа исходных данных как средство повышения качества решения интеллектуальных задач. По сравнению с известными работами, в которых акценты делались на определенные информационные критерии, предлагаемый в данной работе подход позволяет сформировать множество количественных и качественных критериев для формализации источников информации.

## II. МЕТОДЫ ИССЛЕДОВАНИЯ

В общем виде свойства входной информации, требующие контроля, можно представить множеством $M$:

$$M = \{O, D, R, G, T, U\}, \quad (1)$$

где $O$ — объективность информации;

$D$ — достоверность;

$R$ — релевантность;

$G$ — полнота;

$T$ — своевременность;

$U$ — понятность.

Рассмотрим более подробно свойства входной информации и их взаимосвязи с качеством решения выбранной задачи.

Объективность информации $O$ заключается в том, что она отражает внешний мир, существующий независимо от конкретных объектов, субъектов и процессов и носит общепризнанный характер. Объективность – важнейшее свойство информации, которое, к сожалению, крайне редко бывает абсолютным, ведь данные – это только один компонент информации. Второй компонент – информационные методы – связан с источником или потребителем информации и имеет субъективную природу. В зависимости от того, какой компонент превалирует в информационном процессе, результирующая информация может быть объективной более или менее. В науке принято считать объективной информацию воспроизводимую. Например, законы химии воспроизводимы, а законы астрологии – нет. Соответственно, химия считается объективной наукой, а астрология — нет. Основной способ повышения объективности информации заключается в увеличении её полноты. Например, оценка уровня квалификации программиста – субъективна, но при увеличении числа независимых экспертов объективность оценки повышается.

Как исходная, так и результирующая информация характеризуются с двух позиций – достоверности (внутренней валидности) и

обобщаемости (внешней валидности, применимости). Случайные ошибки возникают из-за отклонения результата отдельного наблюдения или измерения от его истинного значения, что обуславливается случайностью. Случайные вариации проявляются на любом этапе решения задачи и связаны с индивидуальной вариабельностью свойств изучаемых объектов, случайными ошибками измерения и недостаточным объёмом выборки. В отличие от систематических ошибок, случайные ошибки нельзя устранить, но можно свести к минимуму. Этого достигают правильным планированием процесса решения задачи, увеличением размера исходной выборки, требованием многократного получения данных от одних и тех же источников, и, кроме того, путём оценки вероятности случайной ошибки с использованием выбранных методов.

Достоверность *D* (внутренняя валидность) исходных данных определяется тем, насколько они соответствуют методам и инструментам, выбранным для решения поставленной задачи. Достоверность полученных результатов решения задачи определяется тем, насколько структура решения соответствует поставленным целям, и в какой степени полученные данные справедливы в отношении изучавшейся выборки. Исходя из этого, достоверным нужно считать решение, в котором возможность возникновения систематических и случайных ошибок сведена к минимуму. При этом объективная информация всегда является достоверной, а субъективная информация не всегда достоверна.

Выбор методов оценки достоверности определяется подходом, который наилучшим образом учитывает структуру и контекст решаемой задачи. Например, при использовании статистического подхода к решению задачи выделяют параметрические и непараметрические методы оценки достоверности результатов, позволяющие перенести результаты выборочного решения на генеральную совокупность.

Параметрическими называют количественные методы статистической обработки данных, применение которых требует обязательного знания закона распределения изучаемых признаков в совокупности и вычисления их основных параметров. В тех случаях, когда имеется малое количество наблюдений и характер распределения неизвестен, когда кроме количественных характеристик результаты выражаются полуколичественными, а иногда описательными характеристиками (например, для медицинского исследования – тяжесть заболевания, интенсивность реакции, результаты лечения), параметрические методы становятся непригодными. В этих ситуациях следует использовать непараметрические методы оценки достоверности.

Непараметрическими являются количественные методы статистической обработки данных, применение которых не требует знания закона распределения изучаемых признаков в совокупности и вычисления их основных параметров. В то же время следует отметить, что назначение применения непараметрических методов гораздо шире, чем только оценка достоверности результатов исследования (в том числе они применяются и для характеристики одной выборочной совокупности, и для изучения связи между явлениями). В рамках данной статьи, говоря о непараметрических статистических методах, исследуется только оценка достоверности результатов исследования.

Как параметрические, так и непараметрические методы, используемые для сравнения результатов исследований, т.е. для сравнения выборочных совокупностей, заключаются в применении определенных формул и расчете определенных показателей в соответствии с предписанными для того или иного метода алгоритмами. В конечном результате высчитывается определенная числовая величина, которую сравнивают с табличными пороговыми значениями. Критерием достоверности будет результат сравнения полученной величины и табличного значения при данном числе наблюдений (или степеней свободы) и при заданном уровне безошибочного решения.

Следует заметить, что достоверность информации является обобщенным показателем качества информации, обозначающим её точность и полноту. В общем смысле точность можно рассматривать как степень приближения истинного значения рассматриваемого параметра процесса, значения величины и т.д. к его теоретическому номинальному значению. Точность измерений –

характеристика качества измерений, отражающая близость к нулю погрешностей их результатов. Высокая точность измерений соответствует малым составляющим погрешностей всех видов (как случайных, так и систематических). Количественно точность может быть выражена значением, обратным модулю относительной погрешности измерения. Например, при относительной погрешности измерения, равной 2%, или 0,02, точность измерений равна 1/0,02 = 50.

В общем виде достоверность информации – это характеристика ее неискаженности. Решение многих сложных задач подразумевает разбиение их на этапы, которые зачастую идут последовательно один за другим. При этом информация, которая является результирующей для одного этапа задачи, может поступать в качестве входной для решения следующего этапа и т.д. Никакие методы и инструменты, задействованные на каждом этапе решения задачи, не могут повысить достоверность информации, являющейся входной для данного этапа. Напротив, для выбранного набора методов и инструментов обработки можно оценить степень искажения первоначальной информации. Таким образом, в ходе решения задачи от этапа к этапу, достоверность информации может лишь снижаться.

Говоря об оценке достоверности информации, в ряде случаев целесообразно проводить это оценивание на промежуточном этапе. Это имеет смысл, когда известны допустимые пороговые значения достоверности для каждого этапа. При этом можно отследить этап, при котором уровень достоверности информации стал недостаточным, а, значит, выполнять оставшиеся этапы бессмысленно. Такой подход, с одной стороны, требует временные и вычислительные ресурсы для получения промежуточных оценок, но, с другой стороны, при недостаточном уровне достоверности информации позволяет сэкономить аналогичные ресурсы за счет сокращения этапов решения задачи. Знание того, на каком этапе уровень достоверности информации стал недопустимым, может помочь в исправлении ситуации.

Если нет возможности получить количественную оценку достоверности информации, как входной, так и результирующей, то ее оценивают по шкалам: чаще всего надёжная, полностью надёжная, довольно надёжная и так далее до абсолютно ненадёжной и той, у которой статус не определён. Информация, которую невозможно проверить на достоверность, является бессмысленной.

Важным показателем качества информации является ее релевантность *R*. Релевантность информации определяется как степень соответствия цели, для которой она требуется. Принято выделять полезную информацию и информационный шум. Их взаимное отношение и определяет степень релевантности информации. Одна и та же информация может быть полезной для решения одних задач и бесполезной для решения других. Таким образом, вычисление релевантности должно быть связано с определенным классом решаемых задач.

То, какая информация является полезной, а какая – нет, следует решать априорно, либо же рассматривать процесс решения задачи как систему с отрицательной обратной связью, и тогда степень полезности информации будет коррелировать с уровнем неопределенности получаемого решения.

Информация считается полной *G*, если ее достаточно для снятия неопределенности при решении задачи. Процесс решения можно считать оправданным, если поступающая входная информация уменьшает существовавшую ранее неопределенность результата [27]. При этом количество информации *I* может быть выражено как разность между априорной $H_{apr}$ и апостериорной $H_{apost}$ энтропиями:

$$I = H_{apr} - H_{apost}. \qquad (2)$$

С использованием вероятностного подхода, согласно Шеннону [28], энтропия исследуемого информационного атрибута *Attr* может быть выражена как

$$H_{Attr} = -\sum_{i=1}^{k_{Attr}} P_{i\,Attr} \log_2 P_{i\,Attr} \ , \qquad (3)$$

где $P_{iAttr}$ — вероятность определенного значения для атрибута *Attr* входной информации (точное соответствие, соответствие с учетом допуска, вхождение в диапазон, принадлежность категории и т.д.);

$k_{Attr}$ — количество разных значений для атрибута *Attr*, $i$=1..$k_{Attr}$.

Как видно, степень полноты информации нелинейно зависит от имеющихся вероятностей значений атрибутов. При этом для однократных пропусков данных величина вероятности значения определенного атрибута может быть восстановлена в любом случае, а для многократных – только при равновероятных значениях атрибута, причем для случая многократных пропусков должно быть известно общее число измерений значения атрибута.

Своевременность информации *T* означает, что она присутствует в системе в тот момент времени, когда ее необходимо обрабатывать. Часто решение некоторой сложной задачи распадается на несколько этапов, для каждого из которых известны момент времени начала информационной обработки и исходная для этого этапа информация. При этом следует учесть, что вся информация, поступившая на вход некоторого этапа после того, как процесс обработки информации уже начался, не может быть использована при решении. Заметим, что чрезмерно раннее получение входной информации приводит к необходимости ее хранения до момента использования, что требует дополнительных системных ресурсов.

Понятность информации *U* – это качество, которое основывается на степени соответствия вида представления информации виду ее восприятия. Здесь нужно четко разграничивать категории пользователей данной информации. Если пользователем является техническая система, то информация должна быть представлена в виде сигналов. Если же пользователем является человек, то качество понятности может быть выражено через качество человеко-машинного интерфейса. Низкий уровень понятности сводит на нет многие усилия по получению качественной информации.

Объективности, а также определенного уровня достоверности и полезности входной информации можно достичь, предъявляя четкие требования к уровням компетентности источников информации. При этом следует учитывать механизм получения входной информации для решения задачи: формулирование мнения эксперта на основании его опыта решения подобных задач, съем показаний измерительного прибора, подсчет величин, формулировка заключения на основании значений статистических характеристик, знания аналитических зависимостей и др. Методы обеспечения качества информации для этих категорий совершенно разные. Далее будет рассмотрена только ситуация, при которой входная информация определяется мнением эксперта на основании его опыта решения подобных задач.

Рассмотрим задачи, решением которых является получение значения целевого свойства объекта (наименования класса, интегральной оценки) на основании значений свойств, описывающих этот объект. При этом разные решения будут соответствовать разным поставленным целям.

Пусть, например, в рамках решения задачи классификации пользователя интересует одно наиболее вероятное значение класса, к которому относится исследуемый объект, и при этом классы нельзя сравнивать между собой и, соответственно, упорядочивать. Тогда качество полученного решения определяется логической функцией соответствия найденного класса теоретическому номинальному с учетом вероятности принадлежности решения найденному классу. В случае, когда классы сравнимы между собой, качество решения можно определить, как величину, обратную расстоянию между центральными характеристиками (центр тяжести, математическое ожидание) найденного и теоретического номинального классов также с учетом вероятности принадлежности решения найденному классу.

В другой ситуации пользователь может отталкиваться от минимально допустимого значения вероятности правильного решения. Для задачи классификации это означает определение набора классов, вероятности принадлежности объекта к которым больше либо равны минимально допустимому значению, и качество решения определяется полнотой такого набора.

Для задачи интегрального оценивания состояния сложного объекта также можно рассматривать разные ситуации. Например, определение наименьшего диапазона значений целевой характеристики с заданной вероятностью, или же определение вероятности попадания значения целевой характеристики в заданный диапазон значений. Как видно, в каждой из описанных ситуаций качество решения задачи

выражается по-своему, но все они имеют вероятностную составляющую.

В случае, когда имеющийся уровень компетентности источника информации не позволяет обеспечить требуемое качество, необходимо либо повышать уровень компетентности (прямая задача), либо снижать требования к качеству (обратная задача), либо, если это невозможно, использовать альтернативные подходы для решения поставленной задачи.

Компетентности источников информации могут определяться априорно или апостериорно. Если источник поставляет информацию многократно для решения одной и той же либо подобной задачи, его апостериорная компетентность принимается за априорную.

Априорная компетентность источника информации может определяться одним из способов либо их совокупностью:

– назначением, сделанным ответственным лицом. Например, для случая экспертного оценивания – лицом, принимающим решение, либо сторонним экспертом, исходя из их собственных представлений о компетентности этого источника;

– самооценкой, при которой источник информации самостоятельно задает свой уровень компетентности;

– при помощи группового экспертного оценивания [29, 30].

Последний метод предполагает обработку полученных данных экспертного оценивания, которые характеризуют обобщенное мнение и степень согласия индивидуальных оценок экспертов. Обработка данных экспертов служит исходным материалом для синтеза прогнозных гипотез, проведения классификации, формирования заключения о состоянии исследуемого объекта или системы и т.д.

Наиболее распространенными экспертными методиками при классификации по признаку оценки преимущества при принятии решений являются следующие:

– методика ранжирования;

– методика непосредственного оценивания;

– методика сопоставления, включающая две разновидности: методику последовательного сопоставления и методику попарного сравнения.

Все они имеют много общего, а их отличие состоит, преимущественно, только в том, что оценивание исследуемых объектов осуществляется разными способами. При этом каждая методика имеет свои преимущества и недостатки.

Концепция групповой экспертизы рассматривает содержательную часть экспертизы как процесс разрешения проблемы компетентности, в котором обычно различают следующие этапы: формирование коллектива экспертов; получение экспертной информации; обработка экспертной информации с целью принятия решения [31, 32]. Проблема подбора экспертов является одной из наиболее сложных. Очевидно, в качестве экспертов необходимо использовать тех людей, чьи суждения наилучшим образом помогут принятию адекватного решения. Формирование результата группового экспертного оценивания заключается в нахождении для поставленной задачи верного (искомого) решения, обладающего достаточной объективностью.

Числовым выражением согласованности мнений экспертов является коэффициент конкордации. Оценка согласованности мнений экспертов необходима в первую очередь потому, что мнения экспертов могут сильно расходиться по оцениваемым параметрам. Зачастую изначально оценку проводят по ранжированию показателей и присвоению им определенного коэффициента значимости (весомости). Несогласованное ранжирование приводит к тому, что данные коэффициенты будут статистически недостоверными. Мнения экспертов при их необходимом количестве (более 7-10) должны быть распределены по нормальному закону. Коэффициент конкордации – это безразмерная величина, показывающая в общем случае отношение дисперсии к ее максимальному значению. Коэффициент конкордации выражается числом от 0 до 1, показывающим согласованность мнений экспертов при проведении ранжирования каких-либо свойств. Чем ближе это значение к 0, тем согласованность считается более низкой. При величине данного коэффициента менее 0.3 мнения экспертов считаются несогласованными. При нахождении величины коэффициента в диапазоне от 0.3 до 0.7 согласованность считается средней, при величине более 0.7 – высокой.

Если об источнике информации нет предварительных данных, позволяющих априорно оценить его компетентность, можно задать для него некоторое среднее значение компетентности.

В работах [33, 34] было предложено вычислять апостериорную компетентность эксперта-источника информации на основании знания его априорной компетентности и коэффициента доверия, который может быть определен автоматически путем сравнения информации, поступающей от него и от других источников. Разумеется, эксперт, чье мнение сильно отличается от мнения других экспертов, может оказаться ближе к истине, чем остальные, но вероятность такой ситуации гораздо меньше, чем обратной. Поэтому такой эксперт чаще всего либо оказывается менее компетентным, чем остальные, либо намеренно искажает информацию.

Вычисление коэффициента доверия основывается на понятии опорного факта. В качестве опорного факта принимается либо усредненное значение факта (данных, знаний), полученное от группы источников, либо один, но точно установленный факт. Оценивается расстояние между значением, полученным в i-й итерации от определенного эксперта, и опорным фактом. Измеряемые характеристики могут быть получены на числовой шкале, либо же на шкалах порядка или наименований в нормированном евклидовом пространстве посредством приведения их все к той же числовой шкале. Если величина полученного расстояния превышает некоторый порог, то значение, предоставленное в данной итерации экспертом, считается несоответствующим истине. В результате коэффициент доверия для такого эксперта должен понизить его уровень компетентности. И наоборот, если значение, полученное от эксперта, достаточно близко к опорному факту, то компетентность эксперта должна возрасти. Коэффициент доверия, как шаг изменения уровня компетентности, может быть как фиксированной величиной, так и зависеть от степени несоответствия значения, полученного в i-й итерации от определенного эксперта, опорному факту.

Отдельно следует решать вопрос как поступить, если уровень компетентности эксперта столь низок, что выходит за область допустимых значений. Необходимо выбрать наилучший подход в зависимости от условий, сопутствующих решению задачи. Например, такого эксперта можно сразу исключать из круга лиц, поставляющих исходную информацию. Это целесообразно делать в случаях, когда мнение эксперта бесповоротно признано бесполезным, и обработка получаемых от него данных лишь напрасно потребляет вычислительные ресурсы. Такая ситуация влечет решение вопроса о назначении нового эксперта на место выбывшего либо же уменьшении числа экспертов.

Другим подходом может стать требование получения от такого эксперта нескольких значений за каждую i-ю итерацию решения задачи, усреднения этих значений, и рассмотрения только лишь такого усредненного значения. Этот подход может быть удобен в тех случаях, когда условия решения задачи динамически меняются, и мнение эксперта зависит от этих условий.

Еще одним подходом может стать игнорирование низкого уровня компетентности эксперта и использование получаемых от него значений только в случае их соответствия опорному факту. Такая ситуация может стать полезной, когда требуется выявить область экспертности какого-либо специалиста. Предлагая ему задачи из различных (но чаще смежных) областей знаний, можно провести границы экспертности.

Открытым пока остается вопрос о связи между уровнями компетентности экспертов и их требуемым количеством в рамках решения конкретной задачи. Логичным кажется факт, что при наличии экспертов, уровень компетентности которых стремится к максимальному значению, нет смысла в их большом количестве.

## III. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Рассмотрим зависимость качества решения задачи обработки данных от требований, предъявляемых к источникам исходных данных, и выбранных методов обработки и анализа данных на примере задачи прогнозирования. Исследуем зависимость оценки точности прогноза от степени полноты исходных данных с использованием статистических методов.

В качестве исходных данных используем результаты, полученные украинскими

абитуриентами при прохождении внешнего независимого оценивания, которое предполагает получение по каждому из сдаваемых предметов целочисленного значения результата по шкале от 0 до 200 баллов. В качестве исследуемого предмета выбрана «математика», анализируемый временной период – 2010–2018 годы. Исследуемую выборку составили выпускники трех школ Одессы общим количеством 127 человек. Данные взяты с открытого ресурса Украинского центра оценивания качества образования [35].

В связи с тенденцией к повышению нижних пороговых результатов, необходимых для поступления в высшие учебные заведения Украины, интерес представил анализ процентной части абитуриентов с результатами внешнего независимого оценивания в диапазоне [150; 200] баллов, а также прогностические оценки по данному диапазону на 2019 год с учетом предположения о возможности распространении прошлых и настоящих тенденций и закономерностей на будущее развитие выбранного объекта прогнозирования. Первоначальное оценивание прогноза проводилось на полной выборке абитуриентов путем последовательной экстраполяции методами скользящей средней, наименьших квадратов и экспоненциального сглаживания. Затем с использованием тех же методов был вычислен прогноз на неполных исходных данных, полученных по двум школам из трех (83 человека), и проведено оценивание точности рассчитанных прогнозов в точках экстраполяции.

Применение метода скользящей средней позволило элиминировать случайные колебания и получить с его помощью краткосрочные прогностические значения, соответствующие влиянию главных факторов. В таблице 1 приведено прогностическое решение методом скользящей средней на основании полных и неполных выборок за 2010–2018 годы проведения внешнего независимого оценивания. Использованы следующие обозначения:

$V$ – данные временного ряда на полной выборке абитуриентов, результат которых попал в диапазон [150; 200] баллов;

$m$ – скользящая средняя временного ряда на полной выборке;

ε – средняя относительная ошибка;

$V'$ – данные временного ряда на неполной выборке абитуриентов с результатом от 150 до 200 баллов;

$m'$ и ε' – скользящая средняя и средняя относительная ошибка на неполной выборке соответственно.

Таблица 1[1].

Решение методом скользящей средней[2]

| $t$ | $V$, % | $m$, % | ε, % | $V'$, % | $m'$, % | ε', % |
|---|---|---|---|---|---|---|
| 2010 | 88.1 | | | 85.2 | | |
| 2011 | 93.1 | 88.04 | 5.43 | 88.5 | 84.10 | 9.67 |
| 2012 | 82.93 | 87.15 | 5.09 | 78.6 | 82.40 | 0.64 |
| 2013 | 85.42 | 85.75 | 0.38 | 80.1 | 79.53 | 6.89 |
| 2014 | 88.89 | 87.41 | 1.66 | 79.9 | 81.53 | 8.28 |
| 2015 | 87.93 | 88.19 | 0.30 | 84.6 | 84.77 | 3.60 |
| 2016 | 87.75 | 84.29 | 3.94 | 89.8 | 83.20 | 5.19 |
| 2017 | 77.2 | 74.75 | 3.17 | 75.2 | 74.40 | 3.63 |
| 2018 | 59.3 | | | 58.2 | | |
| Прогноз (forcast) 2019 | 68.78 | | | 68.73 | | |

[1,2] Appendix 1

К имеющимся полной и неполной выборкам применили следующий типовой подход. Для расчета прогнозного значения сперва рассчитывали скользящую среднюю, а затем прогнозируемый показатель. Расчет скользящей средней выполнялся по формуле:

$$m_t = \frac{V_{t-1} + V_t + V_{t+1}}{n}, \qquad (4)$$

где $V_{t-1}$, $V_t$ и $V_{t+1}$ – результаты оценивания абитуриентов за прошлый, текущий и следующий годы соответственно;

$n$ – число уровней, входящих в интервал сглаживания, принято $n$=3.

Прогнозируемый показатель для полной и неполной выборок определялся стандартным образом:

$$V_{t+1} = m_{t-1} + \frac{1}{n}(V_t - V_{t-1}), \qquad (5)$$

где $t + 1$ – прогнозный год;

$t$ – год, предшествующий прогнозному;

$V_{t+1}$ – прогнозируемый результат оценивания абитуриентов;

$m_{t-1}$ – скользящая средняя за два года до прогнозного;

$V_t$ – фактическое значение результата оценивания за предыдущий год;

$V_{t-1}$ – фактическое значение результата оценивания за два года, предшествующих прогнозному.

Расчет средних относительных ошибок по каждому году для полной ($\varepsilon_t$) и неполной ($\varepsilon_t^{'}$) выборок выполнялся по формулам:

$$\varepsilon_t = \frac{|V_t - m_t|}{V_t} * 100\%, \quad \varepsilon_t^{'} = \frac{|V_t - m_t^{'}|}{V_t} * 100\%, \qquad (6)$$

где $V_t$ – результат оценивания абитуриентов за текущий год, взятый из полной выборки:

$m_t$ и $m_t^{'}$ – скользящие средние за текущий год для полной и неполной выборок соответственно.

Итоговая средняя относительная ошибка для полной выборки показала высокую точность результата прогнозирования:

$$E = \frac{1}{k}\sum_{t=1}^{k}\varepsilon_t = 2.85\% \qquad (7)$$

Расчеты средней относительной ошибки для неполной выборки выявили некоторое снижение точности прогностического решения, что, однако, не является достаточно критичным:

$$E^{'} = \frac{1}{k}\sum_{t=1}^{k}\varepsilon_t^{'} = 5.41\% \qquad (8)$$

Это означает, что общность закономерностей, которые проявляются в результатах внешнего независимого оценивания, позволяет строить достаточно точный прогноз методом скользящей средней в условиях неполных данных.

Для тех же наборов данных построены прогнозы еще двумя методами.

Применение метода наименьших квадратов состоит в минимизации суммы квадратических отклонений между наблюдаемыми и расчетными величинами, которые находятся по подобранному уравнению – уравнению линейной регрессии со скалярной переменной. Чем меньше расстояние между фактическими значениями и расчетными, тем более точен прогноз, построенный на основе уравнения регрессии. В таблице 2 приведено прогностическое решение методом наименьших квадратов на основании полных и неполных выборок соответственно. Прогнозируемое значение показателя для полной (и аналогично для неполной) выборки вычислялось по формуле:

$$V_{t+1} = a * X_{t+1} + b, \qquad (9)$$

где $t + 1$ – прогнозный год;

$V_{t+1}$ – прогнозируемый показатель;

$a$ и $b$ – коэффициенты;

$X_{t+1}$ – условное обозначение времени.

Расчет коэффициентов $a$ и $b$ осуществляется следующим образом:

$$a = \frac{k\sum_{t=1}^{k}(V_t * X_t) - \sum_{t=1}^{k}X_t\sum_{t=1}^{k}V_t}{k\sum_{t=1}^{k}X_t^2 - (\sum_{t=1}^{k}X_t)^2} = -2.51, \qquad (10)$$

$$b = \frac{\sum_{t=1}^{k}V_t - a * \sum_{t=1}^{k}X_t}{k} = 95.96,$$

где $k$ – число уровней временного ряда.

Аналогичные вычисления для неполной выборки дали $a'$= -2.02, $b'$=90.09.

Расчет средних относительных ошибок по каждому году для полной ($\varepsilon_t$) и неполной ($\varepsilon_t^{'}$) выборок на основании расчетных значений $m_t$ и $m_t^{'}$ выполнялся по формулам (6).

Таблица 2[3].

Решение методом наименьших квадратов[4]

| *t* | *X* | *V*, % | *m*, % | ε, % | *V*', % | *m*', % | ε', % |
|---|---|---|---|---|---|---|---|
| 2010 | 1 | 88.1 | 93.45 | 6.08 | 85.2 | 88.08 | 0.03 |
| 2011 | 2 | 93.1 | 90.94 | 2.32 | 88.5 | 86.06 | 7.56 |
| 2012 | 3 | 82.93 | 88.43 | 6.63 | 78.6 | 84.04 | 1.34 |
| 2013 | 4 | 85.42 | 85.91 | 0.58 | 80.1 | 82.03 | 3.97 |
| 2014 | 5 | 88.89 | 83.40 | 6.17 | 79.9 | 80.01 | 9.99 |
| 2015 | 6 | 87.93 | 80.89 | 8.01 | 84.6 | 77.99 | 11.30 |
| 2016 | 7 | 87.75 | 78.38 | 10.68 | 89.8 | 75.98 | 13.42 |
| 2017 | 8 | 77.2 | 75.86 | 1.73 | 75.2 | 73.96 | 4.20 |
| 2018 | 9 | 59.3 | 73.35 | 23.70 | 58.2 | 71.94 | 21.32 |
| Прогноз (forcast) 2019 | | 70.84 | | | 69.93 | | |

Вычисления итоговых средних относительных ошибок для полной ($E$) и неполной ($E'$) выборок по формулам (7) и (8) также показали достаточно высокую точность результатов прогнозирования: $E$=7.32%, $E'$=8.12%. Это хуже, чем при использовании метода скользящей средней, однако, если установлено требование, чтобы ошибка не превышала 10%, то применение метода наименьших квадратов при отсутствии части исходных данных можно считать приемлемым.

Применение метода экспоненциального сглаживания обеспечивает простоту процедуры вычислений. Для расчета прогноза использована рабочая формула метода экспоненциального сглаживания:

$$m_{t+1} = \alpha V_t + (1-\alpha) m_t, \qquad (11)$$

где $t$ – год, предшествующий прогнозному;

$t$ +1 – прогнозный год;

$m_{t+1}$ – прогнозируемый показатель;

α – параметр сглаживания;

$V_t$ – фактическое значение исследуемого показателя за год, предшествующий прогнозному;

$m_t$ – экспоненциально взвешенная средняя для года, предшествующего прогнозному.

При прогнозировании данным методом возникает два затруднения: выбор значения параметра сглаживания α; определение начального значения $m_0$.

Значение параметра сглаживания определяли с учетом числа наблюдений, входящих в интервал сглаживания:

$$\alpha = \frac{2}{k+1}. \qquad (12)$$

Начальное значение $m_0$ определяли двумя способами: I способ – приравнивали к среднему арифметическому всех фактических значений исследуемого показателя, II способ – использовали первое фактическое значение исследуемого показателя.

В таблице 3 колонки $m_1$ и $\varepsilon_1$ содержат прогностические значения и их средние относительные ошибки, рассчитанные I способом, а колонки $m_2$ и $\varepsilon_2$ – II способом соответственно.

Вычисления итоговых средних относительных ошибок для полной ($E_1$, $E_2$) и неполной ($E_1'$, $E_2'$) выборок показали снижение точности результатов прогнозирования: $E_1$ = 9.04%, $E_2$ = 8.41%, $E_1'$ = 10.64%, $E_2'$ = 9.15%. Как видно, для неполной выборки ошибка превысила 10%, что не позволяет интерпретировать соответствующую точность как «высокую».

Таким образом, если к сформулированной задаче прогнозирования процентной части абитуриентов с результатами внешнего независимого оценивания в диапазоне [150; 200] баллов по дисциплине «математика» на изучаемой выборке предъявляются требования по точности, метод экспоненциального сглаживания на неполных данных не обеспечивает заданное качество решения.

[3,4] Appendix 1

Таблица 3[5].

Решение методом экспоненциального сглаживания[6]

| $t$ | $V$, % | $m_1$, % | $\varepsilon_1$, % | $m_2$, % | $\varepsilon_2$, % | $V'$, % | $m_1'$, % | $\varepsilon_1'$, % | $m_2'$, % | $\varepsilon_2'$, % |
|---|---|---|---|---|---|---|---|---|---|---|
| 2010 | 88.1 | 83.4 | 5.33 | 88.10 | 0.00 | 85.2 | 80.01 | 9.18 | 85.20 | 3.29 |
| 2011 | 93.1 | 84.34 | 9.41 | 88.10 | 5.37 | 88.5 | 81.05 | 12.94 | 85.20 | 8.49 |
| 2012 | 82.93 | 86.09 | 3.81 | 89.10 | 7.44 | 78.6 | 82.54 | 0.47 | 85.86 | 3.53 |
| 2013 | 85.42 | 85.46 | 0.05 | 87.87 | 2.86 | 80.1 | 81.75 | 4.29 | 84.41 | 1.18 |
| 2014 | 88.89 | 85.45 | 3.87 | 87.38 | 1.70 | 79.9 | 81.42 | 8.40 | 83.55 | 6.01 |
| 2015 | 87.93 | 86.14 | 2.04 | 87.68 | 0.28 | 84.6 | 81.12 | 7.75 | 82.82 | 5.81 |
| 2016 | 87.75 | 86.50 | 1.43 | 87.73 | 0.02 | 89.8 | 81.81 | 6.77 | 83.17 | 5.22 |
| 2017 | 77.2 | 86.75 | 12.37 | 87.73 | 13.64 | 75.2 | 83.41 | 8.05 | 84.50 | 9.45 |
| 2018 | 59.3 | 84.84 | 43.07 | 85.63 | 44.40 | 58.2 | 81.77 | 37.89 | 82.64 | 39.36 |
| Прогноз (forcast) 2019 | | 79.73 | 9.04 | 80.36 | 8.41 | | 77.05 | 10.64 | 77.75 | 9.15 |

## ВЫВОДЫ

Таким образом, в работе было сформировано множество (1) количественных и качественных критериев для формализации источников информации. В это множество вошли такие критерии, как объективность, достоверность, релевантность, полнота, своевременность и понятность. Требования, предъявляемые к источникам информации в соответствии с этими критериями, позволяют обеспечить заданное качество решения таких задач интеллектуальной обработки данных как квалиметрия, классификация, таксономия, диагностика, прогнозирование, принятие решений и др.

Были выявлены и сформулированы в обобщенном виде требования, предъявляемые к компетентности источников информации. Выявлены проблемы, возникающие при формировании результата группового экспертного оценивания, а также проанализированы методики определения компетентности экспертов с целью их учета при нахождении искомого решения поставленной задачи, которое обладает достаточной объективностью. Рассмотрены задачи, решением которых является получение значения целевого свойства объекта (интегральной оценки, наименования класса) на основании значений, получаемых от экспертов.

Для практического исследования зависимости качества решения задачи обработки данных от требований, предъявляемых к источникам информации, и от использованных информационных методов обработки и анализа данных выбрана задача прогнозирования. Исследование зависимости оценки точности прогноза от степени полноты исходных данных на примере результатов внешнего независимого оценивания показало целесообразность использования метода скользящей средней и метода наименьших квадратов для получения качественной прогностической оценки. При этом, несмотря на отсутствие почти трети исходных данных от компетентных источников, средняя относительная ошибка прогноза не превысила 10%.

Дальнейшее изучение процесса оценивания компетентности источников информации требует разработки четко выстроенных моделей исходных данных с учетом требований к их качеству, а также формализации целевых процессов обработки данных.

Наличие формальных моделей позволит частично автоматизировать процесс анализа и оценивания источников информации в задачах интеллектуальной обработки данных. При этом следует учесть такие ситуации, при которых входные данные являются разнородными, противоречивыми либо еще каким-нибудь образом ухудшают качество решения задачи.

**APPENDIX 1 (ПРИЛОЖЕНИЕ 1)**

[1,2]**Table 1.** Moving average solution.

[3,4]**Table 2.** Least-squares solution.

[5,6]**Table 3.** Exponential smoothing solution.

[5,6] Appendix 1

**Литература (References)**


[1] Boutkhoum O. Multi-Agent Based Modeling Using Multi-Criteria Decision Analysis and OLAP System for Decision Support Problems. *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering*, 2015, vol. 9, no. 12, pp. 2553-2560.

[2] Koncilia C., Morzy T., Wrembel R., Eder J. [Interval OLAP:Analyzing interval data] *Proceedings of the 16th International Conference on Data Warehousing and Knowledge Discovery, DaWaK*. Munich, 2014, pp. 233-244.

[3] Guerra J., Schunn C. D., Bull S., Barria-Pineda J., Brusilovsky P. Navigation support in complex open learner models: assessing visual design alternatives. *New Review of Hypermedia and Multimedia*, 2018, vol. 24, no. 3, pp. 160-192. doi 10.1080/13614568.2018.1482375

[4] Ishaq R., Nasim R. Enhancing information extraction techniques from structured database using artificial intelligence. *International Journal of Computer Science and Information Security*, 2018, vol. 16, no. 11, pp. 140-143.

[5] Hoifung P., Domingos P. Joint Inference in Information Extraction. *Association for the Advancement of Artificial Intelligence, USA*, 2015, vol. 34(5), pp. 171-176.

[6] Krisilov V.A., Gorodnichaya K.O., Huy Vu.N. [Method of adapting content by the volume of transmitted information on the Internet] *IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT*. Lviv, 2018, vol. 2, Article № 8526647. - Proceedings.

[7] Makarova Y., Krisilov V., Vu H.N., Langmann R. [User profile creation and training mode determination in the 'Smart lab' system] *4th IEEE Global Engineering Education Conference*, EDUCON, 2014, Article № 6826110, pp. 315-320.

[8] David J.C. MacKay. Information Theory, Inference, and Learning Algorithms. Cambridge University Press, 2005, 640 p.

[9] Spirtes P., Glymour C., Scheines R. Causation, prediction, and search. *Lecture Notes in Statistics*, New York: Springer-Verlag, 1993, vol. 81. 44 p.

[10] Parsaye K. A Characterization of Data Mining Technologies and Processes. *The Journal of Data Warehousing*, 1998, vol. 1, pp. 11-28.

[11] Michalski R.S. Machine Learning and Data Mining, Methods and Applications. N.Y.: John Wiley & Sons, 1998. 472 p.

[12] Hongjun L., Setiono R., Liu H.. NeuroRule: A Connectionist Approach to Data Mining. *Computer Science Learning, Cornell University Library, UK*, 2017, vol. 6(1), pp. 40-47.

[13] Chen B., Li P., Wu H., Husain T., Khan F. MCFP: a Monte Carlo simulation-based fuzzy programming approach for optimization under dual uncertainties of possibility and continuous probability. *J Environ Inform*, 2017; vol. 29(2), pp. 88-97.

[14] Tagasovska N., Andritsos P. Distributed clustering of categorical data using the information bottleneck framework. *Information Systems*, 2017, vol. 72, pp. 161-178.

[15] Lee I. H., Mahmood M. T. Adaptive outlier elimination in image registration using genetic programming. *Information Sciences*, 2017, vol. 421, pp. 204-217.

[16] Cohen P. R., Feigenbaum E. A. The handbook of artificial intelligence. *Butterworth-Heinemann*, 2014, vol. 3, 587 p.

[17] Krisilov V. A. Ocenka slozhnyh ob'ektov – osnovnoj mekhanizm pri reshenii zadach kolichestvennogo obosnovaniya reshenij [Evaluation of complex objects as the main mechanism for solving problems of quantitative substantiation of decisions] *Trudy Odesskogo politekhnicheskogo universiteta [Proceedings of the Odessa Polytechnic University]*. – Odessa, 2003, vol. 1 (19), pp. 102-106.

[18] Komlevaya N. O., Komlevoy A. N., Chernega K. S. Designing of the specialized computer system for making pulmonology diagnosis. *CEUR Workshop Proceedings, 9th International Conference of Programming, UkrPROG*, Kyiv, 2014, vol. 1843, pp. 253-263.

[19] Komleva N.O., Cherneha K.S., Tymchenko B.I., Komlevoy O.M. Intellectual approach application for pulmonary diagnosis. *Proceedings of the 2016 IEEE 1st International Conference on Data Stream Mining and Processing, DSMP*, 2016, Article № 7583505, pp. 48-52.

[20] Awawdeh S., Edinat A., Sleit A. Enhanced K-means Clustering Algorithm for Multi-attributes Data. *International Journal of Computer Science and Information Security (IJCSIS)*, 2019, vol. 17, no. 2, pp. 1-6.

[21] Rokach L. A survey of clustering algorithms. *Data mining and knowledge discovery handbook*, Springer, Boston, MA, 2009, pp. 269-298.

[22] Firdaus S., Uddin, M. A. A Survey on Clustering Algorithms and Complexity Analysis. *International Journal of Computer Science Issues (IJCSI)*, 2015, vol.12, is. 2, pp. 62-85.

[23] Goswami J. A Comparative Study on Clustering and Classification Algorithms. *International Journal of Scientific engineering and Applied Science (IJSEAS)*, 2015, vol. 1, is. 3, pp. 2395-3470.

[24] Huang K, Dai L, Yao M, Fan Y, Kong X. Modeling dependence between traffic noise and traffic flow through an entropy-copula method. *J Environ Inform*, 2017, vol. 29(2), pp. 134-151.

[25] Krisilov V. A., Komleva N. O., Prigozhev O. S. Navchalnyi posibnyk po dystsyplini "Teoriia informatsii ta koduvannia" dlia studentiv spetsialnosti 6.050103 «Prohramna inzheneriia» [Teaching manual on discipline "Theory of information and coding" for students of specialty 6.050103 "Software engineering". Grif of the Ministry of Education of Ukraine: No. 1/11-4058 dated 26.03.12]. Odessa, ONPU, 2012, 178 p. (In Ukrainian).
[26] Hamming R.W. Coding and Information Theory. *Prentice Hall*, 1986, 259 p.
[27] Cover T.M., Thomas J.A. Elements of Information Theory. 2006, 748 p.
[28] Shannon C. E. Matematicheskaya teoriya svyazi. Raboty po teorii informacii i kibernetike [Mathematical theory of communication. Works on information theory and cybernetics]. M., 1963. 830 p. (In Russian).
[29] Saaty T.L. The analytic hierarchy process. N.-Y.: McGraw Hill, 1980, 288 p.
[30] Ishizaka A., Labib A. Analytic Hierarchy Process and Expert Choice: Benefits and Limitations, ORInsight, 2009, vol. 22(4), pp. 201-220.
[31] Larichev O.I. Teoriya i metody prinyatiya reshenij: Uchebnik [Theory and methods of decision making: Textbook]. - M., 2002, 392 p. (In Russian).
[32] Drake P.R. Using the Analytic Hierarchy Process in Engineering Education. *International Journal of Engineering Education*, 1998, vol.14 (3), pp. 191-196.
[33] Zagorujko N.G. Prikladnye metody analiza dannyh i znanij [Applied methods of data and knowledge analysis]. Novosibirsk, 1999, 270 p. (In Russian).
[34] Zagorujko N.G. Doverie k informacii i ee istochniku v ekspertnoj sisteme [Confidence in information and its source in the expert system]. Ekspertnye sistemy i raspoznavanie obrazov [Expert systems and pattern recognition]. M., 1988, vol. 126, pp. 3-23. (In Russian).
[35] Ukrainian Center for Educational Quality Assessment. http://testportal.gov.ua/ (accessed 26.04.19).

**Сведения об авторах.**

**Крисилов Виктор Анатольевич**,
д. т. н., проф.,
заведующий кафедры «Системное программное обеспечение» Института компьютерных систем Одесского национального политехнического университета
krissilovva2014@gmail.com

**Комлевая Наталия Олеговна**,
к.т.н., доцент кафедры «Системное программное обеспечение» Института компьютерных систем Одесского национального политехнического университета
nkomlevaya@gmail.com