

ОЦЕНКА УРОВНЯ ВИРТУАЛЬНОЙ ИНТЕЛЛЕКТУАЛЬНОСТИ ПРИКЛАДНОЙ ПРОГРАММНО-ТЕХНИЧЕСКОЙ СИСТЕМЫ НА ОСНОВЕ АНАЛИЗА ЭРГОНОМИЧЕСКОЙ МОДЕЛИ

Abstract: It is noted shortcomings of approaches to estimations of the system intelligence with the elements of the artificial intellect based on the ideology IQ coefficient in combination with expert estimations of their components. It is proposed the development of these approaches in the direction of ergonomic model of the interaction process "user-computer" and analysis of the intellectual load on user.

Key words: level of intellectuality of systems, IQ ideology, ergonomic models, intellectual load of user, user interface.

Аномація: Відзначені недоліки підходів до оцінки інтелектуальності систем з елементами штучного інтелекту, що базуються на ідеології коефіцієнтів IQ у сполученні з експертними оцінками їхніх складових. Пропонується розвиток цих підходів у напрямку оцінки порівняльного рівня віртуальної інтелектуальності на основі побудови ергономічної моделі «користувач-комп'ютер» і аналізу інтелектуального навантаження на користувача.

Ключові слова: рівень інтелектуальності систем, ідеологія IQ, ергономічні моделі, інтелектуальне навантаження користувача, інтерфейс користувача.

Аннотация: Отмечаются недостатки подходов к оценке интеллектуальности систем с элементами искусственного интеллекта, которые базируются на идеологии коэффициентов IQ в сочетании с экспертными оценками их составляющих. Предлагается развитие этих подходов в направлении оценки сравнительного уровня виртуальной интеллектуальности на основе построения эргономической модели процесса взаимодействия «пользователь-компьютер» и анализа интеллектуальной нагрузки на пользователя.

Ключевые слова: уровень интеллектуальности систем, идеология IQ, эргономические модели, интеллектуальная нагрузка пользователя, интерфейс пользователя.

1. Введение

Идея количественной оценки уровня «естественного» интеллекта человека с помощью системы тестов и определяемого на их основе коэффициента интеллектуальности IQ впервые была предложена и методически развита еще в начале прошлого века. С тех пор IQ успешно используется в мире для разнообразных целей, включая отбор в армию, специализацию и профориентацию, оценку степени подготовленности обучаемых, формирование критериев социологических исследований и т.п.

Вопросы оценки уровня интеллектуальности компьютерных систем и информационных технологий (ИТ) пока остаются мало исследованной областью. Сами понятия и термины «интеллектуальная», «интеллектуализированная», «интеллектуализация» применительно к ИТ используются в большом количестве работ по теории и практике создания информационных и информационно-аналитических систем разных классов и назначений. Однако пока что единого толкования этих терминов в контексте количественных оценок меры интеллектуальности (интеллектуализированности) не сложилось. Общим недостатком ряда предлагаемых определений [1–6] является использование терминов, связанных с интеллектуализацией, исключительно в интуитивном, неоперационном понимании.

2. Постановка задачи

Один из возможных подходов к количественной оценке интеллекта программно-технических систем заключается в развитии и приложении идеологии IQ к определению понятия «интеллектуализированные системы» и оценке уровня их интеллектуальности.

В частности, подход, предложенный авторами в [7], заключается в использовании оценок, основанных на:

- сравнении характеристик рассматриваемой системы с общим набором базовых показателей некоторой идеализированной интеллектуальной системы (ИИС), по своим характеристикам отвечающей нашим теперешним представлениям про таковую. Результатом такого сравнения должно быть значение абсолютного коэффициента интеллектуальности AAIQ;
- определении степени соответствия значений выбранных проблемно-ориентированных показателей интеллектуальности, характеризующих системы заданного класса, заданным (желаемым) значениям. В результате такого сравнения оценивается относительный коэффициент интеллектуальности RAIQ.

В качестве основы предложенного набора $W^{(A)}$ из 6 базовых показателей использованы предложения Мичи (Mitchi) относительно определения искусственного интеллекта [8], которые считались наиболее операциональными [9]; набор $W^{(R)}$ проблемно-ориентированных показателей формируется в зависимости от класса и функций оцениваемых систем.

Подход [7] и некоторые близкие частные проблемно-ориентированные решения [12] имеют тот общий недостаток, что они базируются на экспертных оценках. Если для AAIQ экспертные оценки пока единственно возможны (в частности, из-за значительного разнообразия и незначительного количества экземпляров однотипных систем, близких к ИИС), то для оценки интеллектуальных свойств прикладных проблемно-ориентированных систем возможны более объективные оценки. Предлагаемое решение, дополняющее экспертные оценки [7], основано на том очевидном положении, что основным назначением элементов «интеллектуальности», вводимых в прикладную компьютерную систему, является снижение интеллектуальной нагрузки на пользователя системы (так же, как для робототехнических систем и физической нагрузки). Поэтому степень снижения интеллектуальной нагрузки может быть использована в качестве критерия кажущейся виртуальной интеллектуальности прикладной программно-технической системы, независимо от ее структуры и внутренних свойств.

3. Оценка интеллектуальной нагрузки

Определим интеллектуальную нагрузку $IH(R_x)$ на пользователя в реализации R_x некоторой задачи (системы, технологического этапа) как сумму

$$IH(R_x) = \sum_{s(x)} \sum_{g(x)} m_{s(x), g(x)} r_{s(x), g(x)}, \quad (1)$$

где $m_{s(x), g(x)}$ – количество операций класса $g(x)$, которые должны выполняться человеком в функции (процедуре) $s(x)$ реализации R_x ;

$r_{s(x), g(x)}$ – интеллектуальная трудоемкость операции класса $g(x)$ функции $s(x)$ реализации R_x при ее выполнении человеком (в выбранных адекватных единицах).

Значение показателя CAIQ представим в виде

$$CAIQ = \frac{IH(R_0) - IH(R_x)}{IH(R_0)}, \quad (2)$$

где R_0 – некоторая базовая реализация, с которой производится сравнение.

Как можно видеть из (2), для $IH(R_x) = 0$ значение $CAIQ = 1$, для $IH(R_0) = IH(R_x)$ значение $CAIQ = 0$, для $IH(R_x) > IH(R_0)$ значение $CAIQ < 0$. Последнее означает, что оцениваемая реализация R_x хуже базовой (в рассматриваемом смысле).

Для выбора функций $s(x)$, классов операций $g(x)$ и оценок трудоемкостей $g(x)$ целесообразно построение эргономических моделей выполнения операций человеком при решении рассматриваемой задачи.

Относительно выбора модели можно сделать следующие замечания.

Существует много методов количественного анализа интерфейса «человек-компьютер» [10]. В качестве прототипа модели для поставленных нами целей целесообразно использование модели типа GOMS [10, 11]. Моделирование GOMS позволяет оценить затраты времени пользователя на выполнение отдельных базовых операций взаимодействия с компьютером. Поскольку речь идет о сравнительных оценках реализаций R_0, R_x , представляется возможным ограничиться упрощенной моделью, в которой оценки $g(x)$ производятся не через абсолютные значения затрат времени, а через количество типовых операций, «нагружающих» пользователя, и их теоретическую трудоемкость.

Рассмотрим в качестве примера, иллюстрирующего использование упрощенных моделей когнитивной эргономики человека, оценку значений $CAIQ$ для задачи клавиатурно-ориентированного ввода и корректировки данных. Решение задачи заключается в том, чтобы:

- а) установить факт отсутствия или наличия ошибки для каждого вводимого слова ($s1$);
- б) идентифицировать и исправить ошибку в ошибочном слове ($s2$).

В качестве оцениваемого варианта примем использование системы автоматического (программного) контроля и «ручного» исправления ошибок с помощью клавиатуры. В качестве базового варианта – визуальный контроль и ручное исправление. В табл. 1 приведены сводные варианты реализации функций рассматриваемой задачи.

Таблица 1. Функции для сравнения

R	s	
	($s1$)	($s2$)
R_0	Визуально (монитор – исходный текст)	Визуально – вручную (монитор – исходный текст – клавиатура)
R_x	Программно	Визуально – вручную (монитор – исходный текст – клавиатура)

Процедура визуального контроля очередного слова (функция $s1$ реализации R_0) заключается в визуальном сравнении контролируемого слова с соответствующим словом исходного текста (документом) группами по α символов, где α – объем в символах «буферной» памяти человека при посимвольном сравнении двух слов.

Процедура «ручного» исправления ошибок (функция $s2$) состоит из следующих операций:

- подвод курсора к ошибочному символу в исправляемом слове;
- «стирание» ошибочного символа;
- выбор на клавиатуре символа замены;
- вставка нового символа;
- повторение предыдущих операций в соответствии с кратностью ошибки;
- подтверждение завершения исправления.

Приняв для упрощенной эргономической модели в качестве единицы измерения трудоемкость сравнения символов (по траекториям «монитор-документ», «документ-клавиатура» и опуская промежуточные оценки отдельных операций для ориентированных значений G_0, G_1 , получим

$$IH(R_0) = N[F_{\text{конт.}} + \sigma F_{\text{испр.}}]; \quad (3)$$

$$IH(R_x) = N[(1 - \beta) \cdot F_{\text{конт.}} + \sigma F_{\text{испр.}}], \quad (4)$$

где

$$F_{\text{конт.}} = n / \alpha + 1;$$

$$\sigma = n \cdot \pi_c;$$

$$F_{\text{испр.}} = \left(\frac{n+1}{2} + \log_2 q + 3 \right).$$

В (3), (4) приняты следующие обозначения и основные упрощающие допущения:

N – общее количество обрабатываемых слов;

n – количество символов в слове;

q – алфавит;

π_c – статистическая вероятность искажения символа;

α – объем группы символов в процессе визуального сравнения двух слов;

β – относительное количество слов, поддающихся автоматическому обнаружению.

Учитываются только однократные транскрипции, пропуски и добавления одиночных символов, т.е. наиболее вероятные ошибки, искажающие один символ.

Из (2), (3) вытекает

$$CAIQ = \frac{N \beta F_{\text{конт.}}}{N(F_{\text{конт.}} + \sigma F_{\text{испр.}})} = \frac{\beta(n + \alpha)}{n + \alpha(\sigma F_{\text{испр.}} + 1)}.$$

Приняв эмпирическое значение $\alpha = 3 \div 4$, $n = 10$, $\pi_c = 8,8 \cdot 10^{-3}$ [11], для значения CAIQ при $\beta = 1$ получим $CAIQ_{\max} \approx 0,78 \div 0,75$.

Полученное значение имеет тот смысл, что для принятых данных оцениваемый вариант по сравнению с базовым снижает интеллектуальную трудоемкость на 75%–78%. Принятое значение $\beta = 1$ соответствует потенциальному (т.е. максимально возможному) значению CAIQ. Реальное значение CAIQ линейно зависит от β и равно 0 для $\beta = 0$ (т.е. в этом случае функция $s1$ для R_x "не работает"). Последняя ситуация имеет место, например, при вводе исключительно числовых неконтролируемых данных.

Кажущаяся парадоксальной зависимость интеллектуальности от значения β находит свое объяснение в концепции «востребованности» интеллекта: для $\beta < 1$ потенциально – интеллектуальная система ввода и контроля применяется там, где ее возможности не могут быть использованы в полной мере, а лишь частично.

Примечание. Последнее замечание в совокупности с аналогичными рассуждениями относительно зависимости от значений π_c приводит к естественной постановке задачи о приемлемости цены, т.е. затрат на повышение CAIQ. Это – отдельный вопрос, общий для всех способов оценки степени интеллектуальности программно-технических систем. Однако расчетные значения именно CAIQ вносят объективную составляющую в решение задачи.

4. Заключение

В целом, показатели интеллектуальности AIQ имеют следующий смысл:

- значение AAIQ определяет, насколько рассматриваемая интеллектуализированная система близка к ИИС;
- значение RAIQ определяет, в какой степени система отвечает конкретным проблемно-ориентированным требованиям, предъявляемым к ее интеллектуальным возможностям;
- значение CAIQ определяет, в какой мере рассматриваемая система снижает интеллектуальную нагрузку на пользователя в сравнении с аналогичными системами соответствующего функционального назначения.

Такие оценки, особенно RAIQ и CAIQ, дают потенциальную возможность выявления «слабых звеньев» и перспективных направлений повышения уровня интеллекта прикладных программно-технических систем и, соответственно, основанных на их использовании ИТ. В частности, приложение результатов оценки RAIQ, CAIQ могло бы, наряду с [12], дополнить известную методику NSTL (National Software Testing Laboratory, USA) по тестированию программных продуктов. Термин «виртуальная интеллектуальность» применим ко всем приведенным коэффициентам, т.к. оценки производятся по прагматическим внешним кажущимся проявлениям интеллектуальности вне зависимости от внутренней структуры (например, наличия «развитой» базы знаний) и используемых алгоритмов. Примерно таким же подходом мы пользуемся, когда оцениваем ум человека по его материальной обеспеченности («если ты такой умный, то почему ты такой бедный»).

Приведенный пример основан на упрощенной, более или менее грубой эргономической модели. Для иллюстраций предлагаемого подхода это обстоятельство не имеет принципиального значения, тем более, что при сравнительных оценках оба варианта находятся в примерно одинаковых условиях. Для других задач и реализаций (например, автоматического исправления ошибок по словарю допустимых слов [14]) эргономические модели могут быть более точными, в частности, могут учитывать различные классы типовых ошибок пользователя, особенности языка вводимых сообщений и т.п.

В заключение авторы хотели бы выразить надежду, что изложенные подходы, при всем их осознанном несовершенстве, создают мотивы для дальнейших исследований в этом направлении и получения более удачных решений относительно методов и критериев количественной оценки уровня интеллекта ИС и ИТ.

СПИСОК ЛИТЕРАТУРЫ

1. Толковый словарь по искусственному интеллекту / Сост.: А.Н. Аверкин, М.Г. Гаазе-Рапопорт, Д.А. Поспелов. – М.: Радио и связь, 1999. – 256 с.
2. Великий тлумачний словник сучасної української мови. – Київ, Ірпінь: ВТФ «Перун», 2004. – 1440 с.
3. Суворов В.В. К оцениванию интеллекта систем // Труды Международной конференции «Интеллектуальное управление: новые интеллектуальные технологии в задачах управления (ICIT). – Переяславль – Залесский, 1999. – 6-9 декабря. – С. 36–42.
4. [http://chemphys.edy.ru/pdf\(2003-12-29-001.pdf\)](http://chemphys.edy.ru/pdf(2003-12-29-001.pdf)).
5. <http://www.dvags.ru/uk/4.6.html>.
6. <http://www.informag.ru/journals/j082r/25422.html>.
7. Кузьменко Г.С., Литвинов В.А. Прагматичний підхід до оцінки рівня інтелекту інтелектуалізованих систем // Математичні машини і системи. – 2003. – № 3, 4. – С. 75–79.
8. Michie D., Sibert E.E. Formation and Execution of Plans by Machine // Artificial Intelligence and Heuristic Programming, Findler N.V. and Meltzer B. (eds), Edinburgh University Press. – P. 101–124; Интеллектуальные роботы / Пер. с англ. – М.: Мир, 1975. – Вып. 2. – С. 378–405.
9. Попов Э.В., Фирдман Г.Р. Алгоритмические основы интеллектуальных роботов и искусственного интеллекта. – М.: Наука, 1976. – 455 с.
10. Раскин Д. Интерфейс. Новые направления в проектировании компьютерных систем. – Санкт-Петербург – Москва: Символ, 2006. – 268 с.
11. John, Bonnie E. “Why GOMS?” // Interaction. – 1995. – October. – P. 80–89.
12. Айвазян С.А. Интеллектуализированные инструментальные системы в статистике и их роль в построении проблемно-ориентированных систем поддержки принятия решений // Обзорение проблем прикладной математики. – М.: Наука, изд-во «ТВП», 1997. – Т. 4, № 2. – С. 47–58.
13. Литвинов В.А., Крамаренко В.В. Контроль достоверности и восстановление информации в человеко-машинных системах. – Киев: Техніка, 1986. – 200 с.
14. Алгоритми і моделі автоматичної ідентифікації на критерії типових помилок користувача на основі природної надмірності / Г.Є. Кузьменко, В.А. Литвинов, С.Я. Майстренко та ін. // Математичні машини і системи. – 2004. – № 2. – С. 134–148.

Стаття надійшла до редакції 22.02.2008