

МЕТОДИКА КІЛЬКІСНОЇ ОЦІНКИ ІНФОРМАЦІЇ В СИСТЕМАХ ЕЛЕКТРОННОГО НАВЧАННЯ

Обґрунтована необхідність розробки уточненої методики чисельної оцінки кількості інформації на основі семантики її вмісту. Запропоновано підходи до побудови такої методики та математичні засади її використання. Пропонується здійснювати оцінку у відносних одиницях (відсотках), причому кількість інформації у певному конкретному факті, формулі, схемі, і т.д. рівна відношенню часу викладання цього матеріалу на занятті до всього обсягу дисципліни в годинах. Можливий варіант – знаходження відношення площі, що займає у підручнику викладення даного матеріалу, до всієї його площі. Описана практична галузь застосування методики, зокрема в системах дистанційної освіти.

Ключові слова: кількість інформації, електронне навчання, дистанційна освіта.

Постановка проблеми. Під час навчання людина весь час отримує нову інформацію, в чому, в цілому, і полягає суть цього процесу. Часто було б зручно кількісно оцінити інформацію, видану суб'єкту (наприклад, при розробці навчальних планів, робочих програм, тощо) або сприйняту ним (наприклад, під час перевірки та контролю знань). Особливо різко ця необхідність проявляється в системах електронного навчання. В цьому випадку існує можливість автоматизованого контролю і управління процесом навчання, для чого, однак, необхідно оперувати чисельними характеристиками навчальної інформації. Таким чином, існує потреба у чисельному визначенні кількості інформації, що отримує людина під час навчання.

Аналіз останніх досліджень і публікацій. У традиційній літературі з інформаційних технологій кількість інформації найчастіше оцінюють з позицій теорії інформації у синтаксичному аспекті та приймають як міру знятої ентропії після опрацювання повідомлення (наприклад, [1]):

$$I_{\beta}(\alpha) = H(\alpha) - H_{\beta}(\alpha), \quad (1)$$

де $H(\alpha)$ – апіорна ентропія системи α (до отримання інформації β , кількість якої оцінюємо);

$H_{\beta}(\alpha)$ – ентропія системи α після опрацювання повідомлення β ;

$I_{\beta}(\alpha)$ – кількість інформації, що надає повідомлення (інформація) β про систему α .

Незручність такого підходу для оцінки кількості інформації, що міститься у навчальному матеріалі очевидна, адже сформувавши функцію ентропії за Шенноном для неформалізованих певних предметних відомостей та знань не можливо.

Більш адекватним поставленій задачі є варіант визначення кількості інформації за її семантикою. Однак, один із важливих недоліків існуючих підходів полягає у достатньо широкому свавіллі особи, що проводить оцінку. Так, конкретне значення коефіцієнта змістовності C ([2]) однієї і тієї ж інформації може дуже сильно відрізнятися при його оцінці різними експертами.

Використання розповсюдженої тезаурусної міри оцінки кількості інформації [3] не є зручним для систем освіти, так як дозволяє визначити кількість інформації саме для одного конкретного суб'єкта, який сприймає інформацію (відповідно до його особистого тезауруса). Наприклад, якщо якийсь студент пропустив заняття з вищої математики про теорію поля, то йому практично не можливо буде зрозуміти рівняння Максвелла з курсу загальної фізики, тобто кількість інформації при її оцінці за тезаурусною мірою буде невеликою. Однак, зрозуміло, що знання про рівняння Максвелла складають суттєву частину сучасної загальної фізики і кількість інформації, що в них закладена є величиною об'єктивною, постійною.

Відповідно до постановки задачі та аналізу існуючих методів її вирішення, можна сформулювати наступну мету дослідження: розробити методику завдання кількості інформації, що міститься в навчальному матеріалі, на основі його семантичного вмісту.

Основна частина. Першим кроком такої методики є формування множини V видів найменших нерозривних частин навчального матеріалу з даного предмету. Наприклад, із фізики можна виділити наступні види таких «атомів»: $V = \{\text{«означення фізичної величини»}, \text{«опис досліду»}, \text{«формула»}, \text{«закон чи правило у текстовій формі»}\}$. З предмету «історія» можлива наступна множина видів елементарної інформації: $V = \{\text{«факт»}, \text{«подія»}, \text{«біографія історичної особи»}, \text{«опис певного соціуму (держави, братства, ложі, і т.д.)»}\}$. В цілому можлива ієрархічна організація видів навчальної інформації. Наприклад, Перша світова війна може вважатися подією високого рівня, що включає події нижчого рівня, факти, біографії, тощо. Однак, ієрархічність ніяк не впливатиме на кількісну оцінку навчальної інформації, як це буде зрозуміло із викладення наступного матеріалу.

Другим кроком є структуризація всього навчального матеріалу з даної дисципліни, тобто розбиття його на певні невеликі нерозривні частини тих типів, що містяться у множині V (окремі факти, формули, положення, тези, і т.д.). Цю задачу відносно легко може виконати викладач відповідного предмету, що має певний досвід (експерт). В результаті формується набір множин V_1, V_2, \dots, V_i кількістю, рівною потужності множини V . У кожній такій множині міститься вичерпний перелік усіх неподільних елементів навчального курсу відповідного i -того типу (наприклад, у множині V_1 можуть бути усі формули з курсу механіки, а у множині V_2 – усі означення, що використовуються у цьому курсі).

Кінцево кожному елементу усіх множин V_i слід поставити у відповідність деяку чисельну величину, що характеризує кількість інформації I_{ij} , що містить у собі елемент. Зручно задавати таку величину у відносних одиницях (у порівнянні з повним обсягом курсу). Цей третій крок може бути виконаний двома шляхами.

По-перше, можливий кількісний аналіз часу t_{vij} , що відводиться на викладання певного j -того елемента V_{ij} із V_i (враховуючи пояснення, коментарі, приклади, іншу будь-яку допоміжну інформацію, що його стосується) у порівнянні із загальним обсягом часу t , відведеного на даний предмет:

$$I_{ij}^{(1)} = t_{vij} / t. \quad (2)$$

Цей підхід вимагає наявності дуже детальної робочої програми дисципліни та великого досвіду лектора, що має викладати матеріал збалансовано, вкладаючись у встановлені рамки часу та при цьому, який приділяє увагу усім фактам, причому пропорційно до їх важливості у курсі. Отже, в даному випадку значною є залежність кількісної характеристики інформації, закладеної у певному факті, від особи викладача, що проводить заняття.

Другий підхід у більшій мірі позбавлений суб'єктивізму експерта. Аналіз часу викладання певного факту може замінитися аналізом об'єму викладення даного «атому» в одному із загальноприйнятих підручників. Тут ми спираємося на думку лише одного експерта – автора книги, однак, цей підхід не гірший за попередній, а кращий, оскільки компетентність цього одного експерта зазвичай набагато більше, ніж у великої кількості простих викладачів. Отже, для визначення кількості інформації слід знайти відношення:

$$I_{ij}^{(2)} = S_{V_{ij}} / S, \quad (3)$$

$$I_{ij}^{(2)} = N_{V_{ij}} / N, \quad (4)$$

де $S_{V_{ij}}$ – площа у підручнику, відведена на викладення та пояснення факту V_{ij} ;

S – загальна площа підручнику, відведена під навчальний матеріал;

$N_{V_{ij}}$ – кількість рядків у підручнику, відведена на викладення та пояснення факту V_{ij} ;

N – загальна кількість рядків у підручнику, відведена під навчальний матеріал.

Формулу (4) легше застосувати на практиці, так як порахувати кількість рядків легше, ніж виміряти площу у квадратних сантиметрах деякої ділянки сторінки. Однак, варіант (3) дозволяє враховувати інформацію не тільки текстового характеру, а й графічну, табличну, і т.д. Іноді дуже суттєві відомості містяться саме на рисунках, тому підхід з використанням (3) в таких випадках буде єдино правильним (наприклад, в курсі нарисної геометрії чи живопису). В той же час, якщо пролистати підручник історії чи інших гуманітарних дисциплін, то можна помітити, що практично весь матеріал складається виключно з тексту. Зрозуміло, що тут логічно використовувати (4).

Звернемо увагу, що

$$I_i = \sum_j I_{ij} \neq 1,$$

так як ця сума дає лише частку матеріалів i -того типу із V_i серед усього навчального матеріалу за k -тою дисципліною. Порівняння значень таких сум для різних i дозволяє визначити в якій мірі даний предмет наповнений формулами, рисунками, фактами, тощо. Така інформація може бути корисною для аналізу якості друкованого навчального матеріалу. Наприклад, відомо, що для ліпшого засвоєння інформації один розворот книги повинен містити хоча б один-два рисунки, тобто повинно бути дотримане певне мінімальне співвідношення між текстовими фактами та рисунками. З іншого боку, перевантаженість тексту формулами також може сильно знизити ефективність його сприйняття.

Оскільки формули (3) та (4) дають результат в однаковій одиниці виміру (у відносних одиницях, тобто безрозмірній), то їх можна застосовувати разом (для різних предметів), а результати об'єднувати будь-яким способом (адитивно, мультиплікативно, зважено і т.д.).

Описаний підхід до кількісної оцінки навчальної інформації дозволяє проводити аналіз якості викладання та засвоєння матеріалу. Для оцінки якості викладання достатньо порівняти міру інформації, пораховану з одного боку за (2), а з іншого – за (3) чи (4) в залежності від суті k -тої дисципліни (зокрема, технічна вона чи гуманітарна). Зрозуміло, що відношення цих величин весь час повинно бути близьким до одиниці:

$$Q_{ij} = \frac{I_{ij}^{(1)}}{I_{ij}^{(2)}} \approx 1. \quad (5)$$

Відхилення (5) для деяких V_{ij} в меншу сторону легко дозволяє виявити відомості, які викладач «не полюбляє», а, отже, або погано розуміє, або підзабув, або необгрунтовано (суб'єктивно) вважає малозначущими. Якщо (5) менше одиниці для більшості V_{ij} , то це можливе тільки за рахунок дуже високого значення (5) для невеликої кількості відомостей

V_{ij} , які, відповідно, є улюбленими темами викладача, що він занадто розвиває у збиток іншим. Аналогічно, якщо (5) більше одиниці для більшості фактів з k -того курсу, то слід шукати невелику кількість відомостей з дуже малим значенням Q_{ij} : це і будуть ті V_{ij} , на які слід звернути увагу даному конкретному викладачеві, які він практично не дає на заняттях.

За допомогою (3-4) також стає можливою і більш важлива у системах електронної освіти здатність аналізу навчання окремої особи. Сам процес навчання полягає у вивченні певних відомостей V_{ij} , отже, якщо говорити про один конкретний предмет, то цей процес у першому наближенні може бути представлений функцією:

$$K(t) = \sum_i \sum_j (I_{ij} \cdot h(t - t_{ij})), \quad (6)$$

де $K(t)$ – кількість знань за даним предметом на момент часу t ;

$h(t)$ – функція у вигляді шабля, яка до нульового моменту рівна нулю, а після нього і в цей нульовий момент рівна одиниці;

t_{ij} – час вивчення j -тих відомостей i -того типу відповідно до робочої програми даного предмету; слід звернути увагу, що мова йде про час саме закінчення вивчення матеріалу.

Весь процес вивчення дисципліни при застосуванні (6) представляється ступеньковою функцією – рис.1. Кожен шабелів відповідає засвоєнню певного факту: висота шабля рівна кількості інформації у цьому факті, а його розташування на часовій осі відповідає величині t_{ij} . В кінці процесу вивчення дисципліни $K(t) = 1$.

Розглянутих величин достатньо щоб оперувати такою характеристикою, як швидкість засвоєння знань:

$$u_{ij} = I_{ij} / \Delta t, \quad (7)$$

де u_{ij} – швидкість засвоєння факту V_{ij} ;

Δt – інтервал часу, необхідний на засвоєння факту V_{ij} .

В середньому швидкість засвоєння знань даним конкретним суб'єктом є величиною більш-менш постійною $u_{i1} \approx u_{i2} \approx \dots \approx u_{ij}$ для даного i -того їх типу. Для кожного елемента із V_i швидкість засвоєння може бути різною: рисунки можуть сприйматися відносно швидше, ніж формули, чи текст, або навпаки. В цілому ж для даного конкретного суб'єкта нахил ламаної повинен бути приблизно однаковим, якщо оперувати значними інтервалами часу (щоб нівелювати вплив різного настрою, втоми, конкретного розділу дисципліни і т.д.). Відповідно, ламана $K(t)$ буде наближатися до прямої пропорційності, коефіцієнт якої даватиме загальну швидкість засвоєння знань за даною дисципліною.

Таким чином, система електронного навчання на основі інформації, що запитувалася і пророблена учнем, може вираховувати $K(t)$ на даний момент часу і контролювати u , щоби швидкість засвоєння знань не була занадто високою (користувач мало вникає у навчальний матеріал, швидко проглядає його), або занадто низькою (користувач відволікається, при цьому забуває попередній матеріал, не є «в темі»).

Студент чи учень в процесі навчання працює із кількома дисциплінами, отже одночасно його описує N незалежних функції $K_n(t)$. Тоді поточний стан користувача в системі електронного навчання можна представити точкою у N -мірному просторі знань, координати якої $\{K_1(t), K_2(t), \dots, K_n(t)\}$. Процес засвоєння знань можна представити рухом у такому N -мірному просторі. Система електронного навчання має контролювати цей рух, адже у просторі знань існують заборонені стани, обумовлені взаємозалежністю дисциплін. Наприклад, поки з курсу математики не вивчено поняття похідної, у курсі фізики не можливе вивчення швидкості, як похідної від шляху, прискорення – як похідної від швидкості, і т.д. Більш укрупнено такі залежності (між окремими дисциплінами) видно із структурно-логічної схеми кожної спеціальності.

Таким чином, можна говорити про потенціальне поле, задане у багатовимірному просторі знань. Значення потенціалу рівне сумарній інформації, засвоєній на даний момент часу по всім дисциплінам:

$$K(t) = \sum_n K_n(t) = \sum_n \sum_{i_n} \sum_{j_n} (I_{i_n j_n} \cdot h(t - t_{i_n j_n})). \quad (8)$$

Висновки. В роботі запропоновано підхід по оцінці кількості знань, що міститься в навчальному матеріалі. Кількість інформації, що міститься в конкретному факті, може бути знайдена як відносна величина, рівна відношенню часу викладання цього факту до всього часу дисципліни (або як відношення площі, відведеної на викладання даного факту у підручнику до всієї площі, і т.д.). Даний підхід дозволяє розраховувати чисельні показники процесу засвоєння знань, що при правильному управлінні може покращити його якість та ефективність. В подальшому планується уточнення даної методики шляхом апроксимації процесу засвоєння факту V_{ij} іншими видами залежностей, а не лише функцією у вигляді щабля.

ЛІТЕРАТУРА:

1. Згуровський М.З., Коваленко І.І., Міхайленко В.М. Вступ до комп'ютерних інформаційних технологій: Навчальний посібник. – К.: Вид-во Європейського університету, 2003. – 265 с.
2. Заболотский В.П., Оводенко А.А., Степанов А.Г. Математические модели в управлении: Учеб пособие / СПбГУАП. СПб., 2001. - 196 с.
3. Перевозчикова О.Л. Інформаційні системи і структури даних: Навчальний посібник для студентів вищих навчальних закладів. – К.: Видавничий дім «Києво-Могилянська академія», 2007. – 287 с.

Рецензент: д.т.н., проф. Мочалов О.О., директор Інституту заочної та дистанційної освіти Національного університету кораблебудування імені адмірала Макарова

к.т.н., доц. Гайша А.А.

МЕТОДИКА КОЛИЧЕСТВЕННОЙ ОЦЕНКИ ИНФОРМАЦИИ В СИСТЕМАХ ЭЛЕКТРОННОГО ОБУЧЕНИЯ

Обоснована необходимость разработки уточненной методики численной оценки количества информации на основе семантики ее содержимого. Предложены подходы к построению такой методики и математические основы ее использования. Предлагается осуществлять оценку в относительных единицах (процентах), причем количество информации в определенном конкретном факте, формуле, схеме, и т.д. равно отношению времени преподавания этого материала на занятии ко всему объему дисциплины в часах. Возможен вариант - нахождение отношения площади, которая занимает в учебнике изложения данного материала, ко всей его площади. Описана практическая область применения методики, в частности в системах дистанционного образования.

Ключевые слова: количество информации, электронное обучение, дистанционное образование.

A. Gaisha

TECHNIQUE OF QUANTITATIVE INFORMATION EVALUATION IN ELECTRONIC LEARNING SYSTEMS

The necessity of development of numerical evaluation of information quantity technique based on its contents semantics is proved. The approaches to construct such technique and its mathematical basis is proposed. It is proposed to realize the evaluation in relative units (percents), and quantity of information in some concrete fact, formulae, scheme, etc. equals to relation of time teaching this material at the lesson to the whole discipline volume in hours. Alternative variant – finding relation of square in the book of this material to the whole square of book. Practical field of application of technique is described, particularly in distance education systems.

Keywords: quantity of information, electronic learning, distance education.