

В.В. ЛИТВИН, канд. техн. наук, доц., Національний університет
«Львівська політехніка»

Д.І. УГРИН, канд. техн. наук, доц., Буковинський університет, м.Чернівці

МЕТОД АВТОМАТИЧНОЇ РОЗБУДОВИ АДАПТИВНОЇ ОНТОЛОГІЇ

Розглядається метод автоматичної розбудови адаптивної онтології. Залежно від типу задачі наведено дві процедури визначення ваг важливості концептів та відношень онтології.

Ключові слова: адаптивна онтологія, ваги важливості концептів та відношень, база знань, інтелектуальна система.

Рассматривается метод автоматической перестройки адаптивной онтологии. В зависимости от типа задачи приведены две процедуры определения веса важности концептов и отношений онтологии.

Ключевые слова: адаптивная онтология, веса важности концептов и отношений, база знаний, интеллектуальная система.

The method of the automatic re-erecting of adaptive ontology is examined. Depending on as a task two procedures of determination of vag importance of konceptiv and relations of ontology are resulted.

Keywords: adaptive ontology, scales of importance of konceptiv and relations, base of knowledges, intellectual system.

1. Постановка проблеми у загальному вигляді

Ефективність адаптації онтології бази знань до особливостей предметної області (ПО) визначають закладені в її структуру елементи та механізми її адаптації шляхом самонавчання під час експлуатації. Одним з підходів до реалізації таких механізмів є автоматичне зважування понять бази знань (БЗ) та семантичних зв'язків між ними під час самонавчання. Цю роль беруть на себе коефіцієнти важливості концептів та зв'язків. Коефіцієнт важливості концепту (зв'язку) – це чисельна міра, котра характеризує значущість певного концепту (зв'язку) у конкретній ПО і динамічно змінюється за певними правилами у процесі експлуатації системи. Визначену таким чином онтологію ми назвали адаптивною. Вона адаптується до ПО за рахунок модифікації понять та коефіцієнтів важливості цих понять і зв'язків між ними.

Використання адаптивних онтологій в складі інтелектуальних систем підтримки прийняття рішень (ІСППР) показало ефективність таких систем на практиці [1]. Така онтологія визначається кортежем з п'яти елементів: $O = \langle C, R, F, W, L \rangle$, де C – концепт (поняття), R – відношення між поняттями, F – інтерпретація понять та відношень (аксіоми), W – важливість понять C , L – важливість відношень R . Аксіоми встановлюють семантичні обмеження для системи понять та відношень. Введення двох останніх скалярних величин дозволяє будувати метрику на таких онтологіях. Така метрика успішно використовується в прикладних ІСППР, описаних нами в [2, 3].

Для того, щоб вручну побудувати повну зв'язану онтологію для певної ПО необхідно затратити достатньо багато часу та ресурсів. Причина таких затрат

полягає в тому, що такі онтології повинні містити десятки тисяч елементів, щоб бути придатними для розв'язування широкого кола прикладних задач, які виникають у цих ПО. Тому ручна побудова онтології людиною-оператором – це рутинний процес, який вимагає ґрунтовних знань ПО та розуміння принципів побудови онтологій. Отже необхідно розробити методи та алгоритми автоматичної побудови онтологій [4].

2. Аналіз останніх досліджень та публікацій

Сфера автоматизації побудови онтологій представляє дуже великий інтерес, тому не дивно, що у світі існує багато напрацювань у цьому напрямку. Розглянемо декілька існуючих.

«Ontosophie» – напівавтоматична система для створення онтологій з текстів, розроблена David Celjuska та Maria Vargas-Vera з Knowledge Media Institute (Великобританія). Система дозволяє будувати онтологію із неструктурованого тексту, самонавчаючись в процесі побудови онтології і використовувати отримані знання для аналізу нових текстів.

«On-To-Knowledge» – керована контентом система керування знаннями на основі онтологій. Ця система призначена для аналізу і керування вмістом великих інформаційних мереж рівня корпорацій.

В IBM T.J. Watson Research Center (Нью-Йорк, США) розробляється метод побудови онтологій на основі наукових запитів із використанням технологій аналізу тексту. Цей метод формує онтологічні концепти і зв'язки на основі аналізу результатів пошуку текстів предметної області у мережі Інтернет.

Наша ідея, що лежить в основі автоматичної побудови онтологій, полягає в тому, що опрацьовані тексти із знаннями предметної області використовуються для отримання даних для доповнення існуючої онтології. У той же самий час проміжна онтологія використовується для опрацювання текстів ПО. У результаті отримуємо рекурсивний процес, який можна вважати самонавчанням системи.

Формування цілей: розробити математичне забезпечення розвитку адаптивної онтології на основі базової онтології предметної області та алгоритми автоматичного наповнення цієї онтології новими поняттями та зв'язками, та коефіцієнтами їх важливості на основі аналізу тестових ресурсів, що описують відповідну предметну область.

3. Основний матеріал

Розглянемо процес автоматичної розбудови деякої базової адаптивної онтології $O_{base} = \langle C_b, R_b, F_b, W_b, L_b \rangle$. Вважаємо, що базові терміни та відношення між ними повинні бути введені людиною-експертом в онтологію вручну. Для задання початкових ваг важливості термінів цієї онтології використовується один із двох методів: статистичний або на основі інтелектуального аналізу даних. Детально ці методи нами розглянуті у [5, 6]. Отже процес автоматичної розбудови онтології починається з моменту, коли в ній вже є якісь дані. Такий процес позначитимо:

$$\chi: O_{base} \rightarrow O \quad (1)$$

Онтологія – це мова науки. Мова науки, як структуроване наукове знання, задає собою багат шарове ієрархічне утворення, в якому виділяються блоки:

терміносистема; номенклатура; засоби та правила формування понятійного апарату і термінів.

Отже з точки зору процесу побудови онтології необхідно побудувати її терміносистему O_T та номенклатуру O_N . В нашому підході базова онтологія повинна точно включати в себе частину терміносистеми (див. рис. 1), тобто $O_B \cap O_T \neq \emptyset$.

Енциклопедії, термінологічні та толкові словники, на основі яких будується терміносистема ПО мають чітку структуру і складаються із словникових статей. Тому необхідно дослідити можливі їх структури з метою розпізнавання понять і відношень між ними. Побудова номенклатури складніше. Якщо в словниках терміни деяким чином вже виділені, то в наукових текстах (підручники, монографії і т.д.) їх необхідно виділяти, здійснити пошук властивостей понять і відношень між поняттями. Отже потрібна технологія природньо-мовного опрацювання наукового тексту.

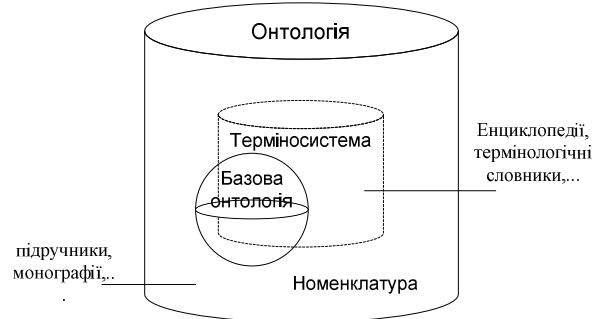


Рис. 1. Архітектура онтології

Позначимо множину властивостей концептів $V = \{v_1, v_2, \dots, v_s\}$. Тоді поняття відношення можна записати як відображення із C в C , зважене V : $R: C \xrightarrow{V} C$. Тоді відношення r_i представляє собою триплет:

$$r_i = \langle C_{i_1}, v_{ij}, C_{i_2} \rangle, \quad (2)$$

Терміни визначимо як клас

$$C = \langle N, R^x, R^y, S, D, A, Ob \rangle, \quad (3)$$

де N – ім'я терміну, R^x – множина відношень в яких клас C є доменем (областю визначення), R^y – множина відношень в яких клас C є множиною значень, S – суперкласи C , D – підкласи C , A – аксіоми визначення C , Ob – екземпляри C .

Отже в процесі розвитку базової онтології O_{base} необхідно будувати триплети r_i та нові поняття C , які задаються сімкою величин. У структуру концепту (3) входить набір аксіом A , однак автоматично побудувати такий набір аксіом дуже складно (принаймі автори не знають жодної такої спроби).

Тому наразі такий процес необхідно проводити вручну. Відношення ієрархії IS-A та PART-OF будемо називати вертикальними відношеннями і позначатимемо їх R^V .

Всі інші відношення називаються горизонтальними і позначатимемо їх R^H . Очевидно $R^V \cup R^H = R$, $R^V \cap R^H = \emptyset$.

Зважений концептуальний граф, ізоморфний адаптивній онтології є орієнтованим (див. рис. 2) [7].

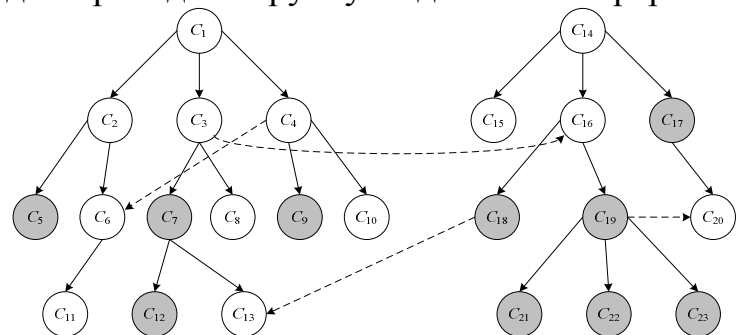


Рис. 2. Концептуальний граф онтології

Суцільними лініями наведені вертикальні зв'язки, пунктирними – горизонтальні. Визначеним класам відповідають темні вершини.

Для опрацювання природомовних текстів з метою розбудови онтології нами використано готові парсери [8]. Також нами розроблено шаблони відношень, на основі яких шукаються відповідні відношення у тексті. У залежності від типу задачі ваги відношень та методи обчислення цих ваг відрізняються. Розроблені нами шаблони, а також їх значущість для різних задач наведені у табл. 1. Групи відношень взяті із [10].

Табл. 1. Групи та ваги важливості відношень

Група відношень	Відношення	Шаблон (предикат)	Множина значень першого аргумента	Вага важливості L (семан./озн. задачі)
Ієрархія	Рід↔вид	PHier(a,x,y)	{Class, Kind}	0,9 (1,2)
	Ознака↔значення ознаки		{Category, Value}	0,9 (1,2)
	Інваріант↔варіант		{Invariant, Variant}	0,9 (1,2)
Агрегація	Ціле↔частина	PAggr(a,x,y)	{Whole, Part}	0,9 (1,2)
	Об'єкт↔простір реалізації (локалізації) об'єкта	PExist(a,x,y,z)	{Being, Location, Position, Order}	0,4 (1,2)
	Об'єкт↔властивість/ознака	PProp(a,x,y)	{Property, Indication, Character, Parameter, Factor, Criterion}	0,4 (1,2)
	рівень↔одиниця рівня	PAggr(a,x,y)	{Level, UnitLevel}	0,4 (1,2)
Функціо-нальні	Об'єкт дії↔дія↔суб'єкт дії	PFun(a,x,y,z)	{Function, Causal, Condition, Event, ActState, ObjectState, Tool, Data, Quantity}	0,3 (1)
	причина↔наслідок			0,3 (0,9)
	умова↔дія			0,3 (0,9)
	явище↔дія			0,3 (0,9)
	стан↔дія			0,3 (0,9)
	явище↔стан			0,3 (0,9)
	інструмент↔дія			0,3 (0,9)
	дані↔дія			0,3 (0,9)
	дані↔величини			0,3 (0,9)
Семіотичні	Термін↔спосіб вираження	PForm(a,x,y,z)	{Expression, Representation, MetaSign}	0,2 (0,2)
	Термін↔спосіб подання			0,2 (0,2)
	Термін↔метазнак терміну			0,2 (0,2)
Тотожності	Термін↔синонім терміну	PEquiv(a,x,y)	{Synonym, Quasisynonym}	1 (-)
Кореляції	Термін↔корелят терміну	PCor(a,x,y)	{Correlate, Oppose}	1 (-)

Для визначення правил обчислення ваг ми пропонуємо користуватись такою евристикою: для семантичних задач вага верхнього терміну ієрархії не може бути меншою за вагу свого нащадка. І навпаки, для ознакових задач, вага нащадка не може бути меншою за вагу свого батьківського поняття. Дійсно така евристика має сенс. Так для семантичних задач очевидно, що якщо мова у тексті йде про інтелектуальні системи, то й значить про штучний інтелект. А для ознакових задач краще мати значення найнижчих за ієрархією понять. Тому, виходячи з цього, перерахунок ваг для семантичних задач будемо здійснювати знизу до верху, а для ознакових навпаки – зверху до низу. Виходячи з таких міркувань нами визначено відповідні ваги відношень. Відношення тотожності та кореляції в ознакових задачах не використовуються, оскільки синонімія та королювання ніяк не впливають на значення ознак. Відразу вважається, що це одна і та ж ознака.

Розглянемо тепер методи задання початкових ваг (коефіцієнтів важливості) понять. Існують такі методи:

- за рахунок експертних оцінок;
- присвоєння випадковим чином;

за рахунок аналізу (статистичного, інтелектуального) інформаційних джерел, які описують ПО в якій функціонує ІА.

Окрім того ці ваги можуть мати обмеження на величину, наприклад їх значення знаходяться у відрізку $[0,1]$ (ймовірнісні методи) або без обмежень на величину (нагромаджувальні методи). Оскільки онтологія формує собою таксономію понять, то використовуючи мову об'єктно-орієнтованого підходу, кожне поняття являє собою клас.

Для семантичних задач нами запропоновано такий метод обчислення ваг понять (класів):

1. Повна вага W_j^i класу онтології дорівнює сумі власної ваги Wo_j^i , ваги підкласів Ws_j^i та ваги суміжних класів Wn_j^i (класів, зв'язаних з даним класом не IS-A зв'язком):

$$W_j^i = Wo_j^i + Ws_j^i + Wn_j^i, \quad (4)$$

де $Ws_j^i = \sum_k Wc_k^{i+1} \cdot L_{j,k}$ – вага k підкласів j -го класу i -го рівня, причому для кореневого класу рівень $i = 0$; $Wc_k^{i+1} = Wo_k^{i+1} + Ws_k^{i+1}$ – вага класу C_k^{i+1} ; $L_{j,k}$ – вага зв'язку між класами C_j^i та C_k^{i+1} .

Перерахунок окремих компонент повної ваги класу відображено на схемі (рис. 3).

2. У момент внесення на $i+1$ -й рівень нового підкласу йому присвоюється власна вага Wo_j^{i+1} , рівна половині власної ваги класу, вищого i -го рівня:

$$Wo_j^{i+1} = \frac{1}{2} Wo_j^i \quad (5)$$

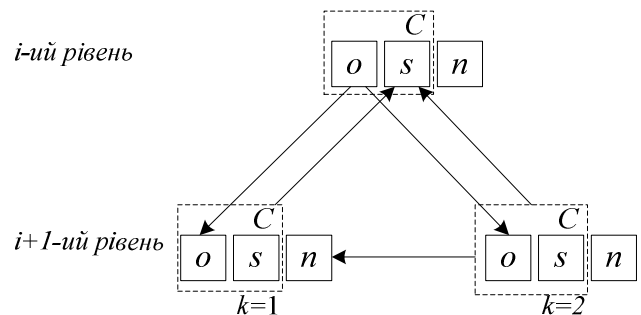


Рис. 3. Схема перерахунку окремих компонент повної ваги класу.

Вага класу Wc_j^i та усіх батьківських класів аж до кореневого збільшується на величину ваги новоствореного підкласу:

$$Wc_j^m = Wc_j^m + Wo_j^{i+1}, \forall m \leq i \quad (6)$$

3. Під час встановлення зв'язку між поняттями k_1 та k_2 між відповідними вершинами графа онтології з'являється ребро, а до ваги суміжних класів Wn_1 додається вага Wc_2 і навпаки – до Wn_2 додається вага нового суміжного до нього класу Wc_1 так, що:

$$Wn_j = \sum_k Wc_k \cdot L_{j,k} \quad (7)$$

Повторне встановлення зв'язків призводить до появи кратних ребер у графі.

4. Кратність ребер відображає частоту зустрічання U пари семантично пов'язаних понять $L_{i+1} = U \cdot L_i$. Кратні ребра після перерахунку не збільшують валентність вершини.

5. Вага екземпляра БЗ дорівнює повній вазі його класу.

Таким чином визначена модель онтології БЗ дає змогу розраховувати вагові коефіцієнти своїх компонентів у процесі їх додавання, вилучення і використання під час експлуатації системи, завдяки чому реалізує механізм адаптації до заданої користувачем ПО.

Розглянемо визначення ваг понять для ознакових задач на основі інтелектуального аналізу даних, а саме на основі побудови дерева рішень (ДР). Як відомо ознакові задачі дозволяють для пошуку релевантних прецедентів будувати ДР. Однак ДР не є панацеєю, оскільки згадувані ознаки, що лежать на відповідній гілці, що задає прецедент, не гарантують врахування повної множини ознак, які необхідно врахувати для знаходження релевантного прецедента. Нами пропонується використовувати ДР для визначення ваг базових термінів, які задають деякий прецедент, а потім на основі онтології ПО розвинути отримані ваги на всю онтологію для відповідного прецедента. Тоді для пошуку релевантного прецедента використовувати значення тих n понять, які для відповідного прецедента мають найбільші ваги.

Розглянемо гілку дерева рішень. Вершини (ознаки) цієї гілки знаходяться на k рівнях. Очевидно, що чим вищий рівень, тим значуща ознака, яка на цьому рівні знаходиться. Ця евристична думка має бути відображена в значеннях ваг цих ознак. Крім того пропонується ці ваги нормувати, тобто щоб їх сума для кожного прецедента (гілки) була рівна 1.

Розглянемо два способи визначення ваг базових ознак, які задовольняють вищі описані два припущення.

1 спосіб. Арифметичні ваги. Визначаються як відношення різниці $(k+1)$ рівня дерева та рівня на якому знаходиться ознака до суми всіх рівнів гілки, тобто базуються на сумі арифметичної прогресії:

$$w_i = \frac{k+1-i}{\sum_{j=1}^k j} = \frac{k+1-i}{\frac{(1+k)k}{2}} \quad (8)$$

2 спосіб. Геометричні ваги. Базуються на сумі геометричної прогресії:

$$w_i = \frac{2^{k-i}}{2^k - 1} \quad (9)$$

Отримані ваги назвемо вагами базових понять і позначимо таку множину ваг W_B . Тепер необхідно їх розвинути на всю онтологію ПО, використовуючи таксономію понять онтології, відношення між поняттями та їх інтерпретацію. Математично (формально) цей процес запишемо у вигляді:

$$W_B \xrightarrow{o} W. \quad (10)$$

Розмноження ваг на всю онтологію залежить від визначення (аксіоматизації) класів, їх ієрархії (вертикальний зв'язок) та горизонтальних зв'язків.

Для ознакових задач нами пропонується такий метод розповсюдження ваг. Спочатку ваги всіх ознак рівні 0. Для ознак, які приймають участь у ДР для відповідного прецедента до нуля додаємо вагу, отриману на основі ДР. Потім розглядаємо функціональні та семіотичні зв'язки, якщо вони визначені в онтології. Функціональні відношення діляться на симетричні R_s та несиметричні R_v . Очевидно, що елементи, які приймають участь у симетричних зв'язках є рівносильними. Тому ваги L симетричних відношень рівні 1. Всі інші функціональні зв'язки мають вагу 0,9. Отже, якщо відома вага W_i терміну C_i і цей термін має симетричний зв'язок із терміном C_j , вага якого невідома, то $W_j = L \cdot W_i$. Для несиметричних зв'язків $C_i \rightarrow C_j$ отримаємо аналогічне співвідношення, якщо відомо вагу C_i і $W_i = \frac{W_j}{L}$, якщо відомо вагу C_j .

Висновки

У роботі наведено базові положення математичного забезпечення автоматичної розбудови базової адаптивної онтології. Це забезпечення служить для пошуку відношень згідно визначених шаблонів та перерахунку ваг важливості концептів. Нами формально визначено поняття концепта в онтології та відношення між концептами, що дозволяє ефективно розбудовувати існуючу онтологію, уникаючи протиріч. Робота в більшій мірі спрямована на опис методів перерахунку ваг онтології залежно від типу задачі. Зазначимо, що розроблений метод стосується лише таксономії концептів та відношень між ними, однак не дозволяє автоматично будувати аксіоми, тобто визначати концепти. Ця задача є предметом наших подальших досліджень.

Список літератури: 1. В.В.Пасічник Інтелектуальні системи, базовані на онтологіях // Д.Г.Досин, В.В.Литвин, Ю.В.Нікольський, В.В.Пасічник – Львів: „Цивілізація”, 2009. – 414с. 2. Литвин В.В. Мультиагентні системи підтримки прийняття рішень, що базуються на прецедентах та використовують адаптивні онтології / В.В.Литвин // Радіоелектроніка, Інформатика, Управління. – Запоріжжя, 2009. – №2(21). – С. 120–126. 3. Крайовський В.Я. Використання адаптивних онтологій в інтелектуальних системах прийняття рішень / В.Я.Крайовський, В.В.Литвин, Н.Б.Шаховська // Східноєвропейський журнал передових технологій. – №4/3(40) – Харків, 2009. – С.7-12. 4. Литвин В.В. Автоматизація процесу розвитку базової онтології на основі аналізу текстових ресурсів / В.В.Литвин // Комп'ютерна та математична лінгвістика. Вісник НУ “Львівська політехніка” № 673, 2010. – С. 319-325. 5. Даревич Р.Р. Оцінка подібності текстових документів на основі визначення інформаційної ваги елементів бази знань / Р.Р.Даревич, Д.Г.Досин, В.В.Литвин, З.Т.Назарчук // Искусственный интеллект. – Донецк. – 2006. – № 3. – С. 500–509. 6. Даревич Р.Р. Метод автоматичного визначення інформаційної ваги понять в онтології бази знань / Р.Р. Даревич, Д.Г. Досин, В.В. Литвин // Відбір та обробка інформації. – 2005. – Вип. 22(98). – С. 105–111. 7. Литвин В.В.

Проектування інтелектуальних агентів прийняття рішень в просторі ознак з використанням онтологічного підходу / В.В.Литвин, Р.Р.Даревич, Д.Г.Досин, Н.В.Шкутяк // Штучний інтелект. – Донецьк-Кацивелі. – 2010. – т.2. – С. 100-104. 8. Литвин В.В. Проектування інтелектуальних агентів на основі адаптивних онтологій / В.В.Литвин, Н.Б.Шаховська, А.С.Мельник, О.Ю.Пшеничний, Ю.В.Ришковець // Міжнародна наукова конференція „Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту” ISDMCI’2010. – Євпаторія. – Т.2. – С.401-404. 10. Найханова Л.В. Технология создания методов автоматического построения онтологий с применением генетического и автоматного программирования / Л.В.Найханова. – Улан-Удэ: Издательство БНЦ СО РАН, 2008. – 244с.

Поступила в редколлегию 19.03.2011

УДК 664:519.8

О.І. ТОРЯНИК, докт. хім. наук, проф., ХДУХТ, м. Харків

О.Г. ДЬЯКОВ, канд. техн. наук, доц., ХДУХТ, м. Харків

Ж.В. ВОРОНЦОВА, канд. пед. наук, доц., ХДУХТ, м. Харків

ВИЗНАЧЕННЯ ТОЧОК ЕКСПЕРИМЕНТУ ПРИ НЕЛІНІЙНИХ МОДЕЛЯХ

Розглянуті питання визначення точок проведення експерименту коли модель нелінійна. Показано, що оптимальні точки проведення досліджень залежать від значень коефіцієнтів моделі та її структури

Ключевые слова:

Рассмотрены вопросы определения точек проведения эксперимента, когда модель нелинейная. Показано, что оптимальные точки проведения эксперимента зависят от значений коэффициентов модели и её структуры.

The questions of determination point realization experiment for nonlinear model are considered. It is shown, that optimal value realization experiment depend from defined value and structure of model.

Постановка проблеми у загальному вигляді

При проведенні досліджень харчових продуктів, у зв'язку з ускладненням речовин для дослідження, все більше використання набувають моделі які є суттєво нелінійними, тобто такі моделі, які не можуть бути спрощені шляхом розкладання у відповідні ряди чи шляхом використання спеціальних перетворень [1,2]. Тому питання побудови планів експерименту та подальшого знаходження коефіцієнтів моделі з відповідною точністю є на даний час актуальними. Принципова складність визначення плану експерименту для нелінійної моделі полягає у тому, що для визначення оптимального плану (плану, який дає можливість визначити коефіцієнти моделі з мінімальними похибками) треба мати відповідну уяву про значення істинних коефіцієнтів моделі. На перший погляд це виглядає повним протиріччям, але є цілком природним. Для ефективного планування експерименту необхідно добре знати його поведінку об'єкта в області планування. Тому у більшості випадків спочатку проводиться та обробляється невелика кількість дослідів з метою знаходження початкових оцінок параметрів, які необхідні для подальшого ефективного планування експерименту. У якості прикладу, де доцільно проводити планування точок експерименту, можна