УДК 519.81/.23

М. В. Андрєєв, д-р фіз.-мат. наук
Національний технічний університет України «КПІ», м. Київ

ТЕОРІЯ РІШЕНЬ В ЗАДАЧІ ВИБОРУ МОДЕЛІ ДЛЯ ПРОГНОЗУВАННЯ МАЙБУТНЬОГО СПОСТЕРЕЖЕННЯ

Стаття присвячена застосуванню теорії прийняття рішень в задачі вибору моделі для прогнозування майбутнього спостереження. Концептуальна основа є одним зі способів пошуку стохастичної моделі цього рішення. Вибір необхідної моделі робиться відповідно до відкритої чи закритої альтернатив. Ці випадки досліджуються засобами Байєсової теорії рішень.

Ключові слова: *байєсова теорія рішень, порівняння статистичних моделей, байєсів фактор, прогнозний розподіл, метод Монте-Карло, пересічне оцінювання.*

У рамках теорії прийняття рішень можливі різні концепції в залежності від того, які поняття вважаються основними при аналізі процесу прийняття рішень. Згідно з теоретико-ігровою концепцією проблема прийняття рішень представляє собою вибір найбільш привабливої альтернативи з множини можливих альтернатив. У рамках цієї концепції розглядається низка проблем оптимального вибору: деякі включають тільки вибір моделі; деякі включають вибір моделі, за яким слідує термінальна дія, така як прогнозування за вибраною моделлю; інші включають тільки термінальну дію. Всюди проводиться чітка різниця між трьома надто різними проблемами оптимального вибору: по-перше, припускається випадок, коли низка моделей, що розглядається, включає «істинну» модель; по-друге, випадок, коли розглядається низка моделей, аби забезпечити апроксимацію для визначеної, але важкодоступної дійсної моделі довіри; нарешті, випадок, коли розглядається низка моделей за відсутності специфікації фактичної моделі уподобань. При цьому встановлюються зв’язки процесу прийняття рішень з проблемами перевірки гіпотез, перевірки значущості статистик та пересічного оцінювання і прогнозування [1].

Порівняння і вибір статистичних моделей як проблема теорії рішень. У цій статті досліджується застосування теорії рішень до задачі статистичного висновку, що включає вибір моделі із множини альтернативних моделей $\mathbf{M} = \{M_i, i \in I\}$. Далі припускається, що дані спостережень, на яких базуються рішення, позначаються через $\mathbf{x}$, а вибір моделі M_i, як бази для послідовної відповіді на проблему статистичного висновку позначається через $m_i, i \in I,$ з подальшим прийняттям рішень $a_j,\ j \in J_i$ стосовно невідомого стану природи ω.

Елементи теорії рішень для цієї задачі представлено на рис. 1 у вигляді дерева рішень:

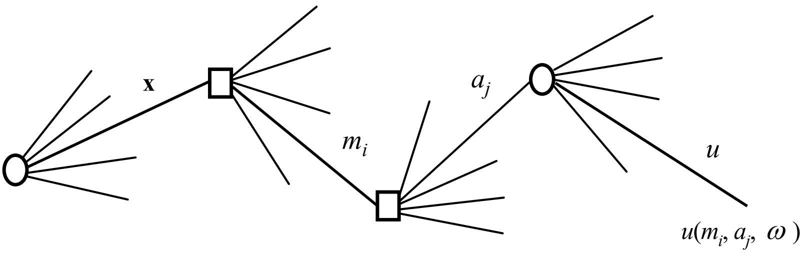

Рис. 1. *Дерево рішень стосовно вибору моделі та послідовного висновку*

Застосування критерію очікуваної вигоди приводить до оптимального вибору m^*, для якого виконується рівність:

$$\bar{u}\left(m^* \mid \mathbf{x}\right) = \sup_{i \in I} \bar{u}\left(m_i \mid \mathbf{x}\right), \tag{1}$$

де $\bar{u}(m_i \mid \mathbf{x}) = \int u(m_i, a_{\mathbf{x}}^*, \omega) p(\omega \mid \mathbf{x}) d\omega$, очікувана вигода за даних спостережень $\mathbf{x}$ оптимального вибору моделі M_i і рішення $a_{\mathbf{x}}^*$, що забезпечує максимальне значення інтегралу як функції рішення a_j, тобто

$$a_{\mathbf{x}}^* = \left\{ \int u\left(m_i, a_{\mathbf{x}}^*, \omega\right) p_i\left(\omega \mid \mathbf{x}\right) d\omega = \sup_{a_j, j \in J_i} \int u\left(m_i, a_j, \omega\right) p_i\left(\omega \mid \mathbf{x}\right) d\omega \right\}. \tag{2}$$

Прогноз нового спостереження як проблема теорії рішень. Загальна структура теорії рішень стосовно прогнозу майбутнього спостереження представляється даними спостережень **x**, дією m_i, що відповідає вибору моделі M_i для прогнозування, рішенню $a_j, j \in J_i$, що відповідає прогнозу $\widehat{y}_i$ майбутнього спостереження y за моделлю M_i, та критерієм оптимальності прогнозу у вигляді «квадратичних втрат» вигоди прогнозу,

$$u(m_i, \widehat{y}_i, y) = -(\widehat{y}_i - y)^2, \ \ i \in I. \tag{3}$$

З порівняння моделей як проблеми теорії рішень випливає, що оптимальний вибір m^* моделі прогнозування, визначається співвідношенням

$$\bar{u}(m^* \mid \mathbf{x}) = \inf_{i \in I} \int \left(\widehat{y}_i^* - y\right)^2 p_i(y \mid \mathbf{x}) dy, \tag{4}$$

де $\hat{y}_i^*$ — оптимальний прогноз майбутнього спостереження y по моделі M_i, за заданих статистичних даних $\mathbf{x}$, $p_i\left(y \mid \mathbf{x}\right)$ — щільність прогнозного розподілу майбутнього спостереження y, що відповідає моделі M_i і даним спостережень $\mathbf{x}$. Звідси безпосередньо випливає, що

$$\hat{y}_i^* = \int y p_i(y \mid \mathbf{x}) dy = E[y \mid \mathbf{x}, M_i], \tag{5}$$

де $E[y \mid \mathbf{x}, M_i]$ — очікуване значення майбутнього спостереження y, що відповідає моделі M_i і минулим спостереженням $\mathbf{x}$.

Завершення аналізу наразі залежить від визначення повного критерію оптимальності прогнозу

$$\int \left(\hat{y} - y\right)^2 p\left(y \mid \mathbf{x}\right) dy \tag{6}$$

і щільності повного прогнозного розподілу $p\left(y \mid \mathbf{x}\right)$ та обчислення очікуваного розходження

$$E\left[\left(\hat{y}^* - y\right)^2 \mid \mathbf{x}, M^*\right] \tag{7}$$

відносно щільності цього прогнозного розподілу $p\left(y \mid \mathbf{x}\right)$.

Зазначимо, що у випадку, коли $\mathbf{M}$ є повною множиною моделей, тоді відсутні аналітичні методи аналізу оптимального вибору моделі для прогнозування; у цьому випадку просто потрібно здійснити необхідні обчислення, зокрема, провести чисельне інтегрування виразу (5), використовуючи підходящий вигляд виразу для прогнозного розподілу $p\left(y \mid \mathbf{x}\right)$.

У випадку, коли $\mathbf{M}$ є відкритою множиною моделей, проблема порівняння моделей M_i, $i \in I$, як наближень до дійсної моделі, яка апріорі є невизначеною, а тому щільність прогнозний розподіл $p\left(y \mid \mathbf{x}\right)$ не є доступною. У цьому випадку використовується метод Монте-Карло вибору статистичної моделі в умовах стохастичної невизначеності для прогнозування нового спостереження. Детальний аналіз проблеми точкового прогнозу з критерієм квадратичних втрат вигоди, базується на апроксимації пересічного оцінювання, що передбачає розбиття множини даних спостережень $\mathbf{x}$ на дві частини: на базі першої частини спостережень здійснюється ідентифікація моделі для прогнозування, а на базі другої частини — оцінюється точність прогнозу.

Проілюструємо можливий підхід до цієї проблеми детальним розглядом особливого випадку, коли $\omega = y$ — майбутнє спостереження, для якого вимагається отримати точкове передбачення відносно квадратичних втрат або передбачуваний розподіл відносно логарифмічної або квадратичної рахункових функцій.

По-перше, зауважимо, що в усіх випадках, очікувана вигода вибору $M_i, i \in I,$ може бути представлена у вигляді

$$\int u\left(m_i, a_i^*, y\right) p\left(y \mid \mathbf{x}\right) dy = \int f_i\left(y, \mathbf{x}\right) p\left(y \mid \mathbf{x}\right) dy, \tag{8}$$

для деякої функції f_i змінних y та $\mathbf{x}$, залежної від i, вигляд якої може бути явно визначеним. Наприклад, для точкового передбачення за допомогою квадратичної функції втрат вигоди маємо

$$f_i(y, \mathbf{x}) = -\left\{E\left[y \mid M_i, \mathbf{x}\right] - y\right\}^2; \tag{9}$$

для щільності прогнозного розподілу з логарифмічною функцією втрат, ігноруючи членами, що не відносяться до цього випадку, маємо

$$f_i\left(y, \mathbf{x}\right) = \log p\left(y \mid M_i, \mathbf{x}\right); \tag{10}$$

для квадратичної рахункової функції маємо

$$f_i(y, \mathbf{x}) = 2\, p\left(y \mid M_i, \mathbf{x}\right) - \int p^2\left(y \mid M_i, \mathbf{x}\right) dy. \tag{11}$$

По-друге, зауважимо, що для даних спостережень $\mathbf{x} = \mathbf{x}_n = = \left(x_1, ..., x_n\right)$ є n можливих розбиттів у вигляді $\mathbf{x}_n = \left[\mathbf{x}_{n-1}\left(j\right), x_j\right]$, $j = 1, ..., n,$ де $\mathbf{x}_{n-1}\left(j\right) = \mathbf{x}_n - \left\{x_j\right\}$ позначає вибірку $\mathbf{x}_n$ з вилученим x_j, і коли n досить велике та кожне з спостережень x_j замінюється, то таке розбиття ефективно приводить до множини $\mathbf{x}_{n-1}\left(j\right),$ як наближення для множини **x,** та спостереження $x_j,$ як наближення для прогнозного значення $y.$ Якщо наразі випадково вибрати k із цих n розбиттів, то згідно із стандартним законом великих чисел, при $n, k \to \infty,$ маємо

$$\left| \int u\left(m_i, a_i^*, y\right) p\left(y \mid \mathbf{x}\right) dy - \frac{1}{k} \sum_{j=1}^{k} f_i\left(x_j, \mathbf{x}_{n-1}(j)\right) \right| \to 0, \tag{12}$$

тому очікувані вигоди від вибору моделей $M_i, i \in I,$ можна порівняти на базі величин

$$\frac{1}{k} \sum_{j=1}^{k} f_i\left(x_j, \mathbf{x}_{n-1}(j)\right), \quad i \in I. \tag{13}$$

У випадку точкового передбачення, якщо y — майбутнє спостереження та $\widehat{y}_i^*\left(j\right)$ позначає величину $E\left[y \mid M_i, \mathbf{x}\right],$ коли $\mathbf{x}$ замінено на $\mathbf{x}_{n-1}\left(j\right),$ то з цієї апроксимації випливає, що проводиться пошук мінімуму по $i \in I$ значення виразу

$$\frac{1}{k} \sum_{j=1}^{k} \left(\widehat{y}_i^*(j) - x_j\right)^2, \tag{14}$$

який представляє собою середньоквадратичне відхилення точкових прогнозів, обчислених на доступних підмножинах вибіркових даних $\mathbf{x}_{n-1}(\cdot)$, від відповідних їм вибірках розмірності $n-1$.

У випадку вигоди прогнозного логарифмічно-рахункового розподілу проводиться пошук максимального по i значення виразу

$$\frac{1}{k}\sum_{j=1}^{k}\log p\left(x_j \mid M_i, \mathbf{x}_{n-1}(j)\right), \tag{15}$$

який можна назвати усередненою мірою, що базується на логарифмі інтегрованої вірогідності за моделі M_i, і який можна зручно переписати для обчислювальних цілей, у вигляді

$$\log p\left(x \mid M_i\right) - \frac{1}{k}\sum_{j=1}^{k}\log p\left(x_{n-1}(j) \mid M_i\right). \tag{16}$$

У випадку порівняння двох моделей M_1, M_2, цей критерій можна розглядати з цікавим представленням. За вигоди щільності прогнозного логарифмічно-рахункового розподілу, з використанням позначення $p_i\left(y \mid x\right) = p\left(y \mid M_i, x\right)$, критерій порівняння моделей можна представити таким чином, що надання переваги моделі M_1 відповідає умові

$$\int \log \frac{p_1(y,\mathbf{x})}{p_2(y,\mathbf{x})}\, p(y,\mathbf{x})dy > 0, \tag{17}$$

де, однак, для відкритої множини **M**, щільність прогнозної ймовірності $p\left(y \mid \mathbf{x}\right)$ не визначена. Але, як і раніше, можна утворити n розбиттів множини даних спостережень $\mathbf{x}_n = \left[\mathbf{x}_{n-1}\left(j\right), x_j\right], j = 1,\ldots,n,$ таких, що $\mathbf{x}_{n-1}\left(j\right) = \mathbf{x}_n - \left\{x_j\right\}$ для досить великих n є наближенням для **x**, та x_j є наближенням для y. Звідси випливає, що коли випадково вибрати k із цих n розбиттів, то величина

$$\frac{1}{k}\sum_{j=1}^{k}\log \frac{p_1\left(x_j \mid \mathbf{x}_{n-1}(j)\right)}{p_2\left(x_j \mid \mathbf{x}_{n-1}(j)\right)} \tag{18}$$

дає слушну оцінку (при $n \to \infty$) за Монте-Карло виразу для критерію модельного порівняння, поданого вище. Зазвичай, для порівняння моделей часто використовують наступні відношення.

Байєсові фактори. Розглянемо деяку форму інтуїтивної міри попарного порівняння ймовірностей будь-яких двох моделей із множини $\left\{M_i, i \in I\right\}$. Такою мірою порівняння моделей M_i та M_j може бути відношення апостеріорних ймовірностей цих моделей

$$\frac{p\left(M_i \mid x\right)}{p\left(M_j \mid x\right)}=\frac{p\left(x \mid M_i\right)}{p\left(x \mid M_j\right)} \times \frac{P\left(M_i\right)}{P\left(M_j\right)},$$

де, наприклад, $p(x \mid M_i)=\int p_i(x \mid \theta_i) p_i(\theta_i) d\theta_i$.

Визначення байєсового фактора. *За заданих двох гіпотез H_i, H_j, що відповідають альтернативним моделям M_i, M_j, та статистичним даним* **x**, *байєсів фактор стосовно прийняття гіпотези H_i (відхилення H_j) задається відношенням апостеріорних до апріорних ймовірностей цих моделей.*

$$B_{ij}(\mathbf{x})=\frac{p(\mathbf{x} \mid M_i)}{p(\mathbf{x} \mid M_j)}=\left\{\frac{p(M_i \mid \mathbf{x})}{p(M_j \mid \mathbf{x})}\right\} \Big/ \left\{\frac{p(M_i)}{p(M_j)}\right\}.$$

Інтуїтивно, байєсів фактор характеризує міру збільшення або зменшення шансів прийняття гіпотези H_i порівняно з гіпотезою H_j за появи статистичних даних **x.** Отже, $B_{ij}\left(\mathbf{x}\right)>1$ означає, що гіпотеза H_i наразі більш порівняно вірогідна у світлі даних **x**; $B_{ij}\left(\mathbf{x}\right)<1$ означає, що зросла відносна вірогідність гіпотези H_j.

Good у [2] стверджував, що логарифми різних відношень у визначенні байєсового фактору, слід назвати *вагами даних,* тому $\log B_{ij}\left(\mathbf{x}\right)$ відповідає інтегрованій вірогідності ваги статистичних даних на користь (за) M_i (і проти M_j). У цій логарифмічній шкалі, *адитивна* комбінація апріорної ваги даних і байєсового фактору дають апостеріорну вагу даних.

Слід також зауважити, що для моделей $M_i, i \in I$, які відповідають специфікації параметричних розподілів $\left\{p_i\left(\mathbf{x} \mid \theta_i\right), i \in I\right\}$, відношення інтегрованих вірогідностей зводиться до простих відношень вірогідностей.

З урахуванням визначення байєсового фактору, критерій порівняння моделей M_1 і M_2 можна представити у логарифмічному вигляді, за якого надання переваги моделі M_1 відповідає умові

$$\prod_{j=1}^{k}\left[\frac{p_1(x_j \mid \mathbf{x}_{n-1}(j))}{p_2(x_j \mid \mathbf{x}_{n-1}(j))}\right]^{1/k}=\prod_{j=1}^{k}\left[B_{12}(x_j, \mathbf{x}_{n-1}(j))\right]^{1/k}>1, \tag{19}$$

для $j=1,...,k$, де $B_{12}\left(x_j, \mathbf{x}_{n-1}\left(j\right)\right)$ позначає байєсів фактор для надання переваги моделі M_1 в порівнянні з моделлю M_2, який ґрунтується на версіях розподілів

$$\left\{p_i\left(x_j \mid \theta_i\right),\, p_i\left(\theta_i \mid \mathbf{x}_{n-1}(j)\right)\right\} \tag{20}$$

моделей M_1, M_2. Слід зауважити, що у контексті «0–1» — функції втрат і замкнутої множини **M,** роль байєсових факторів базується на версіях щільностей розподілів $\left\{p_i\left(\mathbf{x} \mid \theta_i\right),\, p_i\left(\theta_i\right)\right\}$ моделей M_1, M_2. Хоча тут існує чітка різниця у постановці (**M** — замкнута порівняно з **M** — відкритою множинами, лог-прогнозна функція вигоди порівняно з «0–1» — функцією втрат); цікаво відмітити ту роль, яку знову ж таки відіграє байєсів фактор. Одна цікава відмінність є у посиланні [1], де байєсів фактор оцінює дві моделі M_1, M_2 на базі здатності моделі «передбачити» вектор вибіркових даних **x** без реальних даних (окрім тих, що використовувались для визначення щільності апріорного розподілу $p_i\left(\theta_i\right)$). Навпаки, можна розглядати середнє-геометричне байєсових факторів, які оцінюють моделі M_1, M_2 на базі модельної здатності передбачити одне подальше спостереження за даних $n-1$ спостережень. У першій ситуації робиться наголос на «точності даних спостережень»; у другій — на «силу майбутніх передбачень».

Слід зауважити, що наведена вище апроксимація оптимальної байєсової процедури реалізується природно процесом пересічного оцінювання, результат якого дає преференції для моделей, за яким статистичні дані досягають найвищого рівня «внутрішньої слушності». Так, наприклад, у випадках як квадратичних втрат, так і логарифмічного рахунку, якщо за моделі M_i існує x_j, яке є «несподіваною диковинкою» у світлі $\mathbf{x}_{n-1}(j)$, то це приведе до великих квадратичних розбіжностей або малих величин лог-інтегрованих ймовірностей, відповідно, що, в свою чергу, з точки зору міри якості приведе до штрафу стосовно вибору M_i-моделі.

Процедури вибору та оцінювання моделі, що включають апроксимації пересічного оцінювання, відіграють важливу роль у байєсових постановках для апроксимації очікуваних вигід у проблемах теорії рішень, де порівнюються моделі з множини альтернативних моделей, і не ставиться питання про те, яка одна з моделей цієї множини є «вірною», а також за відсутності специфікації дійсної моделі довіри.

У випадку, коли **M** є замкнутою множиною моделей, щільність повного прогнозного розподілу має вигляд у посиланні [3]:

$$p\left(y \mid \mathbf{x}\right) = \sum_{i \in I} p_i\left(y \mid \mathbf{x}\right) P\left(M_i \mid \mathbf{x}\right), \tag{21}$$

де апостеріорні ймовірності $P(M_i \mid \mathbf{x}),\ i \in I$, згідно з формулою Байєса, мають вигляд

$$P\left(M_i \mid \mathbf{x}\right) = \frac{P\left(M_i\right) p\left(\mathbf{x} \mid M_i\right)}{\sum_{j \in I} P\left(M_j\right) p\left(\mathbf{x} \mid M_j\right)}, \quad p_i\left(y \mid \mathbf{x}\right) = p\left(y \mid M_i, \mathbf{x}\right). \quad (22)$$

Використовуючи вираз повного критерію оптимальності (6) для прогнозу майбутнього спостереження та вираз щільності повного прогнозного розподілу (21), апостеріорний, зважений за усіма моделями M_j, $j \in I$ множини **M** прогноз, має вигляд

$$\widehat{y}^* = \sum_{j \in I} \widehat{y}_j^* p(M_j \mid \mathbf{x}). \quad (23)$$

Модель M_i, $i \in I$ матиме перевагу на множині **M**, коли її прогноз $\widehat{y}_i^*$ буде найближчим до $\widehat{y}^*$. Зокрема, якщо розглянути дві моделі множини **M,** то перевага надається моделі з більшою апостеріорною ймовірністю. Аналогічний результат має місце для наступного простого прикладу:

«0–1» — функція вигоди. Розглянемо проблему вибору із замкнутої множини **M** істинної моделі прогнозування M_t.

У цьому випадку форма функції вигоди має вигляд

$$u(m_i, y) = \begin{cases} 1, \text{ якщо } \; y \in M_i \\ 0, \text{ якщо } \; y \notin M_i \end{cases}. \quad (24)$$

Тоді легко бачити, що згідно з аналізом стосовно вибору моделі та послідовного висновку щодо прогнозу майбутнього спостереження, схематично окресленим на рис. 1, маємо

$$p_i(y \mid \mathbf{x}) = \begin{cases} 1, \quad \text{якщо} \quad y \in M_i, \\ 0, \quad \text{якщо} \quad y \notin M_i \end{cases} \quad (25)$$

та

$$p(y \mid \mathbf{x}) = \begin{cases} P(M_i, \mathbf{x}), \quad \text{якщо} \quad y \in M_i, \\ 0, \qquad\qquad\; \text{якщо} \quad y \notin M_i. \end{cases} \quad (26)$$

Очікувана вигода рішення m_i (вибору моделі M_i) за даних спостережень **x**, набуває вигляду

$$\overline{u}(m_i \mid \mathbf{x}) = \int u(m_i, y) p(y \mid \mathbf{x}) dy = P(M_i, \mathbf{x}),\ i \in I. \quad (27)$$

Звідси випливає таке правило статистичного висновку: для прогнозу майбутнього спостереження слід вибрати ту модель, яка має найбільшу апостеріорну ймовірність.

Список використаних джерел:

1. Bernardo J. M. Bayesian Theory / J. M. Bernardo, A. F. M. Smith. — New York : John Wiley & Sons, 1994. — 555 p.

2. Good I. J. Probability and Weighting of Evidence / I. J. Good. — London ; Griffin ; New York : Hafner Press, 1950. — 340 p.
3. Андрєєв М. В. Лекції з байєсової економетрії. Оптимальні статистичні рішення та системний аналіз проблем прийняття рішень в умовах невизначеності / М. В. Андрєєв. — К. : КІБіТ, 2007. — 464 с.

This paper deals with the problem of a decision making theory application to the prediction of a future observation. The conceptual framework is one of searching for a stochastic model for this problem. The choice of a suitable model is made from open and closed set, respectively. These cases are researched by means of Bayesian approach application.

Key words: *bayesian decision theory, statistical comparison of models, bayes factor, the forecast distribution, Monte Carlo method, the average assessment.*

Отримано: 14.09.2010

УДК 519.23

А. В. Атаманюк, аспірант
Хмельницький національний університет, м. Хмельницький

НОВИЙ ПІДХІД ДО РОЗВ'ЯЗАННЯ ЗАДАЧІ МАРШРУТИЗАЦІЇ

Наведено новий підхід до розв'язання задачі маршрутизації декількох транспортних засобів з часовими вікнами за допомогою алгоритмів допусків.

Ключові слова: *задача маршрутизації, верхній допуск, нижній допуск, «горлишковий» допуск*

Вступ. Під час розгляду задач, що належать до класу NP-складних [1, с. 139—150], таких як задача маршрутизації транспортного засобу з часовими вікнами, важливе місце займає знаходження варіантів розв'язання цих задач за поліноміальний час для певних наборів даних (а не для всіх даних). Найкращим є шлях, який позбавляє від необхідності розгляду повного перебору, виходячи з певних особливостей задачі. Оскільки евристичні правила мають рекомендаційний характер, то евристичні методи не завжди приводять до бажаних результатів розв'язання задачі, тому бажано мати якнайбільший набір евристик. Евристика також використовується для пришвидшення і покращення точності оптимізації алгоритмів в основному через введення ефективних початкових розв'язків.

Постановка проблеми. Аналіз літератури показує, що в даний час задача маршрутизації з часовими вікнами може розв'язуватись за допомогою таких методів: метод гілок та меж (Branch and bound,

