Реоптимізація Узагальнених Задач про Виконуваність з Допомогою Ієрархій Релаксацій як Інтелектуальна Обробка Даних в Системах Штучного Інтелекту

В.О. Михайлюк
кафедра прикладної математики та інформатики
Східноєвропейського національного університету імені Лесі Українки МОНУкраїни
Луцьк,Україна,mikh1709@ukr.net

Reoptimization of Constraint Satisfaction Problems With the Help of the Relaxation Hierarchies as the Data Mining in the Systems of Artificial Intelligence

V. O. Mikhailyuk
Department of Applied Mathematics and Computer Science
Lesya Ukrainka Eastern European National University, Ministry of Education and Science of Ukraine
Lutsk,Ukrainemikh1709@ukr.net

***Анотація*—Обгрунтовано застосування нового підходу для проектування інтелектуального програмного забезпечення для розв'язання узагальнених задач про виконуваність. Інтелектуальнийаналізданихздійснюєтьсянаосновіобгрунтованоїреоптимізаціїоптимальнихрозв'язків**

***Abstract*—The application of a new approach for engineering of intelligent software for constraint satisfaction problems is substantiated. Data Mining is carried out on the basis of reasonable reoptimization of optimal solutions**

Ключові слова—реоптимізація;ієрархії релаксацій; інтелектуальний аналіз даних

Keywords—reoptimization;relaxation hierarchies;Data Mining

I. Вступ

Ціль реоптимізації [1,2], використовуючи наближені методи для задач дискретного програмування, – застосування знання розв'язку початкового екземпляра I задачі для виконання однієї з умов: досягнення кращої якості наближення (відношення апроксимації) I'(змінений екземпляр); створення більш ефективного (по часу) алгоритму для визначення оптимального або близького до нього розв'язку I'; попередні дві умови. Узагальнені задачі про виконуваність (ConstraintSatisfactionProblems, $CSPs$) описують великий клас задач дискретної та комбінаторної оптимізації [1,2].

Означення 1. Узагальнена задача про виконуваність ($MAX\ k-CSP(\mathrm{P})$)може бутизадана скінченими множинами $k \in N, \Sigma$, сімейством предикатів(функцій) P з Σ^k в $\{0,1\}$. Клас $CSP_\Sigma(\mathrm{P})$ складається з усіх екземплярів виду I, де I є підмножиною множини

$$\{f:\Sigma^n \to \{0,1\} : f(x) = P(x_{i_1},...,x_{i_k}) \in \mathrm{P}$$

для $i_1,...,i_k \in [n] = \{1,2,...,n\}\}$.

(Якщо $\Sigma = \{0,1\}$, то підіндекс будемо опускати.)

Значенняекземпляра $I \in CSP_\Sigma(\mathrm{P})$ (позначення val(I)) визначається як $\frac{1}{|I|}\max_{x\in\Sigma^n}\left|\{f \in I : f(x)=1\}\right|$.

Означення 2. Для $k \in N, \Sigma$ скінчених та $\mathrm{P} \subseteq \Sigma^k \to \{0,1\}$ визначимо сімейство предикатів згенерованих P (позначення $<\mathrm{P}>$) як множину

$$\{\mathrm{P}:\Sigma^k \to \{0,1\} : \exists P' \in \mathrm{P}, \sigma_1,..,\sigma_k \in \Sigma : P(x_1,...,x_k) = \\ = P'(x_1+\sigma_1,...,x_k+\sigma_k)\},$$

де $\Sigma = \{0,1,...,|\Sigma|-1\}$ і додавання виконується по модулю $|\Sigma|$.

Нехай $MAX\ k-CSP(\mathrm{P})$ довільна $CSP(\mathrm{P})$ задача. Нехай I є довільний екземпляр задачі, екземпляр I' задачі $MAX\ k-CSP(\mathrm{P})$ отримується з екземпляра I додаванням довільного обмеження $P''\notin<\mathrm{P}>$. Визначимореоптимізаційний варіант задачі $MAX\ k-CSP(\mathrm{P})$.

Задача $Ins-MAX\ k-CSP(\mathrm{P})$**.** **Вхіднідані.**Довільний екземпляр I задачі $MAX\ k-CSP(\mathrm{P})$, x^* єоптимальний розв'язок екземпляра I.

Результат. Знайти оптимальний розв'язок екземпляра I'(отриманого виходячи з I, як описано вище) задачі $MAX\ k-CSP(\mathrm{P})$, використовуючи x^*.

Мета. Знайти x, який максимізує число виконаних обмежень екземпляра I'.

Означення 3.(Цілочислового розриву). Цілочисловий розрив максимізаційної задачі Π визначається як:

$\alpha_\Pi = \sup_{I\in\Pi}\{\frac{OPT_{R_\Pi}(I)}{OPT_\Pi(I)}\}$, де $OPT_\Pi(I)$ є оптимальний розв'язок екземпляра I задачі Π, $OPT_{R_\Pi}(I)$ є оптимальним розв'язком екземпляра I релаксації R_Π задачі Π. Для мінімізаційної задачі Π береться супремум оберненого відношення.

Використовуючи UGC, гіпотезу $P\neq NP$ та ієрархію напіввизначених релаксацій вдалося отримати такий результат.

Теорема 1.[1].Припустимо, що має місце унікальна ігрова гіпотеза (UGC). Нехай $Z-$довільна $MAX\ k-CSP(\mathrm{P})$ задача з цілочисловим розривом $\alpha_Z = \sup_{I\in Z}\{\frac{SDP(I)}{OPT(I)}\}$ і $k=const$, тоді для задачі $Ins-MAX\ k-CSP(\mathrm{P})$ (реоптимізація $MAX\ k-CSP(\mathrm{P})$) існує поліноміальний $\psi(\alpha_Z)$- наближений оптимальний алгоритм, де $\psi(\alpha_Z)=\frac{1}{2-\alpha_Z}$.

II. ІЄРАРХІЇ РЕЛАКСАЦІЙ

Відомо, що обчислення (оцінка) цілочислових розривів опуклих релаксацій (LP або SDP) оптимізаційних задач сама по собі є складною задачею. Згідно схемі загального методу релаксації спочатку задачі формулюються у вигляді цілочислових задач лінійного програмування (IP). IP потім конвертуються в LP (LP-релаксація) або в SDP (SDP-релаксація). Потім здійснюється заокруглення. Якщо, наприклад, розв'язується максимізаційна задача, для якої істинний оптимум є OPT, опукла релаксація дає розв'язок $FRAC$, який не менший ніж OPT (цілочисловий розв'язок є допустимим опуклим розв'язком). Алгоритм заокруглення, використовуючи розв'язок опуклої релаксації зі значенням цільової функції $FRAC$, дає цілочисловий розв'язок зі значенням $ROUND$. Аналіз алгоритму потім зводиться до аналізу цих трьох величин, які задовольняють нерівності $ROUND\le OPT\le FRAC$.

Для максимізаційної задачі цілочисловий розрив визначається як супремум відношення $FRAC/OPT$ (згідно означення 3) по всім екземплярам задачі. Цілочисловий розрив в цьому випадку не менший 1 і «великий» розрив характеризує «погану» якість наближення. Оцінюючи відношення $FRAC/ROUND$, отримаємо оцінку зверху для цілочислового розриву.

Важливим є отримання нижніх оцінок цілочислового розриву. Одним із підходів є застосування ієрархій релаксацій Ловаша-Схрейвера, Шералі-Адамса і Лассерра [1]. Наведемо коротко підхід ієрархії релаксацій Ловаша-Схрейвера . Ловаш і Схрейвер (LovaszandSchrijver) описали метод, відомий як $LS(LS_+)$, посилення (tightening) LP-релаксацій цілочислових 0/1 програм. Метод додає додаткові змінні і допустимі нерівності до LP-релаксацій (причому це можна робити кілька разів послідовно) для отримання послідовності (ієрархії) все більш жорстких релаксацій. Метод, названий Ловаша-Схрейвера ієрархією опуклих релаксацій, дає жорсткі і ефективно розв'язувані релаксації. Для LP-релаксації K позначимо $N(K)$ релаксацію, отриману застосуванням LS методу і $N^k(K)=N^{k-1}(N(K))$ релаксацію, отриману застосуванням k раз ($k>1$) методу LS (або за прийнятою термінологією k раундів або рівнів LS). Якщо початкова лінійна програма K має n змінних і $poly(n)$ обмежень, тоді можливо знайти оптимальний розв'язок $N^k(K)$ за час $n^{O(k)}$, який поліноміальний для $k=const$ і субекспоненціальний для $k=o(n/\log n)$. Ловаш і Схрейвер показали, що $N^k(K)$ забезпечує всі допустимі лінійні нерівності k- кортежів змінних і, зокрема, $N^n(K)$ містить тільки опуклі комбінації допустимих цілочислових розв'язків. Таким чином, опукла програма, отримана після n раундів ієрархії Ловаша-Схрейвера має цілочисловий розрив 1, тобто дає точний розв'язок. Метод LS_+ – більш сильний, оскільки в ієрархії можливі не тільки LP-релаксації, але й SDP- релаксації.

III. ОПИС ЙМОВІРНІСНИХ СИСТЕМ ПЕРЕВІРКИ ДОВЕДЕНЬ

Наведемо формальні означення [1]. Будемо визначати системи доведень з допомогою властивостей перевіряючого. Для перевірки тверджень перевіряючому потрібні помічники, будемо вважати, що він має доступ до одного або більше оракулів. У багатьох варіантах систем доведень обговорюються поняття осіб, що доводять (proovers, прувери) і записаних доведень. Записані доведення еквівалентні доведенням, що використовують оракули, де зчитування i - ого біта відповідає питанню «i ?» до оракулу. Взагалі кажучи, прувери більш могутні, ніж оракули, оскільки можуть бути випадковими й історично залежними.

Означення 4. Оракул є функція $\Sigma^* \to \{0,1\}$ (де Σ^* – множина скінченних рядків в алфавіті Σ).

Типовий перевіряючий $V^{\pi}(x,r)$ є ймовірнісна машина Тюрінга (PTM), де $\pi-$ оракул, $x-$ введення, а $r-$ внутрішня випадкова стрічка. Кажуть, що перевіряючий приймає, якщо він виводить 1 ($V^{\pi}(x,r)=1$), в протилежному випадку він відкидає.

Означення 5. Нехай c і s дійсні числа, такі, що $0 \le s < c \le 1$. Поліноміальна ймовірнісна машина Тюрінга V є перевіряючий в ймовірнісній системі перевірки доведень (PCP-системі) з повнотою c і коректністю s для мови L тоді і тільки тоді, колидля $x \in L$ існує оракул π, такий, що $\Pr_r[V^{\pi}(x,r)=1] \ge c$;для $x \notin L$, для всіх π виконується нерівність $\Pr_r[V^{\pi}(x,r)=1] \le s$.

Означення 6. Перевіряючий V використовує логарифмічну випадковість, якщо існує абсолютна константа c, така, що для кожного входу x і доведення π довжина випадкового рядка r, який використовує V^{π}, оцінюється зверху як $c\log|x|$.

Означення 7. Перевіряючий V читає c біт в PCP-системі, якщо для кожного результату випадкових випробувань і кожного доведення π, перевіряючий V^{π} задає не більше c питань оракулу.

Теорема 2 (PCP-теорема) [1]. Існує універсальне цілєc,таке, що мова в NP маєPCP -перевіряючого V з коректністю $\frac{1}{2}$ та повнотою 1, де V використовує логарифмічну випадковість і читає не більше c біт в доведенні.

Теорема 3 (варіант PCP-теореми) [1]. Нехай $L-$мова в NP і $x-$ рядок. Існує універсальна константа $c<1$, така, що за час поліноміальний відносно $|x|$ можна сконструювати $E3-CNF$ формулу $\varphi_{x,L}$, таку, що якщо $x \in L$, то $\varphi_{x,L}-$виконувана; якщо $x \notin L$, то $\varphi_{x,L}$ не більше, ніж $c-$виконувана. Більше того, кожна змінна зустрічається рівно 5 разів.

IV. ГІПОТЕЗИ ДОСЛІДЖЕНЬ

Основна гіпотеза дослідження – це основна гіпотеза сучасної теоретичної інформатики відносно співвідношення по включенню основних класів складності задач, а саме $P \ne NP$. Відмітимо, що більшість спеціалістів теоретичної інформатики вважають цю гіпотезу істинною. Якщо це не так (тобто $P=NP$), то більшість результатів сучасної теоретичної інформатики (і теорії складності обчислень) втрачають свій зміст.

Ще одна гіпотеза – це так звана унікальна ігрова гіпотеза (UniqueGamesConjecture, UGC) [1]. Вона є деяким посиленням PCPтеореми і була запропонована для доведення результатів з неапроксимованості, де не підходить PCPтеорема. Грубо цю гіпотезу можна сформулювати так. Заданий дводольний граф G, велика константна множина міток $[M]$, перестановка $[M]$, записана на кожному ребрі, розглядається задача знаходження міток вершин (маркування вершин) графа G з $[M]$ так, щоб задовольнялась (виконувалась) кожна перестановка на ребрі. Гіпотеза полягає в тому, що якщо M достатньо велика константа, то являється NP-складним відрізнити екземпляри, де 99% реберних перестановок виконуються, від екземплярів, де 1% реберних перестановок виконуються. Сформулюємо гіпотезу більш точно[1].

Для спрощення роботи UGC часто формулюється з допомогою унікальної задачі покриття мітками (UniqueLabelCoverproblem, ULC) .

Означення 8. Унікальна задача покриття мітками (ULC) $L(V,W,E,[m],\{\sigma_{v,w}\}_{(v,w)\in E})$ визначається так: заданий дводольний граф з лівими вершинами V і правими вершинами W, і множина ребер E. Мета полягає в тому, щоб приписати мітку до кожної вершини графу, де $[m]-$ множина допустимих міток. Маркування повинно виконувати (задовольняти) обмеження, задані взаємно-однозначними відображеннями (бієкціями) $\sigma_{v,w}:[m]\to[m]$. Є одне таке відображення для кожного ребра $(v,w)\in E$. Маркування виконує (задовольняє) ребро (v,w) тоді і тільки тоді, коли $\sigma_{v,w}(label(w))=label(v)$. Оптимальне значення OPT задачі ULC визначається як максимальна доля ребер, що виконується довільним маркуванням.

Гіпотеза 1. Варіант унікальної ігрової гіпотези (UGC, [1]). Для довільних $\eta,\gamma>0$ існує константа $m=m(\eta,\gamma)$, така, що являється NP- складним розрізняти чи ULC задача з розміром m множини міток має оптимум не менший, ніж $1-\eta$ або не більший, ніж γ.

Якщо перейти до розривних версій задач, то гіпотеза стверджує, що $GapULC_{1-\eta,\gamma}$ являється NP- складною.

V. ОСНОВНІ РЕЗУЛЬТАТИ РОБОТИ

Ієрархія Шералі-Адамса визначає ієрархію лінійних програм і є розширенням LS процедури. Ієрархія Лассерра є послідовність «зростаючих посилень» релаксацій напіввизначеного програмування. В даних тезах використовується ієрархія напіввизначеного програмування методу суми квадратів (SumofSquaressemidefiniteprogramminghierarchy) або SOS ієрархія, а також ієрархія Шералі-Адамса [3,4,5]. Завдяки цьому буде справедлива теорема 1 з заміною UGC на загальновідому гіпотезу теорії складності обчислень $P \ne NP$. Будемо говорити [5], що k-місний предикат $P \in <\mathrm{P}>$ підтримує розподіл попарної незалежності μ на $[q]^k$, якщо $\forall j\in[q]\forall i\in[k]\ \Pr_{x\in([q]^k,\mu)}[x_i=j]=1/q$ і для довільних $1\le i\ne j\le k$ та для всіх $a,b\in[q]$ слідує, що

$\Pr_{x\in([q]^k,\mu)}\left[x_i = a, x_j = b\right] = 1/q^2$. Використовуючи результати з [3,4,5] отримаємо.

Теорема4.Припустимо, що $q = 2$ і предикат $P:\{0,1\}^k \to \{0,1\}$ з $<\mathrm{P}>$ підтримує розподіл попарної незалежності μ на $\{0,1\}^2$,Нехай $Z-$довільна MAX $k-CSP(\mathrm{P})$ задача з цілочисловим розривом $\alpha_Z = \sup_{I\in Z}\{\frac{SDP(I)}{OPT(I)}\}$,тоді для задачі $Ins-MAX$ $k-CSP(\mathrm{P})$ існує поліноміальний $\psi(\alpha_Z)$- наближений оптимальний алгоритм, де $\psi(\alpha_Z) = \frac{1}{2-\alpha_Z}$ і $\alpha_Z = \frac{P^{-1}(1)}{q^k}$.

Теорема5.Припустимо, що предикат $P:\{0,1\}^k \to \{0,1\}$ з $<\mathrm{P}>$ підтримує розподіл попарної незалежності μ на $\{0,1\}^q$,Нехай $Z-$довільна MAX $k-CSP(\mathrm{P})$ задача з цілочисловим розривом α_Z,тоді для задачі $Ins-MAX$ $k-CSP(\mathrm{P})$ існує поліноміальний $\psi(\alpha_Z)$- наближений оптимальний алгоритм, де $\psi(\alpha_Z) = \frac{1}{2-\alpha_Z}$ і $\alpha_Z = \frac{P^{-1}(1)}{q^k}$ отримається з допомогою cn рівнів ієрархії напіввизначеної релаксації Шералі-Адамса

Теорема 6. Для довільного $\varepsilon > 0$ і $P:\{\pm 1\}^k \to \{0,1\}$, якщо існує попарно незалежний розподіл μ підтримуваний на $P^{-1}(1)$, тоді існує $\delta > 0$ таке, що для всіх n є множина $I = \{C_1,..., C_m\}$ k-кортежів літералів над $x_1,..., x_n$ така, що

Для довільного $x \in \{\pm 1\}^n$ $E_{C\in I} P(C(x)) \le \frac{\left|P^{-1}(1)\right|}{2^k} + \varepsilon$.

Значення степені δn SOSрелаксації для долі обмежень I, що виконується, рівне 1.

Для задачі $Ins-MAX$ $k-CSP(P)$ існує поліноміальний $\psi(\alpha_Z)$-наближений оптимальний алгоритм, де $\psi(\alpha_Z) = \frac{1}{2-\alpha_Z}$ і $\alpha_Z = \frac{P^{-1}(1)}{2^k}$ отримується з допомогою ієрархії степені δn SOS релаксації.

VI. ІНТЕЛЕКТУАЛЬНА ОБРОБКА ДАНИХ

Для того, щоб провести автоматичний аналіз даних, використовується**DataMining**(здобич (витягання) знань) [6]. Це нова технологія інтелектуального аналізу даних зметоювиявлення прихованих закономірностейувигляді значущих особливостей, кореляцій,тенденцій і шаблонів.Сучасні системи здобичі даних дозволяють знаходитирозчинену в терабайтних сховищах не очевидну, але вельми ціннуінформацію.

Відомо [1], що якщо цілочисловий розрив задачі рівний α_Z, то існує поліноміальний алгоритм з відношенням апроксимації α_Z для її розв'язання. Розглянута серія теорем (теореми 1, 4, 5 та 6) дає змогу в реоптимізаційному варіанті отримати для незначно модифікованих задач поліноміальні алгоритми з відношенням апроксимації $1/(2-\alpha_Z)$, що більше від α_Z (звичайно, при $\alpha_Z \ne 1$). При цьому ці алгоритми є непакращеними та поліноміальними (пороговими, оптимальними) при прийнятті загально відомих гіпотез теорії складності обчислень. Таким чином, даний підхід дає змогу використати DataMining (що в даному випадку представляє собою точний розв'язок задачі, отриманий тим чи іншим методом) для рзв'язання цілого сімейства задач із значно кращою якістю наближення.

ВИСНОВКИ

Як відомо хмарні обчислення виникли як логічна еволюція інформаційних технологій (IT)[6,7]. Рівень DataMining надаєспосіб,якимзнанняпридбане,використовується,знайдено,опубліковано,щобдопомогтикористувачамдосягатисвоїх специфічнихцілей.Тутзнанняподаютьсяякінформація,застосованадлядосягненнямети,вирішенняпроблемиабоприйняттярішення.

ЛІТЕРАТУРА REFERENCES

[1] Михайлюк В.О., Сергієнко І. В. "Постоптимальний аналіз та наближені алгоритми реоптимізації для задач дискретного програмування", Київ: Наукова думка, 2015. 248 с.

[2] BockenhauerH.-J., HromkovicJ., MomkeT., WidmayerP. "Onthehardnessofreoptimization." inSOFSEM, Lecture Notes in Computer Science, V. Geffert, J. Karhumaki, A. Bertoni etc. (eds.), Springer, 2008, vol. no. 4910,pp. 50 – 65.

[3] Ghosh Mrinalkanti, Tulsiani Madhur, "From weak to strong LP gaps for all CSPs."[Online]//Electronic Colloquium on Computational Complexity, 2016. Revision 1 of Report No 117. 29 p. Available:http://eccc.hpi-web.de.

[4] Boaz Barak, Siu On Chan, and Pravesh K. Kothari. "Sum of squares lower bounds from pairwise independence" in Proceedings of the forty-seventh annual ACM Symposium on Theory of Computing (STOC), 2015, pp. 97–106.

[5] Siavosh Bennabas, Konstantinos Georgiou, AvnerMagen, and Madhur Tulsiani. "SDP gaps from pairwise independence", Theory of Computing. 8(12),pp. 269–289, 2012.

[6] Andrei Furda, Colin Fidge, Alistair Barros, Olaf Zimmerman, "Reengineering data-centric information systems for the cloud – a method and architectural patterns promoting multitenancy", in Software Architecture for Big Data and the Cloud, ed. By Ivan Mistrik, Rami Bahsoon, Nour Ali, etc., Cambridge, United States, Elsevier, 2017, CH XIII, Part 3, pp.227 – 251.