

УДК 519.17:512.5

DOI: 10.26661/2413-6549-2019-2-07

## ОСОБЛИВОСТІ ПОШУКУ ОПТИМАЛЬНИХ КЛАСИФІКАЦІЙ: ЕВОЛЮЦІЙНІ АЛГОРИТМИ

I. V. Kozin, Ye. K. Selyutin

Zaporizhzhia National University  
ainc00@gmail.com, wizz92@gmail.com

### Ключові слова:

класифікація, еволюційний алгоритм, генетичний алгоритм, схрещування, мутація, селекція.

Проаналізовано сутність задачі класифікації, а також необхідність виконання умов й відносин на множині. Зазначено властивості відношення еквівалентності. Наведено приклад задачі класифікації, який буде розглядатиметься у даній роботі, а саме: задача покриття графа зірками, яка виникає у багатьох економічних додатках. В якості елементарних фрагментів виступають усі ребра графа. Умови приєднання ребра – це ребро є променем вже існуючої зірки або не має спільних вершин із вже побудованими зірками покриття. Акцентована увага на недостатній оптимальності такого рішення. Була підкреслена актуальність фрагментарної структури задачі. Вказана можливість побудови класів, розглядаючи весь список об'єктів, що класифікуються у певній послідовності. На базі фрагментарної структури запропоновано використовувати еволюційний алгоритм. Були проаналізовані роботи, в яких розглядаються приклади практичного застосування еволюційного (генетичного) алгоритму для вирішення задач класифікації. Оцінена перспектива використання генетичного алгоритму для пошуку оптимальних класифікацій. Вказана поетапна послідовність операцій генетичного алгоритму з прикладами: відбір, схрещування, мутація, селекція. Наведені приклади роботи ключових операторів, а саме кросовера та мутації. Наочно проілюстрований детальний алгоритм еволюційної моделі. Детально описано принцип дії еволюційно-фрагментарного алгоритму. В якості множини допустимих рішень розглядається підмножина максимальних фрагментів на заданій фрагментарній структурі. Визначений механізм перевірки якості генетичного алгоритму на фрагментарній структурі, який зводиться до перебору багатьох варіантів.

## USING OF EVOLUTIONAL ALGORITHMS FOR SEARCHING OPTIMAL CLASSIFICATIONS

I. V. Kozin, Ye. K. Selyutin

Zaporizhzhia National University  
ainc00@gmail.com, wizz92@gmail.com

### Key words:

classification, evolutionary algorithm, genetic algorithm, crossing, mutation, selection.

The article discusses the task of classification, the need to fulfill conditions and relations on the set. The relevance of the fragmentary structure of the task was emphasized. It was illustrated an example of a classification problem that will be considered in this paper, and the task of covering stars with stars that occurs in many economic applications. As elementary fragments all edges appear to the graph. The conditions for joining an edge – this edge is the beam of an already existing star or has no common vertices with already built-up stars of coverage. The emphasis is on the lack of optimality of such a solution. The relevance of the fragmentary structure of the task was emphasized. The possibility of constructing classes is given, considering the entire list of objects classified in a certain sequence. Based on the fragmentary structure, it is proposed to use an evolutionary algorithm. The work analyzed the examples of the practical application of the evolutionary (genetic) algorithm for solving the classification problems. The prospect of using a genetic algorithm for finding optimal classifications is estimated. The following is a step-by-step sequence of operations of the genetic algorithm with examples: selection,

crossing, mutation, selection. Examples of key operators, namely crossover and mutations, are given. The detailed algorithm of the evolutionary model is clearly illustrated. The principle of the evolutionary-fragmentary algorithm is written in detail. As a set of admissible solutions, a subset of maximal fragments on a given fragmentary structure is considered. The mechanism of testing the quality of the genetic algorithm on a fragmentary structure, which is reduced to the enumeration of many variants, is determined.

## 1. Постановка проблеми

Однією з найбільш розповсюджених математичних задач сьогодення є задача класифікації. Вона виникає при аналізі результатів досліджень, при проектуванні й прогнозуванні, при оцінці й прийнятті рішень. У загальному випадку вона є досить складною та суперечливою. Для її вирішення доцільно використовувати метаевристики різних типів [1].

Ефективний виявляється підхід до задачі класифікації на основі теорії фрагментарних структур [2]. Як і будь-яка метаевристика, фрагментарний підхід поступається класичним наближеним алгоритмам й не може гарантувати близькість рішення, що знаходиться з його допомогою, до оптимального. Але простота реалізації й висока швидкість збіжності дозволяє відшукувати за допомогою такого методу гарні наближені рішення багатьох складних оптимізаційних задач [3-5].

У даній роботі розглядається варіант використання комбінації фрагментарного й еволюційного алгоритмів для пошуку оптимальних класифікацій.

## 2. Аналіз останніх досліджень і публікацій

У наш час відома велика кількість різноманітних методів класифікації, ефективність яких суттєво залежить від специфіки предметної області, в якій ця задача сформульована, й особливостей вихідної інформації.

У роботах А. Зенкіна, В. Вапника, Ю. Журавльова розглянуто приклади теоретичного застосування генетичних алгоритмів для вирішення задач класифікації [6-8].

Праці таких вчених, як С. Айвазян, М. Айзерман, Н. Загоруйко, В. Мазуров, Л. Растрингін, присвячені практичному застосуванню генетичних алгоритмів [9-13]. У роботі Д. Голдберга було запропоновано використання оператора схрещування [14].

Ряд підходів до задачі класифікації розглянуто у роботі [15].

Основна ідея пошуку «хорошої» класифікації полягає у введенні критерію оптимальності. У даному випадку задача класифікації зводиться до задачі пошуку оптимального рішення на деякій множині допустимих рішень. Далі починають працювати всілякі методи оптимізації.

## 3. Формулювання ідей

Метою статті є оцінка перспективи використання генетичного алгоритму для пошуку оптимальних класифікацій.

## 4. Виклад основного матеріалу дослідження

Задача класифікації на множині  $X$  – це задача розбиття множини  $X$  на непересічні класи, тобто задача відшукування такого сімейства підмножин  $\{X_\alpha\}$  множини  $X$ , для якого виконуються наступні вимоги:

- 1)  $X_\alpha \subseteq X, X_\alpha \neq \emptyset$ ;
- 2)  $\bigcap_\alpha X_\alpha = X$ ;
- 3)  $X_\alpha \cap X_\beta = \emptyset$  при  $\alpha \neq \beta$ .

Вимога непустоти класів є несуттєвою. Ми вважатимемо однаковими такі класифікації, які відрізняються між собою лише пустими елементами.

Нагадаємо, що відношенням еквівалентності на множині  $X$  називається будь-яке бінарне рефлексивне, симетричне й транзитивне відношення на цій множині, тобто відношення « $\sim$ », яке визначається такими властивостями:

- $$\begin{aligned} \forall x \in X \quad x &\sim x; \\ \forall x, y \in X \quad x &\sim y \Rightarrow y \sim x; \\ \forall x, y, z \in X \quad x &\sim y, y \sim z \Rightarrow x \sim z. \end{aligned}$$

Кожна класифікація породжує природне відношення еквівалентності на множині  $X$ . А саме: два елементи еквівалентні, в тому й тільки тому випадку, коли вони належать одному класу.

Приклад задачі класифікації, який розглядатиметься у даній роботі – це задача покриття графа зірками, яка виникає у багатьох економічних додатках [16].

Нехай заданий граф, ребра якого зважені невід'ємними числами. Задача полягає у знаходженні набору зірок в цьому графі, які містять всі вершини графа й сумарна вага ребер яких мінімальна. Ця задача фактично є задачею класифікації. Класом в цьому випадку є всі вершини зірки. Причому у кожному класі виокремлений представник – центр зірки.

Зірки упорядковуються довільним чином. На черговому кроці алгоритму вибирається перша зірка у послідовності зірок, яка не має спільних вершин з вже знайденими фрагментами.

Ідея фрагментарного алгоритму полягає у наступному: кожне допустиме рішення задачі є фрагментом, який представляється як об'єднання неподільних частин – елементарних фрагментів. На множині фрагментів вводиться операція приєднання, яка дозволяє отримати новий фрагмент шляхом додавання до існуючого одного з елементарних фрагментів. Фрагментарний алгоритм – це алгоритм пошуку максимального по включенню фрагмента.

Більш строго:

а) Визначається множина фрагментів  $\{f\}$  й зовнішня операція  $\oplus$  об'єднання фрагментів. Кожне допустиме рішення задачі складається зі скінченного числа фрагментів, об'єднаних шляхом зовнішньої операції.

б) Задається ряд відношень лінійного порядку  $\{ \}$  на множині фрагментів та ефективна процедура впорядкованих фрагментів, яка дозволяє отримувати впорядкування фрагментів за різними порядками у відповідності з обраними правилами.

в) Задаються умови приєднання фрагмента, які можуть бути як детермінованими, так й динамічними, що змінюються на кожному кроці. Для вже обраної підмножини фрагментів задається ефективна процедура перевірки можливості приєднання фрагмента, що не належить обраній підмножині.

Таким чином, фрагментарний алгоритм дозволяє побудувати допустиме рішення задачі за лінійний за числом фрагментів час.

Покажемо, що задача покриття графа зірками має фрагментарну структуру та, відповідно, будь-яке допустиме рішення задачі може бути отримане шляхом застосування фрагментарного алгоритму.

В якості елементарних фрагментів виступають усі ребра графа. Умови приєднання ребра – це ребро є променем вже існуючої зірки або не має спільних вершин із вже побудованими зірками покриття. Легко показати, що шляхом застосування фрагментарного алгоритму при певному виборі послідовності ребер можна отримати будь-який набір зірок в графі, тобто будь-яке допустиме рішення задачі класифікації. Однак таке рішення не обов'язково є оптимальним. З метою пошуку оптимального рішення опишемо комбінацію еволюційного та фрагментарного алгоритмів.

Еволюційні (генетичні) алгоритми детально розглядалися в чисельних публікаціях [15, 16]. Генетичні алгоритми – це оптимальний метод, що базується на еволюції популяції «особин». У 1975 р. вийшла книга Дж. Холланда «Адаптація в природних та штучних системах», у якій було запропоновано генетичний алгоритм [16].

Для ряду оптимізаційних задач вдалося запропонувати достатньо ефективні процедури пошуку оптимальних рішень, що ґрунтуються на застосуванні еволюційних алгоритмів. Для реалізації еволюційного алгоритму необхідно виділити ряд об'єктів та процедур, сукупність яких називатимемо еволюційною моделлю [15]. Основні складові еволюційної моделі наступні:

– базова множина рішень – множина допустимих рішень  $X$ , на якій відшукується оптимальне рішення задачі;

– оператор побудови початкової популяції: процедура, яка дозволяє виокремити на множині всіх допустимих рішень її підмножину  $Y \subseteq X$  для наступної еволюції;

– критерій селекції – алгоритм, який дозволяє порівнювати за якістю рішення в рамках заданої популяції;

– оператор кросовера  $K: X \times X \rightarrow X$ , що дозволяє за двома допустимими рішеннями-батьками побудувати нове рішення-потомок з множини допустимих рішень;

– оператор мутації  $M: X \rightarrow X$ ;

– оператор відбору, який виокремлює множину пар в  $Y$  для виконання операції кросовера;

– оператор еволюції, що дозволяє будувати нові популяції з множини батьків та потомків;

– правило зупинки, що визначає умову зупинки еволюційного алгоритму.

Опишемо коротко принцип роботи еволюційного алгоритму. На початковому кроці за допомогою оператора початкової популяції будується множина рішень  $Y_0$ . На кожному черговому кроці передбачається заданою деяка множина допустимих рішень – поточна популяція. На першому кроці це множина  $Y = Y_0$ . Для кожного з елементів

множини  $Y$  розраховується значення критерію селекції. Далі за допомогою оператора відбору у поточній популяції  $Y$  обирається множина пар для кросовера. До кожної пари з обраної множини пар застосовується оператор кросовера, а потім до результату кросовера застосовується оператор мутації. Таким шляхом знаходиться множина елементів – потомків  $\tilde{Y}$ .

До проміжної популяції  $Y \cup \tilde{Y}$ , яка є об'єднанням поточної популяції й множини потомків, застосовується оператор еволюції, який виділяє на цій множині нову поточну популяцію. Процес еволюції повторюється до тих пір, поки не буде виконана умова зупинки еволюційного алгоритму. Блок-схема еволюційного алгоритму наведена на рис. 1.

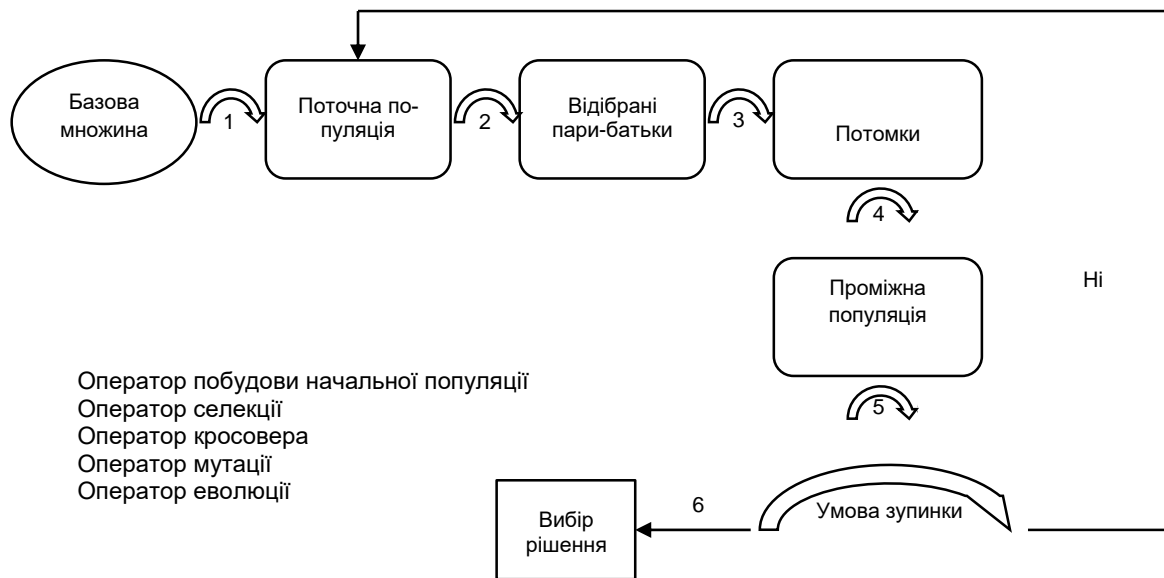


Рис. 1. Еволюційна модель

Властивості фрагментарних структур дозволяють побудувати особливий клас еволюційних алгоритмів на фрагментарних структурах – ЕВФ-алгоритми.

ЕВФ-алгоритм є комбінацією еволюційного й фрагментарного алгоритму. Опишемо еволюційну модель та принцип дії такого алгоритму.

В якості множини допустимих рішень розглядається підмножина максимальних фрагментів на заданій фрагментарній структурі. Кожний фрагмент з цієї підмножини визначається як результат роботи фрагмен-

тарного алгоритму за деякої заданої перестановки елементарних фрагментів. Таким чином, будь-якому допустимому рішенню відповідає визначена перестановка чисел  $1, 2, \dots, N$ , де  $N$  – кількість елементарних фрагментів. Для кожного допустимого рішення визначено значення цільової функції.

Базова множина  $X$  еволюційної моделі – це множина  $S_N = \{i_1, i_2, \dots, i_N\}$  всіх перестановок чисел  $1, 2, \dots, N$ . Оператор побудови початкової популяції виокремлює підмножину заданої потужності  $Q$  з множини  $X$ .

Правило обчислення критерію селекції влаштовано наступним чином: за заданою

перестановкою фрагментів за допомогою фрагментарного алгоритму будується максимальний допустимий фрагмент. Розраховується значення цільової функції задачі для цього фрагмента.

Опишемо оператор кросовера.

Нехай  $U = (u_1, u_2, \dots, u_N)$  і  $V = (v_1, v_2, \dots, v_N)$  – дві довільні перестановки. Перестановка-потомок будується наступним чином: послідовності  $U$  й  $V$  проглядаються з початку. На  $k$ -му кроці обирається найменший з перших елементів послідовності й додається у нову перестановку-потомок. Потім цей елемент видаляється з двох послідовностей-батьків. Наприклад,

$$K((2,3,4,7,8,1,6,5), (3,4,6,2,1,5,8,7)) = \\ = (2,3,4,6,1,5,7,8).$$

Оператор мутації  $M$  виконує випадкову транспозицію (заміну місцями двох елементів) у перестановці.

Оператор селекції обирає випадковим чином набір пар із заданого числа пар у множині перестановок поточної популяції.

Оператор еволюції впорядковує елементи проміжної популяції у послідовність за зменшенням значення критерію селекції. У якості нової поточної популяції обираються перші  $Q$  елементів послідовності.

Звичайне правило зупинки – кількість поколінь досягло граничної межі  $L$ . Найкраща за значенням критерію селекції перестановка з останньої побудованої популяції визначає наближене рішення задачі.

Генетичний алгоритм на фрагментарній структурі є евристичним. Фактично він зводиться до перебору багатьох варіантів. Оцінити якість подібного алгоритму за відсутності точних алгоритмів рішення практично неможливо. Розумно лише порівнювати роботу цього алгоритму з іншими наближеними алгоритмами на великих серіях задач.

Пропонується наступний механізм перевірки якості фрагментарно-генетичного алгоритму на визначеному класі задач.

Випадковим чином будується послідовність задач даного класу.

Для кожної задачі цієї послідовності шукається рішення двома шляхами:

1) впорядкування фрагментів відбувається шляхом застосування генетичних алгоритмів;

2) впорядкування фрагментів відбувається довільним чином.

Кількість розіграшів у другому випадку дорівнює кількості розміру популяції, породженій шляхом застосування генетичного алгоритму у другому випадку.

Якість генетичного алгоритму визначається відношенням числа задач, для яких рішення, отримане за допомогою генетичного алгоритму, краще рішення, отриманого випадковим пошуком, до загального числа задач серії.

Запропонований підхід є універсальним й дозволяє застосовувати один й той самий еволюційний алгоритм до будь-яких оптимізаційних задач на скінченних фрагментарних структурах.

## 5. Висновки

Еволюційний алгоритм відображає синтез алгоритмів класифікацій та оптимізацій функціоналу, що характеризує якість класифікації.

Метод є орієнтованим на обробку багатомірних масивів інформації, особливостями якої є висока розмірність простору ознак та малий об'єм вибірки об'єктів. Пропонований метод класифікації дозволяє не проводити попереднього зниження розмірності простору ознак, що, своєю чергою, дозволяє виключити втрату значимої інформації й врахувати внутрішні зв'язки у інформаційних масивах, що розглядаються.

Результати дозволяють зробити висновок, що підхід до пошуку оптимального рішення ряду дискретних оптимізаційних задач, який ґрунтується на використанні комбінації фрагментарного й генетичного алгоритмів, є перспективним.

Цей підхід особливо зручний у системах підтримки прийняття рішень, коли останнє питання щодо якості того чи іншого рішення вирішується на основі висновку ОПР.

## Література

1. Скобцов Ю. А. Основы эволюционных вычислений: учеб. пособ. Донецк: ДонНТУ, 2008. 326 с.
2. Kozin I. V., Maksyshko N. K., Perepelitsa V. A. Fragmentary Structures in Discrete Optimization Problems. *Cybernetics and Systems Analysis November*. 2017. Vol. 53. P. 931–936.
3. Козин И. В., Батовский С. Е., Сардак В. И. Фрагментарная модель и эволюционный алгоритм 2d упаковки объектов. *Математичне та комп'ютерне моделювання. Серія: Фізико-математичні науки*. 2017. Вип. 15. С. 74–79.
4. Козин И. В., Кривцун Е. В., Пинчук В. П. Эволюционно-фрагментарная модель задачи трассировки. *Кибернетика и системный анализ*. 2015. № 3. С. 35–50.
5. Козин И. В., Полюга С. И. Фрагментарные модели для некоторых экстремальных задач на графах. *Математические машины и системы*. 2014. № 1. С. 143–150.
6. Зенкин А. А., Зенкин А. И. Задача построения оптимальных классификаций. *Сборник работ по математической кибернетике ВЦ АН СССР*. 2010. С. 20–33.
7. Вапник В. Н., Червоненкис А. Я. Теория распознавания образов (статистические проблемы обучения). Москва: Наука, 1974. 313 с.
8. Журавлев Ю. И. Избранные научные труды. Москва: Изд-во Магистр, 1998. 360 с.
9. Айвазян С. А., Бухштабер В. М. Прикладная статистика: Классификация и снижение размерности. Москва: Финансы и статистика, 1989. 276 с.
10. Айзерман М. А., Браверманн Э. М. Метод потенциальных функций в теории обучения машин. Москва: Наука, 1970. 356 с.
11. Загоруйко Н. Г. Прикладные методы анализа данных и знаний. Новосибирск: Изд-во Ин-та математики, 1999. 250 с.
12. Мазуров В. Д. Метод комитетов в задачах оптимизации и классификации. Москва: Наука, 1990. 290 с.
13. Растрингин Л. А., Эренштейн Р. Х. Метод коллективного распознавания. Москва: Энергоиздат, 1981. 325 с.
14. Goldberg D. E. Genetic algorithms in search, optimization, and machine learning. Addison Wesley Longman, 1989. 293 p.
15. Перепелиця В. О., Козін І. В., Терещенко Е. В. Задачі класифікації: підходи, методи, алгоритми. Запоріжжя: Поліграф, 2008. 188 с.
16. Максишко Н. К., Заховалко Т. В. Моделі та методи розв'язання прикладних задач покриття на графах та гіперграфах. Запоріжжя: Полиграф, 2009. 244 с.
17. Holland J. H. Adaptation in natural and artificial systems. Ann Arbor. MI: University of Michigan Press, 1975. 300 p.

## References

1. Skobtsov, Y. (2008). Fundamentals of evolutionary computing: textbook. benefits. Donetsk: DonNTU. (in Russian).
2. Kozin, I. V., Maksyshko, N. K. & Perepelitsa, V. A. (2017). Fragmentary Structures in Discrete Optimization Problems. *Kibernetika i systemniy analiz*, Vol. 53, pp. 931–936.
3. Kozin, I. V., Batovsky, S. E. & Sardak, V. I. (2017). Fragment model and evolutionary algorithm for 2d packing of objects. *Mathematichne ta kompyuterne modelyuvannya. Seriya: Physiko-matematychni nauky*, Vol. 15, pp.74–79. (in Russian).
4. Kozin, I. V., Krivtsun, E. V. & Pinchuk, V. P. (2015). An evolutionary-fragmented model of the trace problem. *Kibernetika i systemniy analiz*, No. 3, pp. 35–50. (in Russian).
5. Kozin, I. V. & Polyuga, S. I. (2014). Fragment models for some extremal problems on graphs. *Mathematicheskkiye mashiny i sistemy*, No. 1, pp. 143–150. (in Russian).
6. Zenkin, A. A. & Zenkin, A. I. (2010). The problem of constructing optimal classifications. *Sbornik rabot po matematicheskoi kibernetike VTS AN USSR*, pp. 20–33. (in Russian).
7. Vapnik, V. N. & Chervonenkis, A. Y. (1974). Theory of pattern recognition (statistical problems of learning). Moscow: Nauka. (in Russian).

8. Zhuravlev, Y. (1998). Selected scientific works. Moscow: Magister Publishing House. (in Russian).
9. Ayvazyan, S. A. & Bukhstaber, V. M. (1989). Applied statistics: Classification and reduction of dimension. Moscow: Finance and Statistics. (in Russian).
10. Aizerman, M. A. & Bravermann, E. M. (1970). The method of potential functions in the theory of machine learning. Moscow: Nauka. (in Russian).
11. Zagoruiko, N. G. (1999). Applied methods of data and knowledge analysis. Novosibirsk: Publishing House of the Institute of Mathematics. (in Russian).
12. Mazurov, V. D. (1990). The method of committees in optimization and classification problems. Moscow: Nauka. (in Russian).
13. Rastrigin, L. A. & Erenstein, R. K. (1981). The method of collective recognition. Moscow: Energoizdat. (in Russian).
14. Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning. Addison Wesley Longman. (in Russian).
15. Perepelitsya, V. O., Kozin, I. V. & Tereshchenko, E. V. (2008). Problems of classification: approaches, methods, algorithms. Zaporizhzhia: Poligraf. (in Ukrainian).
16. Maksishko, N. K. & Zakhovalko, T. V. (2009). Models and methods for developing applied tasks in graphs and hypergraphs. Zaporizhzhia: Polygraph. (in Ukrainian).
17. Holland, J. H. (1975). Adaptation in natural and artificial systems. Ann Arbor. MI: University of Michigan Press.