

# 経験に基づく知識を活用した GA の消込処理問題への適用

奥由輔<sup>†1</sup> 岸田隆視<sup>†2</sup> 佐藤裕二<sup>†1</sup>

消込処理とは、売掛金などが入金された際に明細書と照合し、入金 の遅延や売掛金 が請求額通りであるかを確認する作業を示す。企業間の取引では、請求の合計額と入金額が一致しない場合があるため入金額を基に対応する請求の内訳を求める必要がある。本稿では、遺伝的アルゴリズム(Genetic Algorithm, GA)を用いた解法を提案し、探索能力向上のため企業が持つ顧客の振り込み状況などの、経験に基づく知識を GA の初期値に応用する手法を検討する。

## 1. まえがき

消込処理とは、売掛金などが入金された際に明細書と照合し、入金 の遅延や売掛金 が請求額通りに回収出来ているかを確認する作業を示す。企業間の取引では、振込手数料を抑えるために 100 件を超える請求を纏めて入金することが一般的であり、部署や担当者による取り纏めの時期に差が生じると、請求金額の合計額と入金額が一致しない場合がある。従って、消込処理では、入金額を基に対応する請求金額の内訳を求める必要がある。

現在では、全数探索プログラムや複数の経理関係者による人海戦術で請求金額の内訳を探し出す必要が生じている。この場合、全数探索プログラムでは、請求件数が 50 の場合でも、一般的な PC で求解まで約 1 ヶ月を要する場合があり、人海戦術では期末などの特定の時期のみに、膨大な手作業による人的負荷が集中するなどの問題が生じている。

上記問題は、部分和问题[1]の会計額が大きい特別な例として定義できる。部分和问题は NP 完全問題[2]であり、一般的な部分和问题は動的計画法 (Dynamic Programming, DP) [3]で解けることが示されている。しかし、DP は入金額の大きさに比例した大量のメモリを要するため、企業間の取引で扱うような高額な入金額の場合は適用できない。従って、DP と比べメモリサイズを抑えることができる遺伝的アルゴリズム (Genetic Algorithm, GA) を用いた消込処理問題の解法を提案し[4]、さらに探索能力向上のため企業が持つ顧客の振り込み状況などの、経験に基づく知識を積木[5]と捉えて GA の初期値に応用する手法を検討する。

## 2. 消込処理のための GA プログラムの設計

### 2.1 染色体の定義

消込処理問題を対象とした染色体の定義例を図 1 に示す。各請求項目にそれぞれ異なる遺伝子座を対応させ、遺伝子座の値が 1 である部分の請求に対応した入金がされたと考える。従って、染色体の定義長は請求件数と一致する。図では請求項目 2, 3, 5 に対応する遺伝子座の値が 1 であるため、それぞれ対応した金額 300 円, 250 円, 620 円の金額を合計した 1170 円が本染色体の示す入金額となる。



図 1 染色体の定義例

### 2.2 評価関数

遺伝的操作における評価値は、染色体に対応した請求金額の合計額と入金額の差額の絶対値の逆数を評価値として設定するのが妥当である。しかし、この設定では探索空間の形状が複雑になってしまう場合がある。そこで、0 から 1 の間の値に正規化した式(1)に示す評価関数を定義する。

$$f_e = \frac{f_p}{f_p + f_s} \quad (1)$$

ここで、 $f_p$  は染色体に対応した請求金額の合計額であり、 $f_s$  は  $f_p$  と入金額との差の絶対値を示している。 $f_e = 1$  となった個体が入金額に対応する個体であり、最適解となる。

### 2.3 選択と交差方法

予備実験の結果、ルーレット選択と一様交叉の組合せの時に、請求件数  $N$  に依存せず同一の交叉率、突然変異率で安定した収束性を得られたため、評価実験ではルーレット選択と一様交叉の組合せで実験を行う。

## 3. 経験に基づく知識の応用方法

企業が持つ顧客の振込情報などの経験に基づく知識を利用して、消込処理の件数を絞り込むことを、知識データを用いた処理とここでは定義する。例えば、「A 社からの振込金額は X 円であるが、そのうち B 件からなる Y 円分の振込は定期的に行われる」という場合について考える。この場合、一般に既知である部分を切り分けて「A 件から X 円となる組合せを求める問題」から「(A-B)件から(X-Y)円となる組合せを求める問題」に簡略化することで、探索範囲を絞り込むことができる。一方、この時絞り込みを行った B 件の内に一件でも誤りが有った場合、解なし問題になってしまう。この問題の対策として、知識データにエラーが含まれている場合を考慮した GA の設計を提案する。

GA の初期個体群の一部に、知識データを解の構成要素 (Building Block, BB) として適用する。この場合、積木仮説[5]より探索能力の向上が期待でき、知識データに誤りが有った場合でも、問題無く解探索を行うことが出来る。

請求金額が 3000 円、BB として活用した知識データが

<sup>†1</sup> 法政大学

Hosei University

<sup>†2</sup> (株)日立マネジメントパートナー

Hitachi Management Partner Ltd.

1000 円分であった場合の知識データの適用例を図 2 に示す。GA の初期個体群中の知識データを適用する個体において、1000 円分の知識データに対応した遺伝子を全て 1 とし、それ以外の 2000 円分の遺伝子をランダム生成する。ただし、知識データにエラーが含まれている場合を考慮するため、1000 円分の遺伝子には、正解が 0 である遺伝子座を含める。



図 2 GA の初期個体群への知識データの適用

## 4. 評価実験

### 4.1 評価方法

評価実験の環境を表 1 に、GA の実行パラメータを表 2 に示す。GA の実行パラメータは、選択と交叉方法を決定する際の予備実験を通して決定している。

DP は入金額  $S$  の制限があり企業での消込処理に実用上適用できないため、ここでは全数探索 (Exhaustive Search, ES) と GA の求解までの実行時間を比較する。また、最適な知識データ適用個体数を求めるため、知識データ適用個体数をパラメータに取り、求解までの実行時間を比較する。

ES と GA の比較では、各請求金額の桁数が 8 桁～13 桁までの数値を正規乱数で生成したテストデータを用いて、請求件数  $N$  を 10 から 50 まで変化させて実験を行う。入金件数  $k$  は乱数で設定する。

知識データの最適な適用個体数を求める実験では、各請求金額の桁数が 8 桁～12 桁までの数値を正規乱数で生成したテストデータを用いて、請求件数  $N$  を 30、入金件数  $k$  を 20、知識データ適用遺伝子長を 15、知識データのエラー数を 1 として実験を行う。

### 4.2 実験結果

請求件数  $N$  を変えて ES と GA の求解までの実行時間を比較した結果を表 3 に示す。ES において  $N=30$  までは実測値、 $N=40$  以上は理論上の最悪値を記録し、GA の実行回数は 20 とした。  $N=20$  以下では ES、GA 共に実用上問題ない値である。  $N=30$  を越えると ES は急激に増加するのに対して GA は緩やかな増加を示し、ES と比べて  $N=40$  では約 3,000 倍、  $N=50$  では約 168 万倍高速に最適解を得た。

知識データを適用する個体数を変えて、求解までの実行時間を比較した結果を表 4 に示す。実行回数は 50 とした。平均・最悪実行時間共に知識データは個体数 200 に対し、100 個体適用する場合が最適であり、知識データを適用しない場合と比べ、2 倍程度の処理速度向上を示した。

表 1 評価実験の環境

OS	Windows7
プロセッサ	Intel(R) Core™ i5 CPU 2.50GHz
メモリ(RAM)	6.00 GB
システムの種類	64 ビットオペレーティングシステム
ツール	Visual Studio 2010
プログラム言語	C 言語

表 2 GA の実行パラメータ

個体数	200
遺伝子長	全請求件数 ( $N$ )
交叉率	0.6
突然変異率	0.03
選択方式	ルーレット選択
交叉方式	一様交叉
突然変異方式	単純突然変異

表 3 ES と GA の実行時間

( $N, k$ )	ES	GA (Average)	GA (Worst)
(10, 7)	0.001s	0.001s	0.020s
(20, 14)	1.116s	0.574s	1.860s
(30, 23)	20m43s	1m	4m22s
(40, 17)	350h	2m18s	6m43s
(50, 18)	360,000h	6m58s	12m50s

表 4 知識データ適用個体数

Population of Knowledge Data	Execution Time (sec.)	Worst Execution Time (sec.)
0	35,307	165,063
50	20,582	101,066
100	17,895	78,410
150	19,427	110,212
200	23,684	112,071

## 5. おわりに

本稿では、消込処理問題を部分和问题と定義し、GA を用いた解法について全数探索と GA の速度比較を行い、その有用性を示した。さらに、GA に経験に基づく知識を活用する手法を提案し、速度評価を行った。その結果、誤った請求金額を絞り込んだ場合でも GA では最適解を求めることができ、知識データを適用することで速度向上にも効果があることを示した。

## 参考文献

- 1) 部分和问题  
<http://ja.wikipedia.org/wiki/%E9%83%A8%E5%88%86%E5%92%8C%E5%95%8F%E9%A1%8C> (cited Mar.24, 2013).
- 2) 福田和真, 萩原斉, 中森真理. ナップザック問題の確率アルゴリズムの解析. 数理解析研究所講究録 947 巻, pp. 162-171 (1996).
- 3) 長尾智晴著, "最適化アルゴリズム", 昭晃堂 (2006).
- 4) 武田由美, 奥由輔, 岸田隆視, 佐藤裕二. 消込処理問題への GA の適用とハミング距離を考慮した局所探索能力の向上について. 進化計算シンポジウム 2013.
- 5) D.E.Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wasley Publishing Company (1989).
- 6) Yuji Sato, Yusuke Oku, Masanori Fukuda. Applying GA with Local Search by Taking Hamming Distances into Consideration to Credit Erasure Processing Problems. pp.1183-1190 GECCO-2014.