分散型データベースシステムにおける最適ファイル配置問題に関する研究

埼玉大学大学院政策科学研究科(現在兵庫県企画部企画参事付)　山本　亮三

(修士論文　指導教官　大山達雄　助教授)

1. 研究の目的

現在，地方公共団体のコンピュータシステムの多くは単独のコンピュータに各端末からアクセスする集中型システムを採用している．しかし，情報へのニーズの急速な増加により，今後は分散型データベースシステムへの移行が予想される．分散型データベースシステムでは複数個のファイルをどこに配置するかがシステムの運営上大きな問題となる．本研究は，この分散型データベースシステムにおける最適ファイル配置を，各ファイルへの需要を基礎データとして決定することを目的とする．分析対象は兵庫県自動車税オンラインシステムである．

2. 分析の枠組み

本研究は3モデルからなる．第1は“情報処理需要予測モデル”であり，後につづくモデルの外生データとなる分析対象時点である昭和63年度の情報処理業務の需要予測を行なう．第2は“費用最小化ファイル配置モデル”であり，最適ファイル配置を決定するが，後述のように2つのサブモデルで構成される．第3は“待ち行列シミュレーションモデル”で，最適ファイル配置における各サイトの処理状況を分析する．(図1)

3. 情報処理業務需要予測モデル

本モデルは，1)昭和63年度の需要推計，2)月別変化パターンによる月別需要の推計，3)曜日別変化パターンおよび時間帯特性によるピーク時間帯需要の推計，4)管外比率による各事務所からの各ファイルへの需要推計，という手順で行なったが，詳細な説明は省略する．

4. 費用最小化ファイル配置モデル

17の事務所とファイルについて，システムの費用が最小となる処理機能およびファイルの配置を決定するのが本モデルの目的であるが，同時にこれらを決定しようとすれば，整数変数が膨大となり，適当な時間で解を発見することは困難であり，実用的でない．そこで本モデルを処理機能の配置を決定する「ゾーニングモデル」とファイル配置を決定する「ファイルアロケーションモデル」の2つのサブモデルに分割し，それぞれの最適化を図る．なお，通信回線は公衆回線とし，処理は検索のみとする．

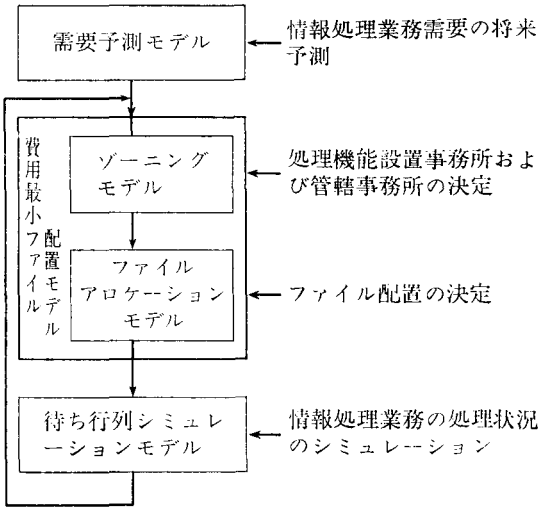


図1　分析手順の概略

4.1 ゾーニングモデル

ゾーニングモデルは，親局（処理機能をもつ）には全ファイルがあるという前提で，以下のように定式化される．なお，$I=\{1,\cdots,17\}$，$J=\{1,\cdots,17\}$ はそれぞれ事務所の集合，親局となる可能性のある事務所の集合を示す．

$$\min.\ \sum_{i\in I}\sum_{j\in J}((PP_j+PT_{ij})\times x_{ij}+\sum_{j\in J}(PF_j\times z_j)$$

決定変数

x_{ij}：i 事務所の需要で j 親局で処理される件数

$z_j=1$：j は親局である．$z_j=0$：j は親局でない．

制約条件

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

1)処理件数に関する制約

$$\sum_{i\in I} x_{ij} \leqq CP_j \times z_j \quad j\in J$$

2)通信回線容量に関する制約

$$x_{ij} \leqq CT_{ij} \times z_j \quad i\in I, j\in J$$

3)親局処理件数下限に関する制約

$$V \leqq \sum_{i\in I} x_{ij} \quad j\in J$$

4)親局数に関する制約

$$1 \leqq \sum_{j\in J} z_j \leqq 17$$

5)その他の制約

a 需要の一致

$$\sum_{j\in J} x_{ij} = T_i \quad i\in I$$

b 親局のみが処理

$$x_{ij} \leqq T_i \times z_j \quad i\in I, j\in J$$

c 整数および非負制約

$$x_{ij} \geqq 0, z_j \in \{0,1\} \quad i\in I, j\in J$$

パラメータ

CP_j：j 親局処理能力

PP_j：j 親局単位処理費用

PT_{ij}：i, j 間単位通信費用

PF_j：j 親局設置単価

CT_{ij}：i, j 間通信回線容量

V：親局処理件数下限

T_i：i 発生需要

計算結果

親局設置単位をパラメトリックに変化させた場合，親局数は3～8局となる．このうち，3～6局におけるゾーニングを示したものが図2である．ところで，目的関数Cは

$$C = \boldsymbol{P}1 \cdot \boldsymbol{Z} + \boldsymbol{P}2 \cdot \boldsymbol{X} + Q$$

$\boldsymbol{P}1$：設置単価ベクトル　$\boldsymbol{P}2$：通信料金ベクトル

$\boldsymbol{Z}$：親局設置ベクトル　$\boldsymbol{X}$：通信量ベクトル

Q：総処理費用

と示される．ところで，制約条件には費用にかかわるものはないので，$\boldsymbol{P}1$ と $\boldsymbol{P}2$ が同率で変化した場合はゾーニングに変化は生じない．したがって，各設置単価におけるゾーニングを調べることにより，図3のように，いずれの組合せにおけるゾーニングを容易に求めることができる．

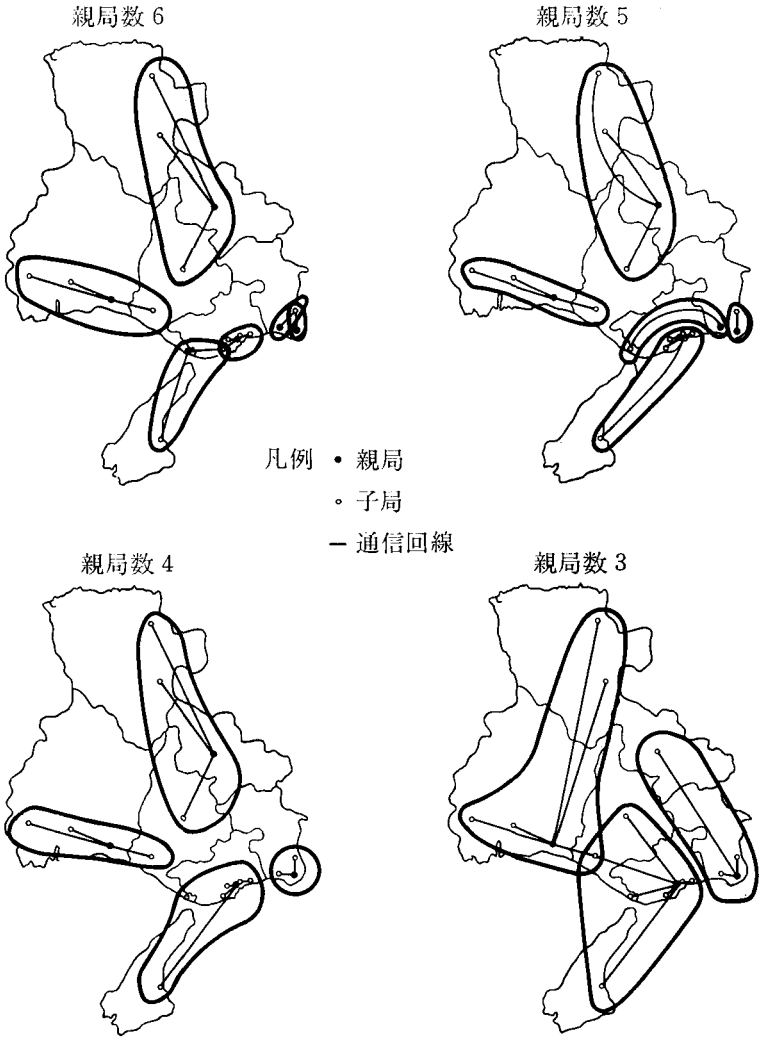


図 2　各親局数におけるゾーニング

4.2 ファイルアロケーションモデル

本モデルは，ゾーニングモデルで得られた5親局の場合について，17のファイルを各ゾーンごとに統合し5ファイルとした場合のファイル配置の最適化を行なう．また，システムの安全性はファイル冗長性により表現している．なお，$I=\{1,\cdots,5\}$，$J=\{1,\cdots,5\}$，$K=\{1,\cdots,5\}$ はそれぞれ需要発生ゾーン，処理ゾーン，ファイルの集合を表わす．

$$\text{min.} \sum_{i\in I}\sum_{j\in J}\sum_{k\in K}(P_{ij}\times x_{ijk}) + \sum_{j\in J}\sum_{k\in K}(SC_k \times y_{jk})$$

決定変数

x_{ijk}：j ゾーンにあるファイル k への i ゾーンからの需要

$y_{jk}=1$：ファイル k は j ゾーンにある．

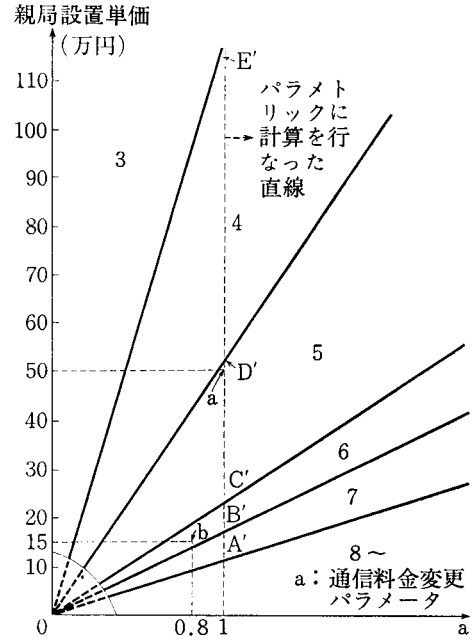
$\quad=0$：ファイル k は j ゾーンにない．

制約条件

(1)需要と処理件数の一致

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

図 3 設置単価と親局数（グラフの中の数字は親局数．8～は8以上を示す



$$\sum_{j\in J} x_{ijk}=D_{ik} \quad i\in I, k\in K$$

(2)ゾーン別親局処理件数に関する制約

$$\sum_{i\in I}\sum_{k\in K} x_{ijk}+A\times\sum_{i\in I(\neq j\in J)}\sum_{k\in K} x_{ijk}\leqq CP_j \quad j\in J$$

(3)ゾーン別ファイル別需要件数に関する制約

$$\underline{M}\times y_{jk}\leqq\sum_{i\in I} x_{ijk}\leqq\bar{M}\times y_{jk} \quad j\in J, k\in K$$

(4)ファイル別コピー数に関する制約

$$\sum_{j\in J} y_{jk}\geqq L \quad k\in K$$

(5)ゾーン別記憶量に関する制約

$$\sum_{k\in K} N_k\times y_{jk}\leqq CS_j \quad j\in J$$

(6)アクセシビリティに関する制約

$$x_{ijk}\leqq F\times y_{jk} \quad i\in I, j\in J, k\in K$$

(7)その他の制約

$$y_{jk}\in\{0,1\},\ x_{ijk}\geqq 0 \quad i\in I, j\in J, k\in K$$

パラメータ

P_{ij}：i, j 間単位通信・処理費用

A：通信処理の検索件数換算パラメータ

L：最低コピー設置数(≧2)

N_k：k ファイル必要容量

F：十分大きな数

SC_k：k ファイル記憶費用

CP_j：j ゾーン処理上限

$\underline{M}, \bar{M}$：各ファイルへのアクセス件数下上限

CS_j：j ゾーン記憶容量

D_{ij}：i から j ファイルへの需要件数

計算結果

図4が得られたファイル配置である．通信，記憶費用を変化させて計算を行なったが，配置に変化は生じない．これは，記憶費用のウエイトが通信費用に比して大きいことと，各ファイルにコピーを2つ設置することによるより近隣のゾーンへのアクセスが可能になるためである．

また，(2)，(3)，(5)の制約なしに変動費用である転送処理費用と通信費用を最小化した場

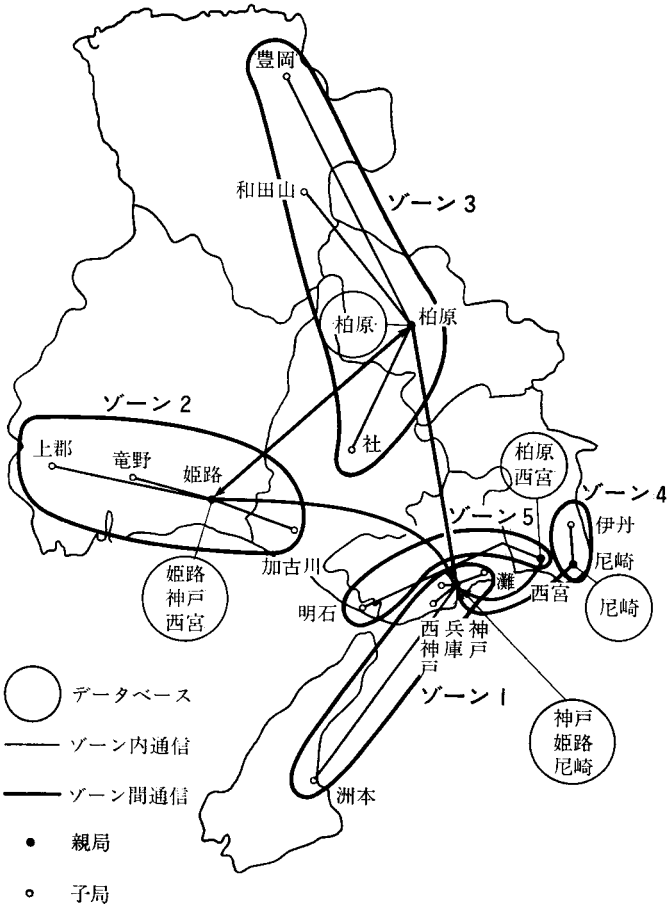


図 4 最適ゾーニングおよびファイルアロケーション

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

合の目的関数値は上記モデルの目的関数値の近傍となっている．

ところで，ファイルをゾーンごとに統合しない場合の制約がないケースにおいても，統合したケースと同じ配置が得られており，統合により解を効率的に求めていると考えられる．

5. 待ち行列シミュレーションモデル

本オンラインシステムの処理形態にはループ処理や二重処理もあり，待ち行列理論の厳密な適用はできない．そこで，上のモデルで得られた各パラメータを使い，GPSSによってシミュレーションを行なった．最悪のケースである到着間隔，処理時間の分布が指数分布であるとした場合，シミュレーション結果はそれぞれのゾーンについてのM/M/1モデルで十分近似しうる(表1)．この場合，需要の変動による処理状況の変化や任意の待ち時間を越える確率を容易に把握することができ，処理状況の評価を行なうについての判断材料を示すことができる．

6. まとめ

上記のモデルにより，効率的に分散型データベースシステムにおけるファイル配置について実用的な近似解を求めることができる．しかし，通信回線およびコンピュータの性能，価格については強い仮定のもとに各パラメータを設定しており，実際の適用においては詳細なデータを使用する必要がある．また，通信回線についても，専用回線の使用を含めた定式化が必要であろう．

表1 シミュレーション結果とM/M/1モデルとの比較

	1ゾーン		2ゾーン	
	SIMULT'N	M/M/1 近似	SIMULT'N	M/M/1 近似
平均処理時間	3.155	3.144	2.803	2.817
平均発生間隔	3.818	3.817	3.504	3.509
平均待ち時間	16.211	14.713	12.132	11.469
平均待ち長さ	4.246	3.855	3.462	3.269
待ち無の比率	17.04	17.610	19.57	19.718
待ち無を除く待ち時間	19.539	17.857	15.084	14.286

	3ゾーン		4ゾーン	
	SIMULT'N	M/M/1 近似	SIMULT'N	M/M/1 近似
平均処理時間	2.829	2.881	2.339	2.347
平均発生間隔	9.483	9.434	9.054	9.174
平均待ち時間	1.464	1.268	1.009	0.807
平均待ち長さ	0.154	0.134	0.111	0.088
待ち無の比率	68.84	69.452	72.86	74.413
待ち無を除く待ち時間	4.699	4.149	3.719	3.155

	5ゾーン	
	SIMULT'N	M/M/1 近似
平均処理時間	2.698	2.770
平均発生間隔	6.587	6.579
平均待ち時間	2.098	2.015
平均待ち長さ	0.319	0.306
待ち無の比率	58.47	57.895
待ち無を除く待ち時間	5.503	4.785

シミュレーション
実施時間
108000秒
(30時間)

単位時間
1/100秒

分布
発生間隔，処理時間とも指数分布

× × ×

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．