クラスタリングにより選定したサンプリングポイントの 需要データを用いた商品マーケットシェア推定の一方法

正 員 小 野 俊 之 (日立製作所)

非会員 中 田 英 樹 (日立製作所)

正 員 小 坂 満 隆 (日立製作所)

A Method of Estimating Product Market Share
Using Demand Data of Sampling Point Selected by Clustering
Toshiyuki Ono, Member (Hitachi, Ltd.), Hideki Nakata, Non-member (Hitachi, Ltd.),
Michitaka Kosaka, Member (Hitachi, Ltd.)

Demand forecasting is a critical success factor for effective supply chain systems. Many demand forecasting methods based on time-series data have been investigated. However these methods can not consider the on-going market change because of its usage of past time series. In this paper, a new method of estimating the product market share based on real time demand sensing data is proposed. This method consists of the two parts, which are (1) sampling point selection using clustering analysis and (2) estimating product market share using sampled demand data. Numerical experiments show the effectiveness of the proposed method and recent network technology makes it easy to implement this method in real demand forecasting systems.

キーワード：クラスタリング分析、需要予測、マーケットシェア推定、サプライチェーン

1. はじめに

消費者ニーズの多様化や企業競争のグローバル化により、企業は消費者ニーズに応じた商品を消費者が欲しい時により安く提供することが求められている。このため、サプライチェーン⁽¹⁾が重要な企業戦略となりつつある。サプライチェーンとは、一企業内での効率化だけではなく、消費者に商品を提供するために必要な調達・生産・物流・販売といった業務の流れを担当する各企業間で戦略的に提携して、全体の効率化を目指そうとするものである。具体的には、消費者の需要に合わせて、供給者・工場・物流拠点・販売店・消費者に至る商品の流れをコントロールし、各企業で消費者サービスを向上しつつ品切れや在庫過多による無駄を低減することである。

サプライチェーンを管理するために最も基本的なことは、生産や物流計画作成の基礎として用いられる商品の需要を正確かつ迅速に把握することである⁽²⁾。従来か

ら需要を把握するために、数多くの定量的な需要予測方法が開発されてきた。例えば、ARIMA⁽³⁾などを用いた時系列予測や、需要に関連する諸要因を関係式として用いる説明変数予測⁽⁴⁾がある。しかし、これらの方法は、過去の時系列データから将来のトレンドを予測するものや、需要に関連する要因のデータから関係式により予測するものであり、現実にマーケットで起こっている状況を直接反映したものではない。

一方、最近の情報通信技術の進展や、POSシステム等のセンサ手段の充実を背景にして、リアルタイムで需要をセンシングし把握することができるようになってきた。たとえば、小売店のPOSや自動販売機の売上データがリアルタイムに売上状況を集計できる手段として活用されている⁽⁵⁾。我々は、リアルタイムな売上のセンシングデータを活用することにより、より現実に近く正確な需要の把握ができなかと考えた。

本論文では、サプライチェーンの効率化のために必要

な需要の把握の一手段として、代表的な店舗や自動販売機のPOSデータをリアルタイムに収集し、これらの売上データから、商品のマーケットシェアを高精度に予測する方法を提案する。マーケットシェアを正確に把握するには、全ての販売ポイントでの競合する全商品に関するの売上データが必要となるが、全ての販売ポイントからのPOSデータを利用できるわけではない。そこで、本論文で提案するクラスタリングにより選定したサンプリングポイントの需要データからマーケットシェアを推定する考え方が有効になる。まず、2章で基本的な考え方とこれを実現するシステム構成を示し、本システムにおけるマーケットシェア問題の定式化を行う。次に、3章でこの問題を解決する基本的な考え方として、(1)全体市場を構成する販売店を似通った販売特性を持つものに分類し、各グループに属する店の中で各グループを代表する店をサンプリングポイントとして選定し(2)選定したサンプリングポイントの店から商品売上情報を収集して、他の店の商品売上情報および全体マーケットシェアを予測する方法を提案する。さらに4章で、数値実験により提案方式の有効性を確認し考察を加える。

2. 提案システムとマーケットシェア予測問題

2.1 マーケットシェア推定システムの概要

本論文で対象とする商品の流通状況を以下のように仮定する。複数のメーカーが、工場で生産した商品を物流センタから各地域の小売店へ配送する。商品は、全て小売店や自動販売機を通しての販売である。対象分野の商品に対して数社が競合している。市場は成熟期にあり、市場の全体需要に大きな変動はなく、各社は商品のシェア確保が重要課題となっている。

小売店や自動販売機での商品販売情報は、幾つかの店についてはネットワーク経由でPOSデータとしてリアルタイムに入手可能である。しかし、コストの関係上、入手できる店舗の数は限られている。また、過去の一定期間の全小売店での各社商品の売上や、シェア情報はデータベースに格納されており必要に応じて利用できる。このような状況のもとで、図1に示すようなマーケットシェア推定システムを構成する。本システムは、過去一定期間の全店舗の商品別売上データをもとに、サンプリングポイントを決める。サンプリングポイントから収集した、リアルタイムの売上データをもとに、各社の商品のマーケットシェアを推定し、サプライチェーン管理に反映する。

2.2 マーケットシェア予測問題の定式化

上記システムにおけるマーケットシェア予測問題を定式化する。本問題は、いくつかの店舗におけるリアルタイムの商品別需要と過去の全店舗の商品別売上情報が与えられたもとで、一定期間の市場全体における商品別シェアを予測することにある。推定すべき (a) 商品別の

図 1. マーケットシェア推定システムの構成
Fig1. Structure of Market Share Estimating System

マーケットシェアと (b) 商品別全体需要、所与となる情報である (c) 過去の商品別店舗別売上情報と (d) サンプリング店の商品別需要の定義を次に述べる。

(a) 商品別のマーケットシェア

シェア推定対象となる期間 $[t, t + p]$ における、対象商品分野での商品 i のマーケットシェア $S(i)$ は、次の式 (1) で表される。

$$S(i) = \frac{D(i)}{\sum_{i=1}^I D(i)} \quad (i = 1, \dots, I) \dots (1)$$

但し、 $D(i)$: 期間 $[t, t + p]$ ($t, p > 0$) に

おける、商品 i の市場全体需要

I : 競合対象の全商品数

(b) 商品別全体需要

期間 $[t, t + p]$ における、商品 i の全体需要 $D(i)$ は、次の式 (2) で表される。

$$D(i) = \sum_{k=1}^K D_k(i) \quad (k = 1, \dots, K) \dots (2)$$

但し、 $D_k(i)$: 期間 $[t, t + p]$ における、店舗 k

での商品 i の需要

K : 全店舗数

(c) 過去の商品別店舗別売上情報

$\overline{D_k(i)}$: 過去の一定期間 $[s, s + p]$ における、店舗 k

での商品 i の需要
但し、 $0 < s < t$

(d) サンプリング店の商品別需要

$D_k(j)$: 期間 $[t, t + p]$ における、店舗 k での商品 j の需要
但し、 $j \in J$ 、 J は、全ての商品 k のリアルタイム需要が把握可能な店舗の集合
また、 $O(J) < I$ 、 $O(J) : J$ に属する要素数

2. 3 提案方法のアプローチ

式 (1) (2) からわかるように、本問題の商品別マーケットシェアを予測するためには、店舗別商品別の需要 $D_k(i)$ が推定できれば良い。しかし、すべての店舗から売上に関するデータを収集することは経済的な制約をはじめ種々の制約で難しい。そこで、サンプルポイント $j \in J$ を選び、そのポイントでの売上データ $D_k(j)$ と過去の商品別店舗別売上 $\overline{D_k(i)}$ とから、データ収集できない店舗の商品別の需要 $D_k(i)$ を推定する。ここで、店舗別商品別の需要 $D_k(i)$ は、次の式 (3) で表すことができる。

$$D_k(i) = S_k(i) \times \sum_{j=1}^J D_k(j) \quad \dots\dots (3)$$

但し、 $S_k(i)$: 期間 $[t, t + p]$ における、
店舗 k での商品 i のシェア

市場の全体需要に大きな変動がないことから、期間 $[t, t + p]$ における店舗 k での全体需要と、期間 $[s, s + p]$ における店舗 k での全体需要とは、近似的に等しいと仮定することができ、次式 (4) が成り立つ。

$$\sum_{i=1}^I D_k(i) = \sum_{i=1}^I \overline{D_k(i)} \quad \dots\dots (4)$$

また、店舗での商品シェア $S_k(i)$ については、店舗の過去の売上データ分析から、同じ様な売れ行き傾向を持つ店の需要傾向は、同じ傾向を持つ他店の需要傾向と似た傾向を示すことを利用して、類似の特性を持つ店をグループ化し、この中でリアルタイムにデータ入手できる店のシェア情報から同一グループ他店のマーケットシェアを推定する。

以上のアプローチを実現するにあたり、(1) どの店舗をサンプリングポイントの店舗とするか、および (2) サンプリングポイントで得られたデータからどのように他店のシェアを推定するか、という2点を具体的に解決

する必要がある。次章にこれらの点に関する提案方法について述べる。

3. マーケットシェア予測方法の提案

提案方式では、サンプリングポイントとしてどの店舗から需要に関する情報を収集するかを決め、次に決定したサンプリングポイントからの需要情報をもとに、全体のマーケットシェアを推定する。以上の考え方にに基づき、サンプリングポイントの決定とサンプリングデータからの全体シェアの推定アルゴリズムを述べる。

3. 1 サンプリングポイントの決定方法

サンプリングポイントの決定基準としては、なるべく全体を代表する店舗をサンプリングポイントとすることである。しかし、全店舗は同質の市場をもつとは限らず、むしろ通常は幾つかの異質な市場で構成されると考えられる。そこで、図2に示すように、店舗を、同じ特性を持つ幾つかのグループに分け (ステップ1)、そのグループの中で、各グループを代表する店舗を一つ選ぶ (ステップ2) こととする。各グループを代表する店舗として、そのグループに属する店舗特性の平均的なものにもっとも近い店舗を選ぶ。

ここでグルーピングに利用する店舗特性とは、当該店舗の商品の売れ行きに関係する市場の特徴を示すデータであり、具体例としては各店舗での各商品の売り上げやシェアのほか、間接的には各店舗の対象市場の人口構成などが考えられる。ここでは、店舗特性として、各店舗の過去の商品売上データを利用する。本論文で対象とする商品の特性は、各店舗の購買層はそれほど変化はなく、過去のデータでグルーピングしてよいと考えられる。次にサンプリングポイントの決定手順を以下に示す。

<ステップ1> 同じ特性を持つ店舗にグルーピング

まず、同じ特性を持つ店舗にグルーピングする。グルーピ

図2. サンプリングポイントの決定
Fig2. Decision of Sampling Points

ングの手法としては、多変量解析の非階層的クラスター分析を用いる。非階層的クラスター分析は、複数の個体を個体間の類似度により、グルーピングする手法である。個体間の類似度を計測する代表的な値としては、ユークリッド距離がある。m個のデータユニットをk個のグループに分類するアルゴリズムは、以下の段階から成り立っている⁽⁶⁾。

- (a) データ集合中のk個のデータユニットをそれぞれグループのメンバとして割り当てる。
- (b) 残りのm-k個のデータユニットを、最近隣距離の重心をもつグループに割り当てる。
- (c) (b)で全てのデータユニットがグループに割り当てられたのち、グループの重心を再計算する。
- (d) 全データユニットを最近隣距離の重心をもつグループに割り当てなおす。
- (e) グループの割り当て直しがなくなるまで(c)(d)を繰り返す。

＜ステップ2＞グループのサンプリングポイント決定
クラスタリングの結果、すべての店舗が分類され、いくつかのグループができる。なおグループの重心とは、グループに属する全店舗を店舗特性の軸にプロットした時の平均座標である。式(5)が最小となる店舗を、各グループの重心から、最も距離の近い当該グループの代表的な店舗と見做して、サンプリングポイントとする。

$$\sum_{i=1}^p (X_i - Y_i)^2 \dots\dots (5)$$

但し、i：店舗特性のインデックス番号

p：店舗特性の総数

X_i：グループ内店舗の店舗特性iに関する座標値

Y_i：グループの重心の店舗特性iに関する座標値

3. 2 サンプリングデータからの全体シェア推定方法

サンプリングポイントから収集した商品別売上情報をもとに、他の店舗のシェアを推定し、全体のマーケットシェアを推定する。全体マーケットシェアを推定する方法として、いくつかの方法が考えられるが本論文では代表的な次の2つについて検討する。

(1) グループ内同一シェア推定

同一グループ内の店舗の特性は、近似的には同一と仮定し。式(6)に示すように同一グループ内の店舗のシェアは当該サンプルポイントのシェアと同一とする。

$$S_k(i) = S_l(i) \dots\dots (6)$$

但し、kとlとは、同じグループに属する店舗

(2) 店舗特性別加重シェア推定

図3に示すように、全サンプルポイントと、シェア推

図3. サンプリングデータからの全体シェアの推定

Fig 3. Market Share Estimating from Sampling Data

定対象の店舗との特性の近さにより、サンプルポイントのシェアを加重平均し、店舗のシェアを推定する。推定する店舗とサンプリング店舗との距離を求める。求めた距離の逆数を、加重平均の重みとする。例えば、サンプリング店舗でのある商品iのシェアがS_k(i)であったとすると、推定対象店舗のシェアS_l(i)は、次式(7)で計算される。

$$S_l(i) = \sum_{j=1}^p \frac{S_k(j)}{X_j} / \sum_{j=1}^p \frac{1}{X_j} \dots\dots (7)$$

但し、S_k(j)：代表店舗の需要

X_j：推定する店舗と代表店舗とのクラスター内での距離

p：サンプリングした店舗数

以上の処理をサンプルポイント以外の全店舗における商品に対して行い、それぞれの店舗でのシェアを推定する。推定された各店舗のシェアを数式(3)へ代入し、数式(1)(2)を用いることで、全体市場における商品別マーケットシェアは計算できる。

4. 提案方法の評価

提案方式の有効性を評価するため、2つの異なる商品のシェア予測に適用実験した。

4. 1 実験方法

(1) 前提条件

予測対象としたのは、商品群Aの北海道地域における一定期間のシェアと、別の商品群Bの東京地域におけるシェアの2つである。商品群Aは8社が販売しており、販売店は北海道地域に218店舗ある。商品群Bは15

社が販売しており、販売店は東京地域に64店舗ある。
クラスタリングは、過去の自他社商品の売上データから各店舗の商品別シェアが全店舗別にわかるため、これを店舗の特性を集約しているとみなして、各商品のシェアをグルーピング変数として用い、初期値重心は全商品の売上の多い店順とした。また、サンプリングポイント数は、コスト面からの最大許容値として商品群Aの北海道地域では15、商品群Bの東京地域では20とした。

(2) 比較方式

提案方式と比較するための方式として、サンプリングポイントの選定に関しては、提案方式と同じ数の15、20をランダムに選定する方式を採用した。全体シェアの予測に関しては、縦軸に推定する現在のシェア、横軸に過去の商品シェアをとり、サンプリングポイントから収集したデータをプロットし、図4に示す2つの推定方式を採用した。一つは、プロットした点に対し3次、4次、5次、6次の多項式関数⁽⁷⁾でモデリングし、情報量基準AIC⁽³⁾が最小となる関数の外挿、内挿で商品シェアを推定する方式で、もう一つは、単純に折れ線でモデ

(1) 折れ線式でのシェア推定

(2) 多項式でのシェア推定

図4. 比較対象となるシェア推定方式

Fig4. Market Share Estimating by the Modeling Method

実験番号	サンプリング	予測方式
実験1	提案方式	提案方式：同一推定
実験2	提案方式	提案方式：加重推定
実験3	提案方式	折れ線予測
実験4	提案方式	多項式予測
実験5	ランダム選定	折れ線予測
実験6	ランダム選定	多項式予測

表5. 比較対象となる実験パターン

Table5. Experiment Pattern

リングして外挿、内挿で商品シェアを推定する方式である。提案方式と比較方式とを組み合わせ、表5に示す6パターンで実験し、結果を比較評価することとした。

(3) 評価指標

評価指標としては、実績値との比較が重要になるため、次式(8)で表される、各社商品のシェア推定値と実績値の乖離平均を合計したものとした。

$$\sum_{i \in A(or, B)} \frac{\sum_{k=1}^K |S_k(i) - R_k(i)|}{K} (\%) \dots \dots (8)$$

但し、 $S_k(i)$: 商品 i の店舗 k での推定シェア
 $R_k(i)$: 商品 i の店舗 k の実績シェア
 K : 店舗数
 A : 商品群Aに属する各社商品の集合
 B : 商品群Bに属する各社商品の集合

4. 2 結果と考察

(1) 結果

実験結果について、図6に商品群別の乖離平均を示す。商品群別の乖離平均を見ると、実験1「サンプリング：提案方式、予測：提案方式、同一推定」が、商品群A、Bそれぞれ34.88%、9.36%と、最も乖離平均が小さい結果となった。提案方式を採用していない実験6と比べると、乖離平均値は、それぞれ約4.2%、5.3%減少した。また、もう一方の提案方式である実験2「サンプリング：提案方式、予測：提案方式、加重推定」についても、他の実験と比較して、乖離平均は減少した。

サンプリング方式のみ提案方式を取り入れた実験3、4については、商品群Bの実験4以外は、いずれも、ランダム選定方式の実験5、6に比べ、それぞれ乖離平均が低減した。

(2) 考察

上記の条件下で、提案するマーケットシェア推定方法の有効性は概ね確認できた。このことは、適切なサンプリングポイントが選定されていれば、有意なシェアの値が推定できることを意味している。

また、実験3、4は、商品群Bの実験4を除いて、実験5、6よりも乖離平均が小さくなっているため、サン

図6. 実験別の全商品の乖離平均
Fig6. The Mean of the Difference

フリリングポイントの選定だけでも予測精度の向上に有効と考える。

5. おわりに

POSデータのようにリアルタイムで入手できる商品売上情報を、サンプリングした代表小売店や自動販売機のPOSから収集し、これを用いて全体の商品売上情報を予測する方法を提案した。本方法は、(1)全体市場を構成する地域市場を、似通った販売特性を持つもの同士でグルーピングし、各グループに属する地域市場の中で各グループを代表する地域市場をサンプリングポイントとして選定し、(2)選定したサンプリングポイントの地域市場から、商品売上情報を収集して、さらに収集した売上情報から、他の地域市場の商品売上情報を推定することが特徴である。さらに、実際の売上データを用いた実験で、提案方式による予測精度を検証し実用に十分供することを確認した。

本提案方式は、シェア推定対象となる商品、店舗が多くなればなるほどその効果が大きくなる。従って、情報ネットワーク社会が進展し、リアルタイムな売上データが利用できる状況になりつつある今日、製造流通業における戦略情報システムの一環として、これを商品のシェア予測に適用すれば、サプライチェーン管理における生産計画や物流計画の効率化を図れるものと考えられる。

参考文献

- (1) 阿保栄司:「ロジスティクスマネジメント」P167-185:同友館(1994)
- (2) P.Kotler: "Marketing Management, Aanalysis, Planning, Implementation, and Control, 9thed.", P130-141: Prentice-Hall (1983)
- (3) S.Makridakis, S.C.Wheelwright:「計画策定と意思決定のための予測手法入門」, P107-138, P159-180:同友館(1995)
- (4) 勝村正鷹、他:「地域販売力評価システムの作成と量産販売施策策定への適用」, Journal of Operations Research Society of Japan, Vol.28, No.2, P134-161(1985)
- (5) 矢作敏行、他:「生・販統合マーケティングシステム」, P23-52:白桃書房(1993)
- (6) M.R.Anderberg:「クラスター分析とその応用」, P201-211:内田老鶴園(1988)
- (7) Norman R.Draper:「応用回帰分析」, P133-134:森北出版(1968)

小野 俊之 (正員) 1969年生。91年3月九州大学理学部数学科卒業。同年(株)日立製作所システム開発研究所に入所。流通物流・金融分野のビジネス情報システムの研究開発に従事。情報処理学会、日本オペレーションズ・リサーチ学会会員。

中田 英樹 (非会員) 1961年生。80年3月長野工業高校電気科卒。同年(株)日立製作所システム開発研究所に入所。製造・流通分野のビジネス情報システムの研究開発に従事。97年6月同社東北支社に異動。現在公共分野を中心に情報システムの拡販に従事。

小坂 満隆 (正員) 1952年生。77年京都大学工学部数理工学修士課程修了。同年(株)日立製作所システム開発研究所に入所。ビジネス情報システムの研究開発に従事。計測自動制御学会などの会員。(工学博士)