

## II-23 LTPP 大規模公開データのための入力プログラム生成システム

上島 壯

Tsuayoshi Kamijima

【抄録】米国の舗装に関する大規模プロジェクト LTPP の公開データ DataPave2.0 はデータテーブル 415 種、フィールド数延べ 10500、レコード総数約 987 万件で、マイクロソフト ACCESS データベースにより提供されている。このようなデータソースからエクスポートされたデータ群の SAS 言語入力プログラムセットを生成する汎用システムを開発した。これを応用して事象年月日データのみを用いて LTPP の試験区間約 2300 ヶ所の測定内容や測定頻度を解析し、公開データの全体像を明らかにするとともに、その結果をデータベース化した。

【キーワード】LTPP, DataPave, SAS, program generator, 舗装マネジメント

## 1 LTPP プロジェクトと配布データ

LTPP(Long Term Pavement Performance)プロジェクトは、1987 より 5 ヶ年計画で行なわれた米国の研究プロジェクト SHRP(Strategic Highway Research Program)の 1 研究分野として、1988 から 20 年間の予定で進行中の、舗装の長期供用性に関する大規模研究プロジェクトである (SHRP 終了後は FHWA に引き継がれている)。これは米国・カナダの 2 千箇所以上の舗装試験区間を、GPS(一般舗装研究)10 件、SPS(特定舗装研究)10 件の研究テーマに割当てている。

1998 年初頭に CD-ROM1 枚に収められた LTPP データがユーザーインターフェースソフトを付属させて DataPave97 として配布が開始され、1999 年に第 2 版の DataPave2.0 が CD-ROM2 枚組で発行された。

DataPave2.0 のデータ本体は MS-ACCESS 形式の 3 ファイルに収められており、次のようなデータ出力プログラムが付属している。

操作手順は、Windows のメニュー画面(図 1)より、州、研究課題、気候区分、降水量、交通量、舗装厚、などの条件を指定する。次にそれらの条件に適合してリストアップされた試験区間から任意の区間群を選択し、解析に利用可能なデータテーブル群を選択して、最後に MS-EXCEL または CSV 形式で外部出力して解析作業に入る。もともになるテーブルは 415 種あり、一般的項目、気象、交通、舗装試験、材料、維持、修繕、研究課題固有の属性、などに分類され

ている。フィールド数は延べ 10500、レコード総数約 987 万件である。

## 2 SAS プログラム生成汎用システムの開発

米国国外の研究者には、「土地勘」がないというハンディを補うためには、研究区間を選択する前にデータの全体像をできるだけ明らかにしておくことが必要と思われるが DataPave のデータベースファイル全てに目を通すためのプログラム作成は手作業では困難である。そこでデータから SAS 用プログラムを生成するジェネレータ GENSAS2 を開発した。なお、付属のデータ選択インターフェースを利用したにしても解析作業手順はユーザー自身が工夫しなければならない。以下、GENSAS2 の概要を示す。

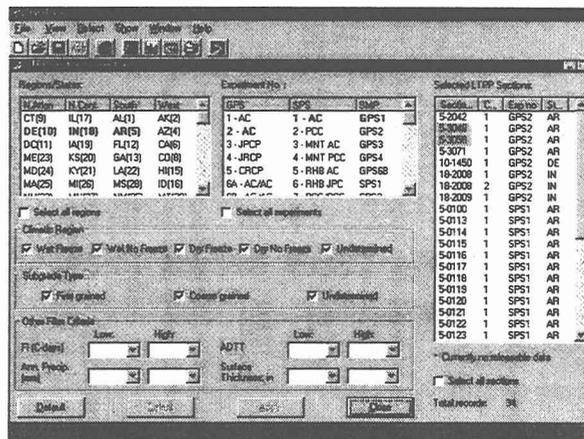


図 1 DataPave 添付のインターフェース画面

```

<ヘッダー部>
%LET SSD = WORK;
%LET DIR1 = C:\LTPPDATA\WDPY;
%LET DIR2 = C:\LTPPDATA\WMSY;
%MACRO LINES1; length _DSN_ $8; _DSN_ = "&DSNAME"; %MEND;
%MACRO LINES2; length _DSN_ $8; _DSN_ = "&DSNAME"; %MEND;
%MACRO STEPS1; proc append base=base1 force; %MEND;
%MACRO STEPS2; proc append base=base1 force; %MEND;
%LET STAT=YES;

<個別プログラム部>
=====
title '(2) DP_02';
filename in "&DIR1.DP_02.txt";
%LET DSNAME = DP_02;
=====
data &SSD. &DSNAME;
length S108 $4 C118 $1;
LENGTH T38 $18;
Length S227 $2 D188 $4;
infile in firstobs=2 missover delimiter='09' x DSD Irec1=91
input S108 S227 C118 T38 D188 P136 N88_1 N88_2;
%LINES1;
label
  S108='SHRP_ID'
  S227='STATE_CODE'
  C118='CONSTRUCTION_NO'
  T38='TEST_DATE'
  D188='DROP_HOUR_MINUTE'
  P136='POINT_LOC'
  N88_1='N_Def1_1'
  N88_2='N_Def1_7'
;
run;
%STEPS1;
%MACRO DO_STAT; %IF &STAT=YES %THEN %DO;
proc means data=&SSD. &DSNAME maxdec=3;
proc freq data=&SSD. &DSNAME;
tables C118;
run;
%MEND; %MEND; %DO_STAT;
run; title;

```

図 2 生成 SAS プログラム例

- a) Windows の DOS プロンプトで用いる。
- b) ACCESS、EXCEL などからエクスポート出力されたコマンドまたはタブ区切りデータで第 1 行が項目の説明行になっているものを対象とする。
- c) 対象データはファイル名リストで指示し、フィールド数の判定、各フィールドのデータ型の判定、略語変数名の設定を行い、SAS 入力プログラム部を作成する。SAS のデータ型は文字型と数値型の 2 種類であるが、通常数値型として扱われる yy/mm/dd hh:mm などの日付時刻、頭部にゼロが付加される整数は文字型に設定し、表記方法で区別した。
- d) SAS はすべての名称が 8 文字までなので変数名を略語化しなければならない。GENSAS2 は生成プログラムと共にデータ解析結果を出力する。これと補助プログラムを用いて統一的な略語変数辞書を作成できる。

GENSAS2 の出力例を図 2 に示す。データ入力部のみであるから、この後の段階から通常のプログラム作成となる。キー変数を用いてテーブル間の結合を行ない解析用データセットを作成することが一般的な作業である。

EXPERIM	CLM 02	CLM 05	INV 17	TST 03	RIB 17
DP 04	CLM 06	CLM 07	INV 18	TST 04	RIB 18
DP 05	CLM 08	CLM 09	INV 20	TST 05	RIB 19
DP 02	CLM 10	CLM 11	INV 21	TST 06	RIB 20
DP 01				TST 07	RIB 21
DP 03				TST 08	RIB 22
				TST 09	RIB 23
MON 03			INV 07		
MON 05			INV 09		
MON 06			INV 10		
MON 07			INV 23		
MON 08					
MON 20					
MON 22			INV 08	TST 07	
MON 24			INV 03	TST 09	
MON 25				TST 10	
MON 26				TST 11	
MON 27				TST 12	
MON 28				TST 13	
MON 04				TST 14	
MON 21				TST 15	
				TST 16	
MON 17				TST 17	
MON 24				TST 18	
MON 09				TST 19	
MON 14				TST 20	
				TST 21	
				TST 22	
				TST 23	
				TST 24	
				TST 25	
				TST 26	
				TST 27	
				TST 28	
TRF 03					
TRF 04					

図 3 文書データベースによる試験区情報の視覚的表示

### 3 特定解析プログラムの作成と出力結果の利用

DataPave テーブルのフィールド数は 3 から 256 ままで分布している。基本的な共通変数は 4 桁の番号と州番号を結合した試験区名である。これと、事象の発生年月日を表すフィールドデータを取り出すコードを付加したプログラムにより、テーブル名、試験区コード、事象年月日の 3 変数からなるデータセットを作成し、試験区コードを持つすべてのテーブルデータの集合を作成した。この内データが収録されていないテーブルが 38 ある一方、路体の状態を評価する FWD 落錘試験レコードは 1884 区間の 485 万件が収録されている。年月日フィールドの選択は準備作業によりテーブルごとに設定されている。

図 3 は試験区間のデータの蓄積状況を把握するために作成したデータの所在を示すマップのある試験区の例である。テーブル名を略語化し、90 年代のデータの取得年を 1 桁の数字で表示している。この試験区では FWD 試験 (MON\_03 など) や路面平坦性の測定 (MON\_22 など) が毎年行なわれ、修繕 (RHB) が 1995 年に行なわたことが示されている。このデータ所在マップはパソコンデータベース「知子の情報」の文書データとして作成したものである。

### 4 あとがき

入力プログラム生成システムの開発により、非常に困難と考えていた LTPP 全データの読込みが可能になり、地点情報、日付情報など最も基本的なデータのみで LTPP データの輪郭を明らかにできた。