GROUPER: An Expert System For Redistricting

Stephen P. Leach and Abraham Kandel

Department of Computer Science
Florida State University
Tallahassee, Florida 32306

Abstract

The process of "redistricting" involves the division of a land surface into two or more pieces. In a political setting, the districts thus formed serve as groups of voters that elect the same public officials. Other types of redistricting problems include the formation of school board districts, water management districts, or transportation districts.

This paper provides a brief overview of the redistricting problem and then describes a PC-based expert system currently being developed that will assist in the process.

1. Introduction

The goal of the redistricting process is to take a single area and divide it into two or more pieces such that these pieces meet some established criteria. When forming voting districts, the criteria is equality of population. For school board districts, one would be concerned only with the distribution of children between the ages of 5 and 18. Likewise, in forming water management districts or transportation districts the main concern would be the distribution of rainfall, surface water, or highway miles. Although this paper specifically addresses the formation of voting districts, the discussion applies equally well to other redistricting problems.

As stated earlier, the primary goal in formation of voting districts will ordinarily be equality of population. In most cases, it is also mandatory that a district be one "contiguous" land mass. Although there are a number of other very desirable features that might be sought in a redistricting process (such as compactness of shape, integrity of municipal boundaries, or the retention of minority voting strength), these other features and their relative importance tend to be user dependent [1].

In nearly all redistricting situations the population figures used are based on those from the U. S. Bureau of Census. The Census, which takes place every 10 years, reports on a large variety of demographic information, including total population and population breakdown by racial categories. This data is made available in machine readable form.

For the purpose of reporting demographic information, the census divides each state into progressively smaller units. Although there are a number of variations and inconsistencies from one state to the next, the following situation is common:

1) The state is divided into counties.

2) Each of these counties is divided into county census divisions (CCD's).

3) Each CCD is divided into tracts.

4) Each tract is divided into block groups (BG's).

5) Each BG is divided into blocks.

Population figures are reported by state, as well as for each county, CCD, tract, BG, and block. It is significant that units of one type are nested within the next larger unit.

The table in Figure 1 contains some information that will help establish the approximate number and relative size of the various Census units. These numbers are derived partially from 1980 data for Florida [2] and will vary from state to state, but the relative size and complexity of the situation is hopefully made clear.

The precision with which one may form districts using the Census units at the block level, as well as the complexity of working with such large sets of data, presents an interesting challenge for a PC-based expert system.

2. The GROUPER Approach

As can be seen from the previous discussion, the sheer volume of units to choose from presents a combinatorial task that is formidable using any type of computer system. The GROUPER approach to this problem attempts to form districts by restricting itself to units at some given size (e.g., by counties), thereby limiting the volume of data dramatically. Thus, the problem (at least temporarily) becomes "are there combinations of whole counties that, when combined together, form an acceptable district?"

Unit Type	# of Units	Avg. Unit Population
State	1	10,000,000
County	67	150,000
CCD	300	35,000
Tract	2,000	5,000
BG	10,000	1,000
Block	200,000	50

Figure 1

One significant problem with this GROUPER approach becomes immediately obvious: a number of larger counties have populations that are much bigger than the "ideal population" of a single district. The solution to this dilemma is to allow combinations of Census units whose total population is close to that of a multiple of the ideal district size. By doing so, a "Group" or "Grouping" can be treated as a separate redistricting problem that could be further refined by working at a finer mesh of Census unit.

The results of this "divide and conquer" approach to the process of redistricting are pleasing. Once a Group of counties has been formed (assuming that the Group is not already of a single district size), one can go down to the next level of Census data without a significant increase in complexity of the redistricting problem. The chart in Figure 2 is intended to demonstrate this concept by showing a hypothetical division of Florida's 67 counties into 120 House districts. Note that groups do not have to be of equal size. The last column in the chart simply indicates the average size of such groups.

3. An Expert System For Grouping

The expert system GROUPER is a generalized system that will locate groups from a given data file of units. If the number of districts to form is not included in the data file, that piece of information is obtained from the user of the system. Although the units within such a data file would ordinarily be of the same Census "mesh" (e.g., counties or CCD's), there is no requirement that this be the case. Data items that reside on this file are the unit name, a unique unit identifier, the unit population, a list of unit identifiers of other units that are contiguous to this unit, and one or more (X,Y) coordinates that are used when graphically displaying a potential grouping.

Grouping Level	# Of Units To Work With At That Level	Number of Groups Formed	Districts Per Group
County	67	5	24
CCD	60	4	6
Tract	100	3	2
BG	300	2	1

Explanation of chart:

If 5 Groupings of counties are found, then each will consist of about 24 districts (120/5). When considering one of these Groups at the CCD level, you would only be working with about 60 CCD's (300/5). If 4 Groupings of these CCD's are found, each will consist of about 6 districts (24/4), and so on. This technique would rarely require looking at Census data at the lowest level (block level).

Figure 2

GROUPER currently consists of 65 rules written using the M.1 [3] expert system shell. A number of useful data file operations are performed by external function calls to routines written in C. The graphical representation of potential groupings is handled through a Turbo Pascal routine that is accessed through one of the external C functions [4].

The actual technique used in locating possible groupings is similar to the exhaustive search technique used to solve the well-known "knapsack" problem [5]. GROUPER solves the problem nonrecursively through the M.1 rule

```
if positiveinteger=LOOP and
   do(set still_looking=yes) and
   grouping_finished
then end
```

which causes the search to continue iteratively until the grouping problem is solved to the user's satisfaction.

Although pure exhaustive search could be used in finding a redistricting solution, the time requirements would usually be prohibitive. To speed up the process, GROUPER uses several rules to short-circuit the exhaustive search. The rule

```
if still_looking and
   hopeless and
   do(set still_looking = no) and
   abort_search and
   1>2
then grouping_finished
```

is one such instance where the GROUPER abandons the current search path and backtracks to look for a more promising one. In this case the criteria for "hopeless" is as follows: no unconsidered units are contiguous to those that are already in the grouping being formed.

A number of other reasons will make a combination of units unacceptable. These are handled (along with the acceptable ones) by rules of the following kind

```
if still_looking and
   choice_acceptable and
   choice_saved and
   major_cleanup and
   1>2
then grouping_finished.
```

Examples of why a group would not be "choice_acceptable" include

1) Population is too small or too big,

2) Grouping is not contiguous,

3) Remaining unused units are not contiguous,

4) Grouping does not satisfy other properties, defined by the user, and

5) Grouping does not have user's final approval.

These other criteria are further defined by additional M.1 rules.

Figures 3 and 4 show a small portion of an M.1 session using GROUPER. Figure 5 shows the results of a complete grouping at the county level in Florida. The table in Figure 6 summarizes the characteristics of this grouping. While there will be reasons that one or more of these groups are not completely satisfactory to certain individuals, preliminary indications are that these particular groupings and others derived through GROUPER have merit.

```
How many districts would you like to form?
>> 120
How close must groupings be to multiples of the ideal district size?

   Give value between 1-100, where 1 means "extremely close"
   and 100 means "closeness of little significance as long
   as within legal bounds".

>> 5

A group of size 7 having a deviation of -35 per distict has been found.
Units in this group are as follows:

escambia
santa_rosa
okaloosa
walton
holmes
washington
bay
calhoun
gulf
```

Figure 3

```
Do you approve of this grouping?
>> why
    M.1 is trying to determine whether the following rule is
    applicable in this consultation:

        kb-45:
            if next_choice = [escambia,santa_rosa,okaloosa,walton,holmes,
               washington,bay,calhoun,gulf] and
               group_size = 7 and
               size_diff = -35 and
               display([nl,'A group of size ',7,' having a deviation ','of ',
               -35,' per distict has been found.',nl,
               'Units in this group are as follows:',nl,nl]) and
               user_shown([escambia,santa_rosa,okaloosa,walton,holmes,
               washington,bay,calhoun,gulf]) and
               picture_shown([escambia,santa_rosa,okaloosa,walton,holmes,
               washington,bay,calhoun,gulf]) is sought and
               user_approves
            then user_acceptable.

    The following entries are also under consideration:

        kb-34   (a rule)
```

Figure 4

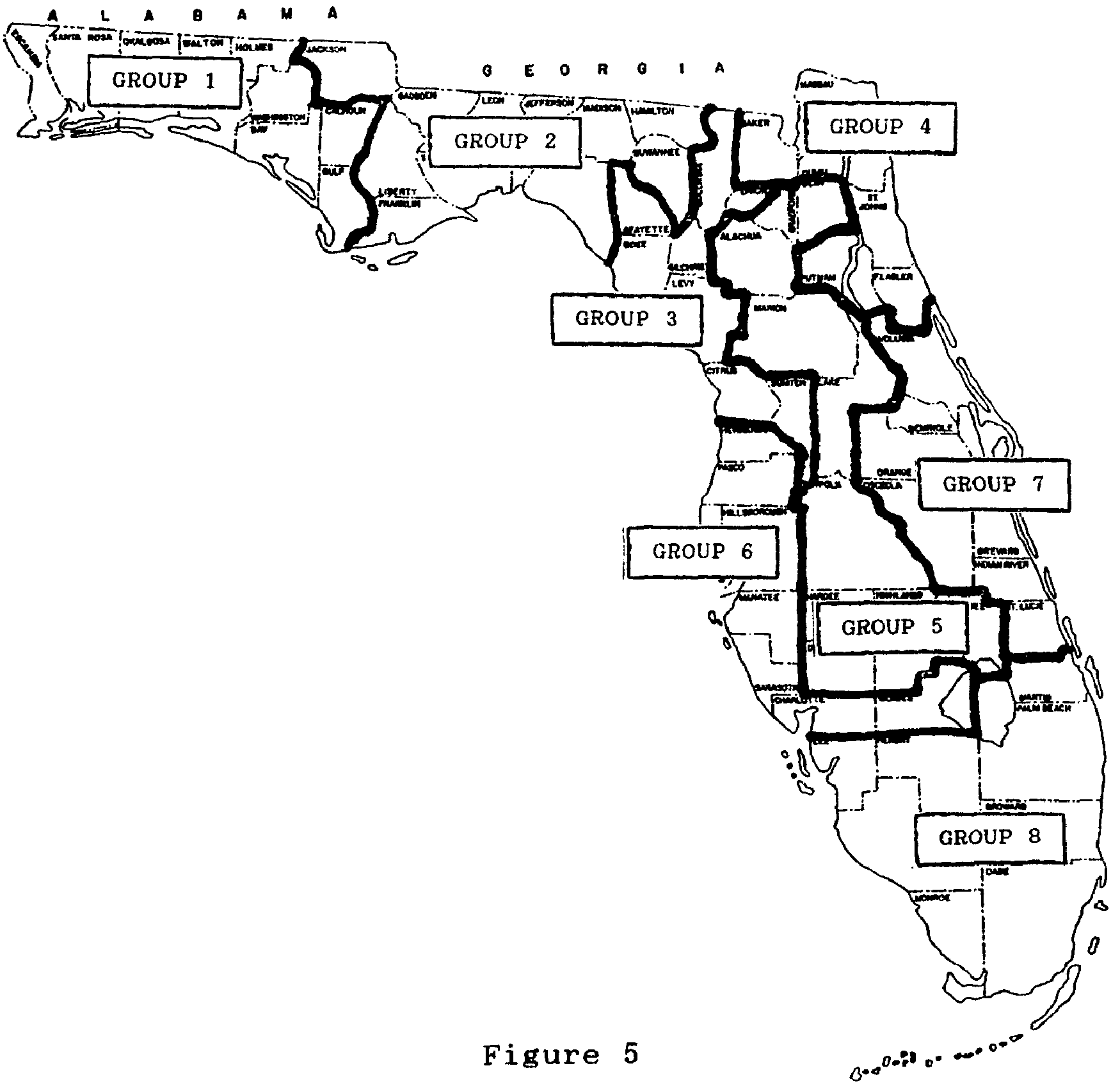

Figure 5

Group	Number of Districts in Group	Deviation From Ideal (%)	(Persons/District)
1	7	0.04	35
2	4	0.21	173
3	2	0.23	186
4	9	0.20	158
5	11	0.09	73
6	25	0.03	23
7	17	0.02	20
8	45	0.07	58

Figure 6

4. Conclusion

GROUPER provides a generalized system for the formation of districts of various types. As an expert system written in M.1, GROUPER rules may be individually tailored to provide the various criteria that a specific user may feel appropriate for satisfactory districts. Preliminary results indicate that groups formed using GROUPER compare quite favorably with those formed using the traditional process (1980) in Florida.

5. References

[1] Leach, S., "A Computerized Analysis of Disenfranchisement from SJR 1E," Florida House Select Committee on Reapportionment, May 1982.

[2] "County Maps and Revised 1980 Census Data," Florida House Select Committee on Reapportionment, March 1982.

[3] M.1 Reference Manual, TEKNOWLEDGE, Inc., 1986.

[4] Turbo Pascal Reference Manual, Borland International, Inc., 1986.

[5] Aho, A., Hopcroft, J., and Ullman, J., Data Structures and Algorithms, Addison-Wesley, 1986, pp. 66-69.