

Phenomenal Data Mining and Link Analysis.

Donal Lyons,
School of Systems and Data Studies,
University of Dublin, Trinity College,
Dublin, Ireland.
Donal.Lyons@tcd.ie

Gregory S. Tseytin,
Intelligent Systems Laboratory,
V. I. Smirnov Research Institute,
St. Petersburg State University, Russia.
tseytin@tseytin.spb.ru

Abstract

Targeted marketing is an increasing trend in advertising with companies attempting to send heavily customised mail shots to only that subset of customers identified as likely to be interested. This requires detailed demographic data on the target group and, when the customer is identified by virtue of having paid by credit card or ordered by mail, such data may be readily available. When customers are anonymous as, for example, at a supermarket checkout, Phenomenal Data Mining has been proposed as a methodology for making demographic and lifestyle inferences from analysis of captured point-of-sale data.

Planners of Public Transport systems require an understanding of patterns of commuter behaviour. This is traditionally derived by expensive survey or diary methods. We show that there are underlying similarities between EPOS and pre-paid ticket data captured by the on-bus Wayfarer system used by the Dublin Bus company and investigate Phenomenal Data Mining as complementary low-cost methodology for analysing this data.

We expect that it will be possible to make statistically valid identifications of commuter trips.

Introduction.

Phenomenal Data Mining (PDM) has been proposed, but not proven, as a methodology for making demographic and lifestyle inferences from analysis of captured point-of-sale (EPOS) data. In this paper, we investigate its utility for understanding patterns of commuter behaviour in data from an on-bus operational system. The dynamics of commuter behaviour is currently an active one - for a good overview, see (Mahmassani 1997) and in particular his discussion of day-to-day dynamics, pp 292-295.

An attractive aspect of PDM is that it identifies individual commuters. When attempting to carry this out, we found Database Visualisation methods were necessary. To help in understanding the patterns of their behaviour, we see Link Analysis as being needed. These techniques are explored later in this paper.

Typically in Data Mining projects a considerable proportion of the overall project time (up to 70-80%) has to be devoted to the preliminary phases. In a later section we look at how this has applied to the present project.

Phenomenal Data Mining.

Data Mining is concerned with the *non-trivial extraction of previously unknown and potentially useful information from databases that may be large, noisy and have missing data* (Piatetsky-Shapiro & Frawley 1991). Contributions have come from both the Machine Learning community (who, as a broad generalisation, have concentrated on pattern recognition within large but relatively clean databases) and from the Statistics community (who, as a broad generalisation, have concentrated on understanding and modelling smaller but noisier datasets). Little of this work has focused on the underlying phenomena that give rise to the observed data. Hence, much published data mining work results in aggregate techniques such as the construction of association rules or Bayesian belief networks rather than in a technique such Link Analysis which works with individual objects and their attributes.

In contrast, Phenomenal Data Mining attempts to find relations between the data and the phenomena which give rise to that data, rather than just relations among the data. A major element is the attempt to model the underlying phenomena and their attributes. John McCarthy, in his unpublished working paper "Phenomenal Data Mining" (McCarthy 1996), investigates

... what can be inferred about phenomena from data and what facts are relevant to doing this. The main technical point is that functions and predicates involving the phenomena should be explicit in the logical sentences and not just present in the mind of the person doing the data mining.

...

Science and common sense both tell us that the facts about the world are not directly observable but can be inferred from observations about the effects of actions. What people infer about the world is not just relations among observations but relations among entities that are much more stable than observations. For example, 3-dimensional objects are more stable than the image on a person's retina, the information directly obtained from feeling an object or on an image scanned into a computer.

...

The extreme positivist philosophical view that science

Record Identifier	Record Type	Main Data
82	Start of Duty	Date
83	Start of Journey	Route Id, Scheduled Start Time, Actual Start Time
84	End of Journey	Stop Time
85	Stage Update	Stage Number and Time
0	Pre-paid Ticket Validation	ID Number, Ticket Type

Table 1: Record Types and Associated Data

concerns relations among observations still influences the design of learning programs, and that's what data miners are. However, science never worked that way, neither do babies and neither should data mining programs. All obtain and use representations of the objects and use observations only as a means to that end.

He proposes a programme to mine information such as is typically captured in an Electronic Point of Sale (EPOS) system, firstly, to identify and label customers by assigning the observed basket data to an inferred phenomenon (customer) and, secondly, to infer likely attributes (age, sex, family . . .) of the phenomenon. It seems useful to consider the former task as a (sometimes essential) precursor and the latter as the core PDM paradigm. In completely anonymous data, it can be difficult to identify customers, however exploratory work in IBM, Almaden Research Laboratory, found some suggestive results in a convenience store database:

1. Packages of functionally related items (like toothbrush, toothpaste and mouthwash) grouped into baskets. This example is of particular interest because these items fell into different categories in the taxonomy used by the store chain, which generally is based on classification of customer needs.
2. Many cases of similar baskets bought in succession, probably by friends.
3. A repeated combination of school stationery items which very probably was due to a local school (almost all such baskets were bought in the same store of the chain).

An interesting aspect of this work is that it did discover some repeated shopping patterns but the underlying phenomena turned out to be other than regular customers. In retrospect, that database can be seen to have been a difficult test of the core PDM paradigm, as it contained many transient customers. It seems reasonable 'a priori', that a programme of using Data Mining techniques to infer demographics from EPOS data can be carried out for many databases, but experiment and algorithm development are needed. However, if successful, methods such as this would be a valuable, relatively low-cost, complement to traditional Market Research data.

One of the current authors (Tseytin) was involved in the Almaden work, the other author (Lyons) had previously been involved in a project to develop a system for identifying suitable locations for Ticket Agents for Dublin Bus using Wayfarer system data on pre-paid ticket usage. It seemed likely

that this database would contain some identifiable phenomena (commuters) and the present paper describes the ongoing work of identifying individual commuters and understanding the patterns of their behaviour.

Although the two databases are superficially different, there is an underlying similarity between:

- A supermarket basket where a single person (with a unique identifier of checkout number and timestamp) purchases various items each identified by a product code.
- A 'journey basket' where a single person (identified by a ticket number) takes a weekly package of trips which are identified by routes and times.

In both cases, it is the combination of items selected that we use as a basis for our inference.

The Wayfarer System.

The Wayfarer system is Dublin Bus's electronic transaction processing system, which is used to collect financial and operational data from its fleet of buses. The system includes:

- Bus-mounted equipment that captures data relating to cash ticket sales, pre-paid ticket validations and individual bus journeys.
- Depot-based and centralised equipment onto which the captured data is downloaded.

Currently, the data is used for financial control, hardware diagnostics and, to a lesser extent, marketing and route planning. While there is recognition of the potential for a more intensive analysis of the data, limited investigation of novel applications has been undertaken by Dublin Bus to date. One reason for this is that, when the system was implemented in 1989, disk storage and RAM were at a premium. The suppliers included extended data at a far-sighted request from Dublin Bus, but their standard software only processed a highly compressed subset of this data. Software modification would be required to analyse the ID number that is associated with a pre-paid ticket.

During November of 1997, Dublin Bus began supplying a portion of the data generated by the system to the School of Systems and Data Studies, Trinity College, Dublin (TCD). The data, some 40Mb per week or 2 Gb per year, are received on a daily basis and fall into two categories:

- Those that are generated at the time of a pre-paid ticket validation and which contain information on an individual ticket. No data relating to cash ticket sales has been requested.

Day	Ticket ID	Type Code	Date/Time	Route	Direction	Stage	Stage Name
Monday	137	671	26/01/98 06:45:00	66	1	46	Kew Park
	137	671	26/01/98 07:11:00	18	0	59	Palmerstown Village
	137	671	26/01/98 16:51:00	18	1	52	Walkinstown Ave
	137	671	26/01/98 17:09:00	66	0	36	Kennelsfort Rd
Tuesday	137	671	27/01/98 06:14:00	18	0	59	Palmerstown Village
	137	671	27/01/98 06:40:00	66	1	46	Celbridge Rd
	137	671	27/01/98 19:19:00	18	1	52	Walkinstown Ave
Wednesday	137	671	28/01/98 06:42:00	66	1	47	Spa Hotel
	137	671	28/01/98 16:54:00	18	1	52	Walkinstown Ave
	137	671	28/01/98 17:13:00	66B	0	36	Kennelsfort Rd
Thursday	137	671	29/01/98 06:54:00	18	0	59	Palmerstown Village
	137	671	29/01/98 17:01:00	18	1	52	Walkinstown Ave
	137	671	29/01/98 17:29:00	66B	0	36	Kennelsfort Rd
Friday	137	671	30/01/98 06:45:00	66	1	48	Celbridge Rd
	137	671	30/01/98 07:16:00	18	0	6	Palmerstown Village
	137	671	30/01/98 17:02:00	66B	0	42	Ballydowd Stores
Saturday	137	671	31/01/98 06:54:00	18	1	47	Stannaway Rd
	137	671	31/01/98 07:09:00	66	1	47	Palmerstown Village
	137	671	31/01/98 12:27:00	66	0	34	Chapelizod
Sunday	137	671					

Table 2: A week in the life of 'Miles Tripp'.

- Those that are generated by the bus drivers as they perform their duties and which contain information such as bus route numbers, times of stage changes etc.

Interpreting the Wayfarer Data.

The most important items of information are shown in table 1 on the preceding page. These data arrive in a hexadecimal transaction file. To begin to identify individual customers, pre-processing is needed to make these data understandable to a human. The first step is to decode the timestamps and stage numbers into a format such as that shown in Appendix B.

This is still not readily interpretable and we require an understanding of bus schedules and also the geographic location of stages. The degree of background knowledge required is shown in the processes needed to understand the weekly journeys of a hypothetical customer we call 'Miles Tripp'. These journeys contain most of the problems in interpretation that occur in the data and are shown in Table 2.

Commuter Journeys.

Table 2 is still not readily understandable by an analyst, and a more appropriate way of displaying it is within a Geographic Information System (GIS). Basic spatio-temporal diagrams such as in the following figures, are, we feel, much more comprehensible.

These figures show, firstly, a typical day for 'Miles Tripp' and, secondly, a day which could be interpreted naively (and incorrectly) or in a more logical way using real-world information.

Monday shows the pattern that is followed, with some exceptions, during the rest of the week - see the figure labelled

Monday on the following page. He (or she) takes an early morning bus in Lucan to Palmerstown, where he switches to one of the relatively rare cross-town services (Route 18) which takes him to his place of work in Walkinstown.

Tuesday is more problematic - the data say that he commences the day by taking an 0645 number 18 bus in Palmerstown - see the figure labelled Tuesday: Door 1 on the following page. This is consistent with his early start in Walkinstown but there are two problems, one minor and the other major, with this interpretation. The minor problem is that the boarding point is quite far from his home, but we could explain this away by his getting a lift - we might be tempted to hypothesise a car journey, were it not for the major problem that, according to the bus schedule, there is no 0645 number 18 bus! So the naïve interpretation of a car journey, followed by the realisation that he had forgotten something at home, a journey back to collect it, leading to him being late to work and working late that evening to make up time, fails the Occam's Razor test in favour of the following interpretation.

The more logical explanation of Tuesday is that the clock on the Southbound 18 bus is wrong by 1 hour - see the figure labelled Tuesday:Door 2 on the following page. This results in a perfectly normal start of day. It should be possible to independently verify the hypothesis of the incorrect clock. In the evening, Miles may have stayed in Palmerstown or his journey home was not recorded for some reason. The latter is more likely, since he starts from home as normal on Wednesday morning. Again, it might be possible to find within the data circumstances to explain the possibly missing journey - for example, we might find that a scheduled Westbound 66 or 66B ceased to record tickets at some stage before Palmerstown, suggesting a machine malfunction.

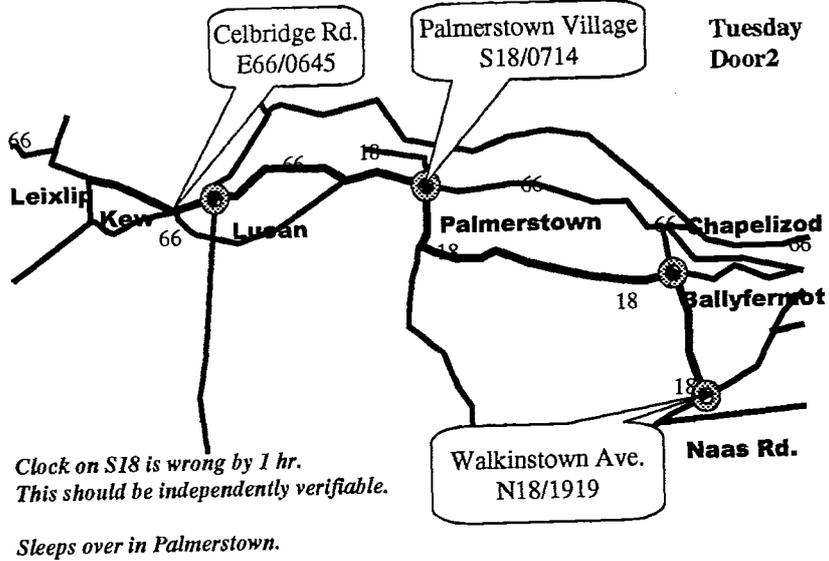
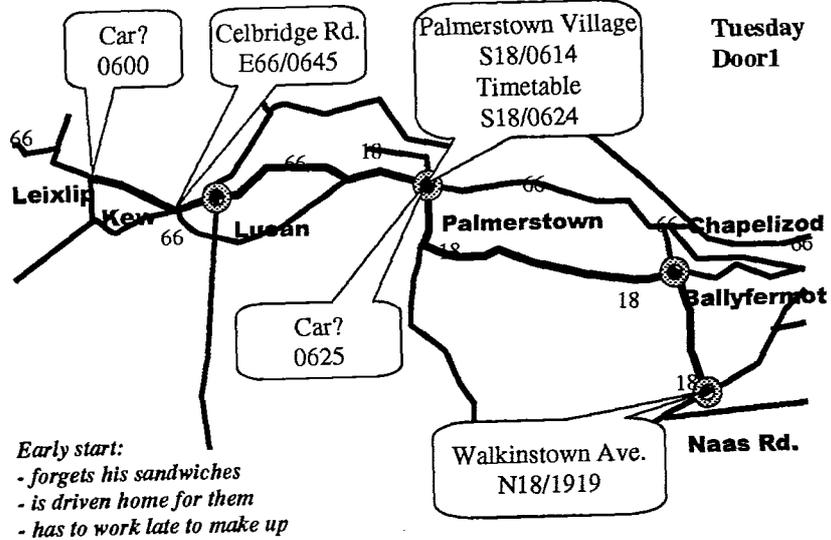
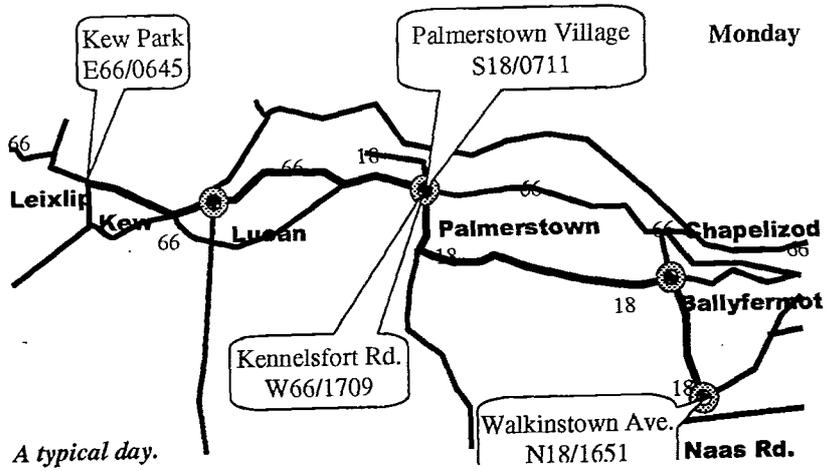


Figure 1: A typical day and one needing interpretation.

Customer Identification.

For the week discussed, we have been dealing with a single ticket and hence, we would expect, with a single customer. However, the 'basket' of journeys identified here can form the basis for customer identification in subsequent weeks. Related work on the discovery of Frequent Episodes in Event Sequences appears in (Mannila, Toivonen, & Verkamo 1997).

If we find a similar pattern of journeys again, we will probably have met Miles once more. We may even be able to identify the time (but almost certainly not the location) of his annual holidays, on the Sherlock Holmes principle of the 'dog not barking'.

The weekly set of journeys constitutes what John McCarthy (op. cit.) calls a 'signature':

Definition: Associated with an assignment and a customer c is a characterization of the putative customer. The characterization may include qualitative characteristics like sex or owning a freezer, quantitative characteristics like age or income group and other customer characteristics like a certain purchase signature. The anomaly $anom(c)$ associated with a customer c depends on the characterization. Thus buying chewing tobacco or baby food is more anomalous for some customers than others. A program that generates assignments will generate characterizations as it groups the baskets by customer. The characterization itself will contribute to the anomaly if it is an unusual characterization.

Definition: A signature is a set of choices among alternate brands or sizes of certain commodities. The commodities most useful for signatures are those for which variety is not normally considered desirable. While a person may want variety in food he is unlikely to want variety per se in dish-washing soap, toilet paper or size of dog food. Signatures are included in the characterization of a customer.

The part of the anomaly $anom(c)$ associated with the putative customer c is computed relative to the characterization. Thus if c is characterized as single young female, a purchase of chewing tobacco should have a higher anomaly score than for a male.

One way of looking at minimizing anomaly of assignments is that we want to explain as much of the purchasing behavior as possible by allowable characterizations of the customers.

We can derive a characterisation of Miles from his weekly journeys. He lives near the Spa Hotel in Lucan - possibly in nearby Kew Park. He works in Walkinstown - possibly in the Robinhood Industrial Estate and works a 0730 to 0430 day, Monday to Friday. On Saturday he works from 0730 to 1230, a 50-hour week. More tenuous lifestyle inferences are that he is young, male and averse to walking.

Miles might be a 'strongly signed' customer - one who falls into a class of his own. As such, perhaps he would not be of immediate interest to Dublin Bus. However, more investigation of the data might show that, in fact, there was

an equivalence class of commuters who make a route change from the 66 to the 18. This class would be of interest, as it indicates a minimum level of demand for a new route. Such new routes are occasionally introduced by Dublin Bus, for example, the very successful cross-town 75 route from Tallaght - Dun Laoghaire. Such 'trip chaining' commuters are discussed in (Mahmassani, Hatcher, & Caplice 1997).

Database Visualisation.

A good introduction to this area can be found in (Keim 1996). So far, we have limited ourselves to basic GIS facilities. Over the coming months we intend to develop this further using a full GIS system.

However, as a result of identifying customers and their journeys, we have a rich network for visualisation. Computer programs for visualisation of networks are under development, for example, the work of Alden Klovdahl (View_Net), Vincent Duquenne (GLAD), Lothar Krempel (NetVis) and David Krackhardt (KrackPlot).

One can imagine some interesting visualisations of the data set, for example, a map of Dublin on which Bus routes are highlighted. This map might allow rotation in three dimensions and zoom to examine particular areas of interest. This would aid in visualising the spatial structure of both buses and ticket validations.

Using the time data, it would be possible to create an animated description of bus and ticket usage over a time period. Three possible uses of such an interface would be:

- To track individual ticket usage, one or more tickets would be selected. According to the time series of their validations, dots might appear on the map. This would give both spatial and temporal information in a readily understandable form.
- Each bus might be represented by a cuboid that grows in height relative to the number of validations associated with the bus journey. The bus would start its journey as a little cube, and as more and more passengers validated tickets, the cube would grow correspondingly. Since there is no record of when a commuter alights, that data would need to be inferred from the commuter's basket of journeys.
- To perform an analysis of passengers who use one or more routes on a regular basis, such tickets would first be identified. A similar visualisation to that above would then be used, resulting in the highlighting of routes that are used by such passengers and, hopefully, the identification of areas where new routes might be appropriate.

Link Analysis.

In our dataset, stages can be considered as nodes and customer journeys as links between them. There is a time dimension, in that the journeys can be looked at daily, weekly, monthly or yearly depending on what specifically is being studied. Our linkage data is of the 'simple but voluminous' class, with a uniformity of node and link types, and a great deal of regularity.

It is worth considering how the traditional questions posed by Link Analysis are relevant to our study:

- *Which nodes are "key" or central to the network?*

The radial nature of Dublin Bus routes means that there is a clear centre to the existing network, with many geographically close high-volume nodes. There also are suburban nodes where several routes terminate close to each other. For many commuters, these routes effectively link two 'super-nodes'. This would not be true for all commuters, as some alight along the course of a route.

The raw data does not allow identification of such commuters since a journey only records the boarding and not the alighting stage. Further processing can identify the 'home' and 'work' stage for each ticket. Our PDM analysis strengthens this further, as we identify the persistent underlying phenomenon of a commuter from the more transient ticket data.

The importance of a node is primarily the volume of commuters it receives but the time distribution of this volume also matters. Some nodes also have the characteristic that they are switching nodes - these have a special importance to that subset of commuters who chain trips.

- *Which links can be strengthened to most effectively enhance the operation of the network?*

Enhancing the capacity of a link consists of providing a more frequent service or creating a new route including the start and end nodes. This question is of considerable interest in our context.

- *Can the existence of undetected links or nodes be inferred from the known data?*

We have already mentioned that identifying the potential demand for new routes is of interest.

- *Are there similarities in the structure of sub-parts of the network which may indicate an underlying relationship?*

It is not clear that this applies in our context.

- *What are the relevant sub-networks within a much larger network?*

There are some clear sub-networks within the Dublin route structure. We expect that our data could infer these even if they were not already known. What is of interest is whether any unknown sub-networks will emerge.

- *What data model and level of aggregation best reveal certain types of links and subnetworks?*

We have options for both spatial and temporal aggregation and intend to investigate which work best in defining commuter signatures.

Relevant phases of a Data Mining project.

Recent texts on Data Mining (Berry & Linoff 1997), (Adriaans & Zantinge 1996), (Cabena *et al.* 1997) are in general agreement on the relevant phases of a project which are:

Data Selection. Initial data selection is carried out by Dublin Bus who run filter programs against the full

dataset. For example, no data relating to cash ticket sales is received. Neither is any data relating to individual drivers.

Our analysis is then concentrated on specific types of tickets - initially, weekly tickets.

Data Cleaning. For an operational system, the data is impressively clean. However, there are inevitably both noise and missing data. One source of noise is due to the physical nature of the ticket. It incorporates a magnetic strip which can be damaged by contact with metal, severe bending etc.

Data files pass through a number of different machines and formats before final storage on an NT server in T.C.D. Occasionally, sections of a file are either not captured or are lost at some stage of the process.

A whole day's data file has even been lost due to human error on the part of the first author!

Various levels of data cleaning have been considered, from straightforward error reporting of unexpected data through to semi-automatic analysis and correction of the more frequent type of error.

Data Transformation. The raw transaction file has to be read as binary data and analysis programs transform this to human readable form.

As an interim measure, other programs were written to populate an Access database, allowing ad hoc queries to be made. It is intended to migrate to an Oracle database during the coming months.

Data Enrichment. In this case, we have added the stage names and reconciled some inconsistencies among them. This allows us to translate stage numbers from the data files to meaningful names for various reports, for example, that of Appendix B.

We are in the process of adding map co-ordinates to these, to provide those basic GIS facilities that we feel are essential in understanding the data.

We are considering adding schedule information that would allow more refined data cleaning.

As a general point, we consider that Data Enrichment is an essential part of any meaningful Data Mining project. Each database mined has its own metadata and associated real-world knowledge which must be incorporated before any true understanding can be reached. Hence, we consider that the "Holy Grail" of complete automation of Knowledge Discovery in Databases is well beyond the current state of the art. What can be contemplated is the crafting of real-world knowledge for limited portions of real databases to allow semi-automated knowledge discovery.

Reporting. Working with large databases creates very real problems of deciding what to report. It is very easy to generate so much paper that not only is the wood not seen for the trees but that the trees themselves are hidden in foliage.

To some extent, PDM addresses this problem, since it attempts to identify underlying phenomena that are sparser than the data.

Acknowledgments. The authors thank Dublin Bus staff who have supported this research effort with generous contributions of their time and data. Mr Gary Thompson, then of the Management Science and Information Systems programme, and now with Price-Waterhouse, provided much of the material on operational aspects and did investigative database work. Our thanks are also due to Professors John Haslett and John McCarthy for discussion and support.

Grants from Trinity College, Dublin's *Provost's Academic Development Fund* and *Visiting Professorships and Fellowships Benefaction Fund* partially supported equipment and development work.

Appendix A: Data and Methods of Collection. On board each bus, there are two pieces of hardware, which record onto a data cartridge. These are:

- A control unit, known as an Electronic Ticket Machine (ETM), mounted beside the driver consists of an electronic clock, a cash ticket issuing machine, a digital display, and a keypad. The driver uses the keypad to enter information concerning the current bus journey, which consists of a start of journey transaction, the route number, and the current stage the bus is at along the route and an end of journey transaction. The control unit combines the drivers' inputs with time data to form transaction records that are stored on the driver's data cartridge.
- A magnetic card validation unit connected to the control unit that is used by passengers to validate pre-paid tickets. A magnetic strip that contains information including an ID number, a ticket type code, and details of the ticket's previous journey is printed on each pre-paid ticket. This information is read by the validation unit and then is passed to the control unit. A transaction is created which is stored in the driver's data cartridge. Normally, all the information on the magnetic strip is read and rewritten.

The Driver removes the data cartridge from the bus at the end of his duty for downloading onto a depot computer. Data from the depot computers is consolidated onto Dublin Bus mainframes, a filter program is run to remove certain types of transaction records and the resulting daily file is mailed to the first author, normally arriving the following morning.

Five main types of Wayfarer transaction records are included in the data set provided by Dublin Bus. These are as shown in table 1 on page 2.

Appendix B: Ticket usage output. The usage of an individual ticket (serial number 8869) over a two-week period, as generated by an analysis program, are shown. To a person familiar with the route structure of Dublin, it is apparent that these two weekly tickets have been used by different commuters. This would, of course, be expected since this ticket has a weekly type. If the ticket had been, for example, a monthly type these data would be unexpected.

Ticket type: Weekly Adult Cityzone, serial no. 8869. Mon Jan 26 07:50 Route 20B, Ardlea Rd. (Ardlea Inn), -> St. Stephen's Green

Mon Jan 26 08:09 Route 16B, Shanowen Rd., -> Beaumont Ave.
Tue Jan 27 07:50 Route 20B, Ardlea Rd. (Ardlea Inn), -> St. Stephen's Green

Tue Jan 27 08:02 Route 16, Shanowen Rd., -> Grange Rd.
Tue Jan 27 17:24 Route 27B, Beresford Pl. (Liberty Hall), -> Castletimon (Ballyshannon Ave.)

Wed Jan 28 07:28 Route 20B, Ardlea Rd. (Ardlea Inn), -> St. Stephen's Green

Wed Jan 28 07:45 Route 16, Shanowen Rd., -> Grange Rd.

Wed Jan 28 17:19 Route 27B, Beresford Pl. (Liberty Hall), -> Castletimon (Ballyshannon Ave.)

Thu Jan 29 07:34 Route 20B, Ardlea Rd. (Ardlea Inn), -> St. Stephen's Green

Thu Jan 29 08:01 Route 16, Shanowen Rd., -> Grange Rd.

Thu Jan 29 13:13 Route 11A, Dorset St. (North Frederick St.), -> Mather Rd. North

Thu Jan 29 17:25 Route 27B, Beresford Pl. (Liberty Hall), -> Castletimon (Ballyshannon Ave.)

Thu Jan 29 21:30 Route 20B, Annesley Bridge Rd., -> Ardlea Rd. (Ardlea Inn)

Fri Jan 30 07:58 Route 20B, Ardlea Rd. (Ardlea Inn), -> St. Stephen's Green

Fri Jan 30 08:23 Route 16A, Shanowen Rd., -> Beaumont Ave.

Fri Jan 30 17:24 Route 27B, Beresford Pl. (Liberty Hall), -> Castletimon (Ballyshannon Ave.)

Sun Feb 01 20:23 Route 22, O'Connell St., -> Drimmagh Rd.

Sun Feb 01 23:27 Route 77, Drimmagh Rd. (Half Way House), -> Jobstown

Mon Feb 02 11:12 Route 77, The Kestrel, -> Eden Quay

Mon Feb 02 11:37 Route 13A, Merrion Sq., -> Poppintree

Mon Feb 02 17:06 Route 19A, The Rise (Ballymun Rd.), -> Bulfin Rd.

Mon Feb 02 17:34 Route 134, Phibsboro Shopping Centre, -> Middle Abbey St.

Mon Feb 02 20:13 Route 155, South Great Georges St., -> St. James Rd. (Shopping Centre)

Tue Feb 03 08:31 Route 50, The Kestrel, -> Crumlin Rd. (Loreto Convent)

Tue Feb 03 09:16 Route 13A, O'Connell St., -> Poppintree

Tue Feb 03 17:18 Route 13, Sillogue Rd. (Ballymun S. C.), -> Merrion Sq.

Tue Feb 03 18:05 Route 77, Eden Quay, -> Jobstown

Tue Feb 03 20:19 Route 56A, Sylvan Dr., -> Eden Quay

Tue Feb 03 20:40 Route 77A, Eden Quay, -> The Square

Wed Feb 04 11:29 Route 77, The Kestrel, -> Eden Quay

Wed Feb 04 12:18 Route 13A, O'Connell St., -> Poppintree

Wed Feb 04 17:57 Route 123, Croyden Pk. Ave., -> Crumlin Hospital

Wed Feb 04 18:13 Route 77A, Eden Quay, -> The Square

Wed Feb 04 20:54 Route 77, Drimmagh Rd. (Half Way House), -> Jobstown

Thu Feb 05 09:55 Route 155, Cromwells Fort Rd./St. Peter's Rd., -> College St.

Thu Feb 05 10:18 Route 13A, Merrion Sq., -> Poppintree

Thu Feb 05 15:53 Route 17A, Ballymun Rd. (Library)/Glasnevin Ave., -> Mellows Rd. (Cnr. Kildonan Rd.)

Thu Feb 05 16:15 Route 134, Finglas (Jamestown Rd.), -> Middle Abbey St.

Thu Feb 05 16:57 Route 77A, Eden Quay, -> The Square

Fri Feb 06 14:25 Route 155, Cromwells Fort Rd./St. Peter's Rd.,
-> College St.
Fri Feb 06 15:25 Route 83, Bishop St., -> College St.
Fri Feb 06 15:34 Route 7, Burgh Quay, -> Loughlinstown Pk.
Fri Feb 06 16:42 Route 18, Sandymount Ave., -> Old Lucan Rd.
(Hollyville Lawn)

References

- Adriaans, P., and Zantinge, D. 1996. *Data Mining*. Addison Wesley.
- Berry, M. J. A., and Linoff, G. 1997. *Data Mining Techniques: For Marketing, Sales, and Customer Support*. John Wiley.
- Cabena, P.; Hadjinian, P.; Stadler, R.; Verhees, J.; and Zanasi, A. 1997. *Discovering Data Mining: From Concept to Implementation*. Prentice Hall.
- Keim, D. A. 1996. Databases and visualization. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*. Tutorial.
- Mahmassani, H. S.; Hatcher, S. G.; and Caplice, C. G. 1997. Daily variation of trip chaining, scheduling, and path selection behaviour of work commuters. In *Understanding Travel Behaviour in an era of change*. eds. Peter Stopher and Martin Lee-Gosselin, Pergamon Press (1997) 351-379.
- Mahmassani, H. S. 1997. Dynamics of commuter behaviour: Recent research and continuing challenges. In *Understanding Travel Behaviour in an era of change*. eds. Peter Stopher and Martin Lee-Gosselin, Pergamon Press (1997) 279-313.
- Mannila, H.; Toivonen, H.; and Verkamo, A. I. 1997. Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*. 1(3):??
- McCarthy, J. 1996. Phenomenal data mining. URL=<http://www-formal.stanford.edu/jmc/data-mining.html>.
- Piatetsky-Shapiro, G., and Frawley, W., eds. 1991. *Knowledge Discovery in Databases*. Cambridge, MA: The MIT Press.
- Stopher, P., and Lee-Gosselin, M., eds. 1997. *Understanding Travel Behaviour in an era of change*. Pergamon.