European Journal of Population (2006) 22: 37–65
DOI 10.1007/s10680-005-5549-0

Timing, Sequencing, and Quantum of Life Course Events: A Machine Learning Approach

FRANCESCO C. BILLARI[1,*], JOHANNES FÜRNKRANZ[2] and ALEXIA PRSKAWETZ[3]
[1]*Institute of Quantitative Methods, Università Bocconi and IGIER, Viale Isonzo 25, I-20135, Milan, Italy;* [2]*Department of Computer Science, Knowledge Engineering Group, TU Darmstadt, Hochschulstrasse 10, D-64289, Darmstadt, Germany;* [3]*Vienna Institute of Demography, Prinz Eugen Strasse 8–10, A 1040, Wien, Austria;*
(Author for correspondence: E-mail: francesco.billari@unibocconi.it)

Billari F. C., Furnkranz J., and Prskawetz A. 2006. Timing, Sequencing, and Quantum of Life Course Events: A Machine Learning Approach. European Journal of Population, **22**: 37–65

Abstract. In this paper we discuss and apply machine learning techniques, using ideas from a core research area in the artificial intelligence literature to analyse simultaneously timing, sequencing, and quantum of life course events from a comparative perspective. We outline the need for techniques which allow the adoption of a holistic approach to life course analysis, illustrating the specific case of the transition to adulthood. We briefly introduce machine learning algorithms to build decision trees and rule sets and then apply such algorithms to delineate the key features which distinguish Austrian and Italian pathways to adulthood, using Fertility and Family Survey data. The key role of sequencing and synchronization between events emerges clearly from the analysis.

Keywords: data mining, event history, life course, machine learning, transition to adulthood

Billari F.C., Fürnkranz J., et Prskawetz A., 2006. Calendrier, séquence et intensitédes événements du cycle de vie : une application des techniques d'apprentissage par machine. Revue Européenne de Démographie, **22**: 37–65

Résumé. Dans cet article, nous appuyant sur des éléments centraux de la recherche en intelligence artificielle, nous appliquons les techniques d'apprentissage par machine pour analyser simultanément le calendrier, la séquence et l'intensité des événements du cycle de vie d'un point de vue comparatif. Nous soulignons le besoin de techniques qui permettent une approche holistique de l'analyse du cycle de vie, en illustrant ici le cas particulier de la transition vers l'âge adulte. Nous présentons brièvement les algorithmes d'apprentissage qui permettent de construire des arbres décisionnels et des ensembles de règles et appliquons ces algorithmes pour identifier, à partir des données des enquêtes Fécondité et Famille, les principaux traits qui distinguent les chemins de transition vers la vie adulte des Autrichiens et des Italiens. L'analyse fait ressortir le rôle clé joué par la séquence et la synchronisation des événements.

Mots clés: analyse biographique, apprentissage par machine, cycle de vie, fouille de données, transition vers l'âge adulte

1. Introduction

Demographers are mostly concerned with the study of major events that shape people's lives such as births, deaths, migrations, and the formation and dissolution of households and families. The life course approach – the theoretical framework behind many recent studies – sees above all the demographic trajectories of individuals as mutually interconnected and in turn linked to other trajectories (of which education and labour market careers are among the most important ones) (van Wissen and Dykstra, 1999). Trajectories are marked by demographic events, or by events which are thought to have an influence on demographic behaviour. The life course approach to the study of demographic behaviour is thus intrinsically characterized by a holistic point of view. Nevertheless, techniques used so far hardly allow to take such a holistic perspective.

One of the fields that has attracted increasing interest in the demographic life course literature of the last years is the study of the transition to adulthood. In this field, the main emphasis is on the study of the *timing*, the *sequencing* (Hogan, 1978), and sometimes the *quantum* of specific events – which usually happen during early adulthood – for a specific cohort of individuals. These events are normally considered to be indicators of the transition from roles typical of youth to roles typical of adulthood. For the sake of simplicity, the age at which events are experienced is taken as an indicator of the timing, the observed order as an indicator of sequencing, and the observed number of events as an indicator of the quantum. In the latter case (quantum), if one focuses on the transition to adulthood, the main issue is whether an event is experienced at all during the life of an individual. As far as the sequencing is concerned, a critical issue is the simultaneity of events, i.e., the experiencing of events during the same time unit, also known as *synchronization* (Mulder and Wagner, 1993). Following a seminal paper by Modell et al. (1976), most of the papers studying the transition to adulthood analyse some specific events: leaving formal education, entering the labour market, leaving the parental home, experiencing the first union (sometimes with a differentiation between marriages and consensual unions), and becoming a parent. This approach is not the only one that could be adopted in a study on the transition to adulthood (Marini, 1987), but it provides a widely used framework when one wants to analyse the determinants and study the dynamics of behaviour within a society by comparing cohorts, genders, social groups, and/or different societies. A holistic perspective on the transition to adulthood would need to take simultaneously into account timing, sequencing and quantum.

The set of statistical techniques which is broadly defined as *event history analysis* constitutes nowadays one of the principal toolkits of demography (see e.g. Courgeau and Lelièvre, 1992). Event history techniques focus on the

time-to-event as the dependent variable. They also allow researchers to study very complex interdependencies between events in the life course, handling unobserved factors underlying these complex interdependencies (Lillard, 1993). Event history analysis does not, however, allow one to adopt a holistic perspective on the life course, i.e., to see the set of events that shape the lives of individuals as a coherent set and to compare this set for different individuals or groups of individuals. More specifically, event history analysis allows for the analysis of the timing and, with some specific assumptions, also of the quantum of events, but it does not allow for the simultaneous study of the timing, sequencing and quantum.

Life course analysis aims at illuminating on causal relationships shaping individual trajectories. However, it also aims at providing ideal-types of trajectories and exploratory tools that allow researchers to read the complexity of life courses in an adequate way (Billari, 2003). In this paper, we use an "algorithmic modelling approach" (Breiman, 2001), of which only rare examples of application to demographic issues exist. An example is an analysis based on *survival trees* by De Rose and Pallara, applied to the search for determinants of the timing of demographic events (De Rose and Pallara, 1997). Similar techniques are based on monothetic divisive algorithms that result in a classification tree (Billari and Piccarreta, 2005); the latter have been proposed for grouping individuals according to their states at different points in time, without explicitly taking into account the order of events.

Our first contribution, with a primarily methodological aim, is a solution for the problem of analysing simultaneously the timing, the sequencing (including the synchronization on a monthly time basis), and the quantum of events in life courses from a holistic point of view. To that end, we devise a representation that makes life course data amenable to conventional algorithmic modelling techniques.

The second main contribution of this paper is an analysis of a dataset in the proposed representation, which allows us to analyse simultaneously the timing, the sequencing, and the quantum of events in life courses. In particular, we explore the suitability of two techniques developed in the machine learning community (Mitchell, 1997; Witten and Frank, 2000), a premier representative of the algorithmic modelling culture, to the problem of life course analysis, namely decision tree learning (Murthy, 1998) and the induction of classification rules (Fürnkranz, 1999). We use these techniques to detect the basic features that differentiate the transition to adulthood in two European countries, Austria and Italy. Our results underline the crucial role of the information about the sequencing of events in the analysis of transition to adulthood. We note that statistical techniques for learning a classification function, such as logistic regression, neural networks, support vector machines and others, could also be used

for analysing life course data in the proposed representation (Hastie et al., 2001). The main advantage of decision trees and classification trees lies in their support for trading off the simplicity and the accuracy of the resulting models. We demonstrate how we used such *pruning* techniques to find accurate and comprehensible models that are based on only a few discriminatory variables.

The paper is structured as follows. In Section 2, we introduce some basic notions of the machine learning and data mining approach, which are partially novel to a social science audience. In Section 3, we present the data and the representation we use and discuss some of the basic features of the transition to adulthood in Austria and Italy. Section 4 introduces the experimental setup, and it provides a presentation and discussion of the results. Section 5 contains some final remarks.

2. Machine Learning and Data Mining

Machine learning is one of the core research areas in Artificial Intelligence. Currently, the most prominent research topic within the field is the inductive analysis of databases. Together with statistics and database technology, this area provides the core methodologies for the rapidly developing field of *Knowledge Discovery in Databases*, also known as *Data Mining* (Fayyad et al., 1995), which has recently attracted the interest of industry and is considered by many to be one of the fastest-growing commercial application areas for Artificial Intelligence techniques. Machine learning and data mining systems are used for analysing telecommunications network alarms, supporting medical applications, detecting cellular phone fraud, assisting basketball trainers, controlling elevators, categorizing celestial bodies, and classifying documents on the World-Wide Web, to name only a few applications. A selection of recent applications in machine learning and data mining can be found in Michalski et al. (1998), and excellent textbooks for the research area are Mitchell (1997) and Witten and Frank (2000). Within the social sciences, however – and demography is no exception – these tools have not yet received much attention despite the importance of data-oriented research.

In the remainder of this section, we will briefly introduce the classification problem we are dealing with and discuss two common approaches to solving it: inducing decision trees and inductive rule learning. It can be safely skipped by readers familiar with these techniques.

2.1. PROBLEM DESCRIPTION

The task that has received the most attention in the machine learning literature is the following: given a database of observations (described with a

fixed number of measurements x_i, so-called *features* or *attributes*) and a designated attribute y, the *class*, find a mapping f that is able to compute the class value $y = f(x_1,...,x_n)$ from the feature values of new, previously unseen observations. While there are statistical techniques that are able to solve particular instances of this problem, machine learning techniques provide a strong focus on the use of categorical, non-numeric attributes, and on the immediate interpretability of the result. They typically also provide simple means for adapting the complexity and comprehensibility of the models to the problem at hand. This, in particular, is one of the main reasons for the increasing popularity of machine learning techniques in both industry and academia.

Table 1 shows a very small, artificial sample database, adapted from Quinlan (1986). The database contains the results of a survey on 14 individuals, concerning the approval or disapproval of a certain issue. Each individual is characterized by four attributes – *Education* (with possible values *primary* school, *secondary* school, or *university*), *Marital status* (with possible values *single*, *married*, or *divorced*), *Sex* (*male* or *female*), and *Has children* (*yes* or *no*) – that encode rudimentary information about the sociodemographic background. The last column, *Approve*?, encodes whether the individual approved or disapproved the issue.

The task is to use the information in this *training set* to derive a model that is able to predict whether a person is likely to approve or disapprove, based on the four demographic characteristics.

Table 1. A sample database

Education	Marital status	Sex	Has children	Approve?
Primary	Single	Male	No	No
Primary	Single	Male	Yes	No
Primary	Married	Male	No	Yes
University	Divorced	Female	No	Yes
University	Married	Female	Yes	Yes
Secondary	Single	Male	No	No
University	Single	Female	No	Yes
Secondary	Divorced	Female	No	Yes
Secondary	Single	Female	Yes	Yes
Secondary	Married	Male	Yes	Yes
Primary	Married	Female	No	Yes
Secondary	Divorced	Male	Yes	No
University	Divorced	Female	Yes	No
Secondary	Divorced	Male	No	Yes

2.2. INDUCTION OF DECISION TREES

The induction of decision trees is one of the oldest and most popular techniques for learning discriminatory models, which has been developed independently in the statistical (Kass, 1980; Breiman et al., 1984) and machine learning (Quinlan, 1986) communities. A *decision tree* is a particular type of classification model that is fairly easy to induce and to understand.[1] Figure 1 shows a sample tree which might be induced from the data of Table 1. Classification of a new example starts at the top node – the *root* – and the value of the attribute that corresponds to this tree is considered (*Marital Status* in the example). Classification then proceeds by moving down the branch that corresponds to a particular value of this attribute, arriving at a new node with a new attribute. This process is repeated until we arrive at a terminal node – a so-called *leaf* – which is not labelled with an attribute but with a value of the target attribute (*Approve*?). For all examples that arrive at the same leaf value, the same target value will be predicted. Figure 1 shows leaves as rectangular boxes.

Decision trees are learned in a top-down fashion: the program selects the best attribute for the root of the tree, splits the set of examples into disjoint sets (one for each value of the chosen attribute, containing all training examples that have the corresponding value for this attribute), and adds corresponding nodes and branches to the tree. If there are new sets that contain only examples from the same class, a leaf node is added for each of them and labelled with the respective class. For all other sets, an interior node is added and associated with the best attribute for the corresponding set as described above. Hence, the dataset is successively partitioned into non-overlapping, smaller datasets until each set only contains examples of the same class (a so-called *pure* node). Eventually, a pure node can always be

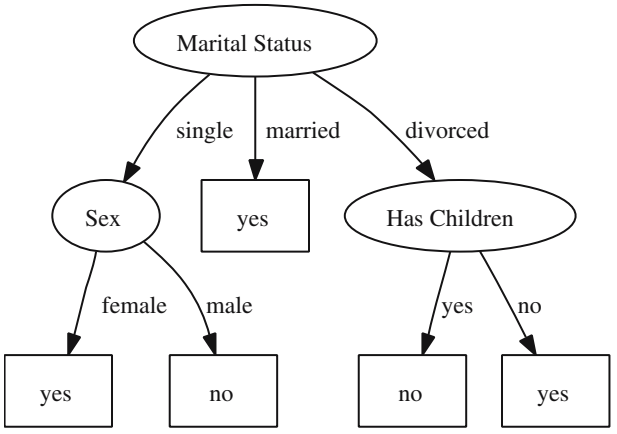

Figure 1. A decision tree describing the dataset shown in Table 1.

found via successive partitions unless the training data contains two identical but contradictory examples, i.e., examples with the same feature values but different class values.

The crucial step in decision tree induction is the choice of an adequate attribute. To see the importance of this choice, consider a procedure that constructs decision trees simply by picking the next available attribute. The result is a much more complex and less comprehensible tree (Figure 2). Most leaves originate from a single training example, which means that the labels that are attached to the leaves are not very reliable. Although the trees in Figures 1 and 2 will both classify the training data in Table 1 correctly, the former appears to be more trustworthy, and it has a higher chance of correctly predicting the class values of new data.[2]

Typical attribute selection criteria use a function that measures the *purity* of a node, i.e., the degree to which the node contains only examples of a single class. This purity measure is computed for a node and all successor nodes that result from using an attribute for splitting the data. The difference between the original purity value and the sum of the purity values of the successor nodes weighted by the relative sizes of theses nodes, is used to estimate the utility of this attribute, and the attribute with the largest utility is selected for expanding the tree. C4.5 uses information-theoretic entropy as a purity measure (Quinlan, 1986), whereas CART uses the Gini index (Breiman et al., 1984).[3] Thus, the final tree is constructed by a sequence of local choices that only take that part of the examples into consideration that end up at the node that is currently split. Of course, such a procedure can only find local optima for each node, but cannot guarantee convergence to a global optimum (the smallest tree).

Also note, that some of the attributes may not occur at all in the tree. For example, note that the tree in Figure 1 does not contain a test on *Education*. Apparently, the data can be classified without making a reference to this variable. This will be automatically found by the algorithm. More generally, one can say that the attributes in the upper parts of the tree (near the root) tend to have a stronger influence on the value of the target variable than the nodes in the lower parts of the tree, because of the way the selection function orders the attributes at each choice point. We will use this property for interpreting our results in Section 4.

The chief problem with this approach is that, as a result of the recursive partitioning of the data, the number of examples that end up in each node decreases steadily. As a consequence, the reliability of the chosen attributes decreases with increasing depths of the tree. As a result, overly complex models are generated, which explain the training data but do not generalise well to unseen data. This is known as *overfitting*. Overfitting can be best understood if one considers that one can perfectly fit any given time series if

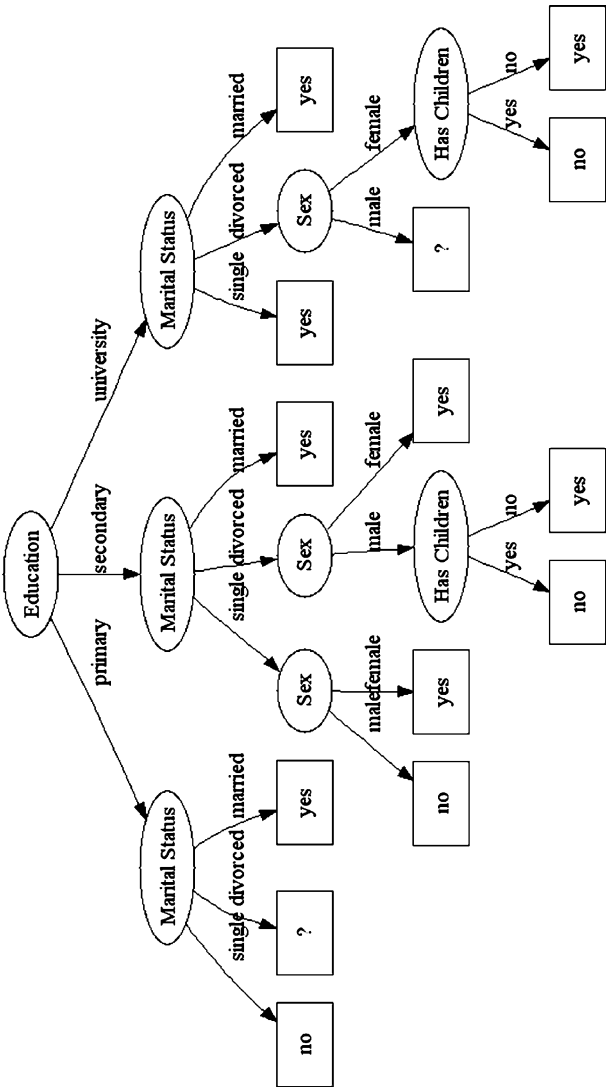

Figure 2. A bad decision tree describing the dataset shown in Table 1.

one can use a polynomial of an arbitrarily high degree. However, the inherent trend (if any) can only be captured if one restricts to models with a lower number of degrees of freedom (e.g., linear). For decision tree models, this basically corresponds to restricting the number of nodes in the tree. This is the main reason why state-of-the-art decision tree induction techniques employ a post-processing phase in which the tree generated with the above procedure is simplified by *pruning* branches and nodes near the leaves. In effect, this procedure replaces some of the interior nodes of the tree with a new leaf, thereby removing the subtree that was rooted at this node. The exact details of this procedure are beyond the scope of this paper (we again refer to Quinlan, 1993), but it is important to note that the leaf nodes of the new tree are no longer pure nodes, i.e., they no longer contain training examples that all belong to the same class. Typically, this is simply resolved by predicting the most frequent class at a leaf. The class distribution of the training examples within the leaf may be used as a reliability criterion for this prediction.

2.3. INDUCTION OF RULE SETS

Another important machine learning technique is the induction of rule sets. The learning of rule-based models has been a main research goal in the field of machine learning since its beginning in the early 1960s. For a detailed survey of rule learning algorithms we refer the reader to Fürnkranz (1999). Recently, rule-based techniques have also received increased attention in the statistical community (Friedman and Fisher, 1999).

Rule sets are typically simpler and more comprehensible than decision trees. To see why, note that a decision tree can also be interpreted as a set of **IF–THEN** rules. Each leaf in the tree corresponds to one rule, where the conditions encode the path that is taken from the root to this particular leaf, and the conclusion of the rule is the label of that leaf. Figure 3 shows the set

IF Marital Status = single **AND** Sex = female
THEN yes
IF Marital Status = single **AND** Sex = male
THEN no
IF Marital Status = married
THEN yes
IF Marital Status = divorced **AND** Has Children = yes
THEN no
IF Marital Status = divorced **AND** Has Children = no
THEN yes

Figure 3. A rule set describing the dataset shown in Table 1.

of rules that corresponds to the tree in Figure 1. Note the rigid structure of these rules. For example, the first condition always uses the same attribute, namely, the one used at the root of the tree.

The main difference between the rules generated by a decision tree and the rules generated by a rule learning algorithm is that the former rule set consists of non-overlapping rules that span the entire instance space (i.e., each possible combination of feature values will be covered by exactly one rule). Relaxing this constraint, i.e., allowing potentially overlapping rules that need not span the entire instance space, may often result in smaller rule sets.

However, in this case, we need mechanisms for tie breaking (i.e., which rule to choose when more than one covers the given example) and default classifications (what classification to choose when no rule covers the given example). Typically, one prefers rules with a higher ratio of correctly classified examples from the training set.

Figure 4 shows a particularly simple rule set which uses two different attributes in its first two rules. Note that these two rules are overlapping, i.e., several examples will be covered by more than one rule. For instance, examples 3 and 10 are covered by both the first and the third rule. These conflicts are typically resolved by using the more accurate rule, i.e., the rule that covers a higher proportion of examples that support its prediction (the first one in our case).[4] Also note that this rule set makes two mistakes (the last two examples). These might be resolved by resorting to a more complex rule set (like the one in Figure 3) but as stated above, it is often more advisable to sacrifice accuracy in the training set for model simplicity to avoid overfitting. Finally, note the default rule at the end of the rule set. This is added for the case when certain regions of the data space are not represented in the training set.

The key ideas for learning such rule sets are quite similar to the ideas used in decision tree induction. However, instead of recursively partitioning the dataset by optimising the purity measure over all successor nodes (in the literature, this strategy is also known as *divide-and-conquer* learning),

IF Marital Status = married
  THEN yes
IF Sex = female
  THEN yes
IF Sex = male
  THEN no
DEFAULT yes

Figure 4. A smaller rule set describing the dataset shown in Table 1.

rule learning algorithms only expand a single successor node at a time, thereby learning a complete rule that covers part of the training data. After a complete rule has been learned, all examples that are covered by this rule are removed from the training set, and the procedure is repeated with the remaining examples (this strategy is also known as *separate-and-conquer* learning). Again, pruning is a good idea for rule learning, which means that the rules only need to cover examples that are *mostly* from the same class. It turned out to be advantageous to prune rules immediately after they have been learned, i.e., before successive rules are learned (Fürnkranz, 1997).

3. Motivation and Data

The transition to adulthood is one of the areas in the sphere of life course events where present-day European countries exhibit a high behavioural heterogeneity (Corijn, 1999; Kiernan, 1999; Billari et al., 2001; Corijn and Klijzing, 2001). In some countries events are experienced at an early age, while they are experienced at much later ages in others. The sequencing of events is also very different, as is sometimes the quantum. These differences, which are linked to cultural and historical patterns, present opportunity structures, and institutional arrangements, are even more clearly visible if one considers neighbouring countries. In this paper, we focus on Austria and Italy. The choice of these two countries is justified by the different patterns of transition to adulthood they exhibit – this provides us with a clear benchmark, and some prior knowledge, with which we confront the method. In Austria, the duration of education is quite standardized, and the vocational training system allows for a potentially smooth transition from school to work. Furthermore, leaving home occurs to a great extent before marriage, and there is a traditionally high share of births outside of cohabiting (married or unmarried) unions. In Italy, the duration of formal education and the entry into the labour market are experienced in a rather heterogeneous way. Leaving home occurs at a late age – the latest age observed among Western countries for which data are available. Leaving home is highly synchronized with marriage, and it is not common to leave home before finishing education. Finally, childbearing outside of marriage is still less common than in other European countries.[5]

3.1. DATA SOURCE

The data for our analysis originate from the Austrian and Italian Fertility and Family Surveys (FFS), which were conducted between December 1995 and May 1996 in Austria and between December 1995 and January 1996 in

Italy. Both surveys were part of a large-scale comparative program co-ordinated by the Economic Commission for Europe of the United Nations. The survey design provided independent samples of men and women in both countries. In Austria, 4,581 women and 1,539 men were interviewed, in Italy 4,824 women and 1,206 men. In Austria respondents were selected from the population aged 20–54, while in Italy the age range was 20–50. Hence, the Austrian FFS covers birth cohorts from 1941 to 1976, while the Italian FFS only includes cohorts from 1946 to 1976. To avoid differences due to sampling design, we opted to restrict the Austrian data set to the same cohorts as covered in the Italian survey. Furthermore, we excluded records with missing or incorrect values for the timing of events considered in our analysis. The final dataset contained 11,107 individuals (*examples* in machine learning terms), 5,325 of which were of Austrian and 5,782 of Italian origin.

In the FFS, retrospective histories of partnerships, births, employment, and education (in a more or less complete fashion) were collected on a monthly time scale, which allows us to study the timing, sequencing, and quantum of events in the transition to adulthood. We analyse the timing and quantum of leaving formal education, entering the first job, leaving the parental home, entering first union, entering first marriage, and having a first child, together with their pairwise sequencing. If individuals have not experienced an event, we consider in our analyses variables explicitly indicating that they are right-censored on this event. Such variables are used as information concerning the quantum of the event.

Two peculiarities of the data need to be mentioned. First, the Austrian FFS only allows one to know when the respondent left home for the last time before the interview, while the Italian FFS explicitly asked when the respondent left home for the first time. This difference should not, however, be a great problem in our comparative analysis: we will be more conservative in comparisons and underestimate differences, as Austrians leave home much earlier than Italians do anyway. Second, several problems arise when one wishes to compare educational histories across countries even using FFS comparative surveys (Dourleijn et al., 2002). For instance, a considerable number of respondents in Austria (1,639 out of a total of 6,020 respondents) have not indicated any educational level beyond ''Pflichtschule'', which is completed at age 15 in Austria. Although education was mandatory until the age of 14 for the Italian cohorts taken into account here (which is already a significant difference from Austria), a significant number of individuals dropped out before 14. Hence, we should expect institutional and drop-outs differences in the timing of education to show up as an important attribute for differentiating between the Austrian and Italian pathways in the transition to adulthood. We will discuss this further in Section 4.3.

3.2. ENCODING OF THE DATA

The key problem that one has to solve in order to make conventional data mining algorithms applicable to this type of data is how to encode the time-sequential nature of the data in a single data table that can be analysed with such algorithms. The approach we propose here is to make the information about timing, sequencing, and quantum explicit by encoding them as separate variables in a data table.

For this particular problem, we encoded the information as is shown in Table 2. We used four general descriptors related to sex, age, and birth cohort (with two potentially different categorizations for cohorts). Binary variables are employed to indicate whether each of the six events used to characterize the transition to adulthood has occurred up to the time of the interview (quantum), similarly to what is done in event history analysis for event/censoring indicators. If an event has occurred, the corresponding timing variable contains the age at which the person experienced the event (computed in years between the birth date and the date at which the event occurred). Finally, to make sequence information accessible to the learning algorithms, we performed pairwise comparisons between the dates at which two events occurred. The sequencing relationship, including synchronization (one date can be smaller or greater than or equal to the other), is encoded as a separate variable.[6] If both events have not occurred, we encode this with a designated value "n.o.". In the case that only one of the two events has occurred at the time of interview, we assume that the one that has occurred occurs earlier, even though the other event might not occur at all in this person's life course. As the time unit for computing sequencing and synchronization between two events, we use the month. That is, we use all the information available in the dataset and place a specific emphasis on events that are truly synchronized.[7]

Note that this encoding methodology is not specific to machine learning algorithms. It would also allow us to use other techniques, both from machine learning and statistics. Our choice to use decision trees and classification rules was made by their ability to produce directly interpretable results by dynamically controlling the complexity of the induced models.

4. Results

4.1. EXPERIMENTAL SETUP

We applied decision tree and rule learning algorithms to the dataset described in the previous section in order to detect the key features which distinguish between Austrians and Italians with respect to the timing, sequencing, and quantum of events in the transition to adulthood. We

Table 2. Variables used in the experiments

General descriptors	
Sex	Female, male
Birth cohort (5 years)	1946–1950, 1951–1955, 1956–1960, 1961–1965, 1966–1970, 1971–1975
Birth cohort (10 years)	1946-1955, 1956–1965, 1966–1975
Age	Age at interview in years
Quantum	
Education finished?	Yes, no
Had job?	Yes, no
Left home?	Yes, no
Formed union?	Yes, no
Married?	Yes, no
Had child?	Yes, no
Timing	
Education	Age at end of education
First job	Age at first job
Left home	Age at leaving home
Union	Age at first union
Marriage	Age at first marriage
Children	Age at the birth of first child
Age is measured in years	
Sequencing	
Education/job	<, >, =, n.o.
Education/left home	<, >, =, n.o.
Education/union	<, >, =, n.o.
Education/marriage	<, >, =, n.o.
Education/children	<, >, =, n.o.
First job/left home	<, >, =, n.o.
First job/union	<, >, =, n.o.
First job/marriage	<, >, =, n.o.
First job/children	<, >, =, n.o.
Left home/union	<, >, =, n.o.
Left home/marriage	<, >, =, n.o.
Left home/children	<, >, =, n.o.
Union/marriage	<, >, =, n.o.
Union/children	<, >, =, n.o.
Marriage/children	<, >, =, n.o.
For each possible combination of timing variables, their relative order is computed, or "n.o." is used if both events have not yet (i.e., before the interview date) occurred	

chose the decision tree learning algorithm C4.5 (Quinlan, 1993) and the rule learning algorithm Ripper (Cohen, 1995). Both algorithms are among the most prominent ones in machine learning, and they are frequently used both in applications and as benchmarks for new algorithmic developments. Their popularity is also due to their wide availability.[8] In addition to the functionalities described in Section 2, both algorithms are able to handle numerical attributes. C4.5 does this by testing the condition $x_i \leq v$ for each possible value v of the attribute x_i and computing the information that is gained by partitioning the data according to the outcome (*true* or *false*) of this test.[9] This can then be directly compared to the information gain values computed for the categorical attributes. The procedure for Ripper is quite similar.

Note that one cannot simply estimate the error rate of the induced model by testing it on the training data because, with increasing model complexity, the algorithms can fit the training data arbitrarily well. However, this fit cannot be expected to hold for new data – the problem of *overfitting* we already mentioned. For this reason, we use 10-fold cross-validation (Stone, 1974; Kohavi, 1995) for estimating the error rates of the learned models. This means that 10 experiments are performed, and in each experiment (each *fold*) a tenth of the data is held out, and a model is learned on the remaining nine-tenths. The resulting model is then tested on the remaining tenth of data. This is repeated 10 times, each time withholding a different tenth of the data. The results measured on these ten disjoint test sets are averaged. Note that the models learned in each of these ten experiments will be slightly different from each other, and also different from the model that has been learned from the entire data. Nevertheless – as the same, deterministic procedure is used for generating these models – their (measured) error rate estimates can be used for approximating the unknown error rates of the tree grown from the full data set.

Our cross-validation folds were *paired* (i.e., the same 10-folds were used for computing the performance estimates of both algorithms, which reduces random fluctuations) and *stratified* (i.e., the number of examples of each class in each fold was fixed in order to maintain the distribution of Austrians and Italians of the original set as closely as possible).

4.2. QUANTITATIVE RESULTS WITH DIFFERENT SPECIFICATIONS

In order to determine the relative importance of the quantum, timing, and sequencing of events that characterize the transition to adulthood, we performed a series of experiments in which we used different subsets of the available features. Each line of Table 3 shows the achieved performance of one particular feature subset. The first column describes which feature subset is used, the second and third columns show the performance estimates for

Table 3. Error rates (in %) and average size (no. of conditions) for C4.5 and Ripper on different problem representations (estimated by paired 10-fold cross-validations)

Feature set	C4.5		Ripper	
	Error	Size	Error	Size
Only general descriptors	45.97	42.1	46.83	15.7
Quantum	33.44	117.4	33.99	38.9
Timing	20.52	778.2	19.15	154.4
Sequencing	17.96	265.7	18.40	48.2
Quantum & timing	19.40	751.7	18.62	188.8
Quantum & sequencing	17.37	267.0	17.94	42.1
Timing & sequencing	15.14	584.9	15.18	116.4
All features	15.05	533.7	14.94	99.4

C4.5 and Ripper, respectively. For both algorithms, we show both the error rate in percentage points and the average size of the model learned (measured by the number of nodes in the decision tree or the number of conditions in the rule set, respectively).[10]

The first line shows the results from using only the four general descriptors shown at the top of Table 2 and none of the quantum, timing or sequencing variables. On this data set, both algorithms achieve error rates that are only slightly better than the error rate of uniformly predicting that all examples belong to the majority class, which has an error rate of 47.94%. These values are included as a benchmark, and in order to check that the relevant information for satisfactory classification performance is captured by the other variables. The next three lines show the results from using each subset independently, i.e., using only quantum, only timing, or only sequencing information. Among these three, sequencing proves to be most important. Using only sequencing information, both learning algorithms are able to discriminate the life courses of Italians and Austrians with an error rate of about 18%. Quantum information – in the simple encoding that shows only the occurrence or non-occurrence of an event – seems to be the least important. These results are confirmed by the next three lines, which show the performance of each pair of variables. The pair quantum & timing – the only one that does not use sequencing information – produces the worst results, while the timing & sequencing pair, which does not use quantum information, performs best. It should be noted, however, that quantum information is still of importance, as can be seen from the last line, which shows the results obtained using all variables shown in Table 2. Adding quantum to the timing and sequencing information further reduces the error rate (although this decrease is not statistically significant) and also results in simpler models.

In general, the error rates achieved by the two algorithms are about equal. Ripper seems to be a little better in handling timing variables, which may be due to minor differences in the handling of numerical attributes in the two algorithms. But the evidence from this dataset is insufficient to confirm this hypothesis, and there is no systematic comparison along this dimension in the literature. One can say, however, that rule sets are considerably simpler, which also means that they are easier to interpret.

The rule model that uses all features still has about 100 conditions and C4.5's decision tree has more than 500 nodes. Rule sets and trees of that size are very hard to interpret as a whole – one would have to focus on the important aspects (e.g., the areas near the root of the tree). Alternatively, one can make use of the pruning mechanisms of the algorithms to find a reasonable trade-off between simplicity and accuracy.

Both algorithms have a parameter that controls the pruning level of the algorithm. The exact details of the pruning procedures are somewhat different in the two algorithms[11] but both pruning parameters have the purpose of controlling the size of the models learned. This can be used for finding a model size that minimizes the estimated error rate as well as for increasing the comprehensibility of the models learned.

Figure 5 shows the error and complexity curves for C4.5. It was used with the parameter settings described at the beginning of the next section except

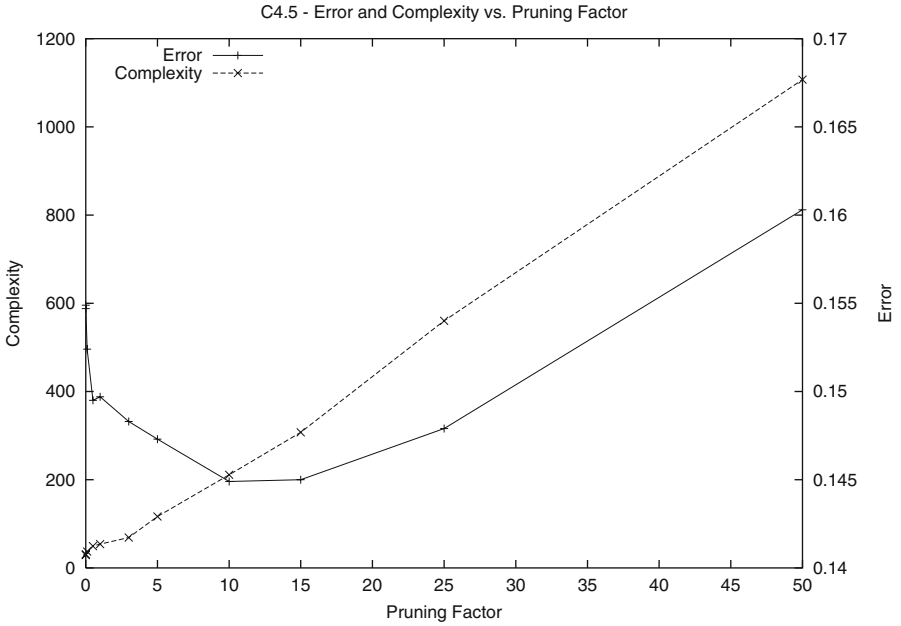

Figure 5. Error rates and model complexity for different settings of C4.5's pruning parameter.

for the pruning parameter, which was varied. The measured error rates and model sizes are plotted over the different values of the pruning parameter. The graph clearly exhibits the typical U-shape of error curves: overly complex models may overfit the data while overly simple models will fail to capture some important regularities in the data. Consequently, the best achievable error rate (about 14.45%) is somewhere in the middle, around 10–15 in our case (but this range may be very different for other datasets). C4.5's default value (25), which has been shown to perform reasonably well over a variety of datasets, is close to this. Systematically varying Ripper's pruning parameter led to similar results.

4.3. QUALITATIVE INTERPRETATION OF RESULTS

In this section we analyse the results of two models, one decision tree and one rule set. These models were obtained by using parameter settings that produce more comprehensible models, which are not necessarily optimal. In particular, we allowed C4.5 to combine branches with different attribute values (in default mode it will generate one branch for each possible outcome of a test, i.e., for each possible attribute value), and we specified that Ripper has to learn a rule set for each class (not only for the minor class, which classifies everything else by a default rule). We chose rather aggressive settings for the pruning parameter (0.001 for C4.5), which do not produce models with maximal accuracy. A quick inspection of the graphs in Figure 5 shows that the optimal tree size would be around 200 nodes, which is too big for an exhaustive interpretation. It should be noted, however, that the differences in accuracy are not that large and, more importantly, that the more accurate, complex models differ from the less accurate, simpler models only in the lower parts of a tree. The upper part, i.e., the choice of the most significant variables, is the same. It is not quite as simple for rule sets, but the starting conditions of the rules are typically more significant and are identical in simpler and more complex models.

Moreover, we will often take some liberties in interpreting the found models, and not strictly adhere to the automatically induced trees or rule sets. Most notably, we will primarily focus on the top levels of the tree and ignore some of the lower levels. This is justified, as we can often observe parts in the tree where a certain split appears to be irrelevant. For example, in Figure 6, the third-level split on *education*/*left home* on the very left, which produces an empty successor node, and a node with three Austrians in a set of examples that is dominated by Italian examples, does not appear to be justified. Ignoring this particular split will not have a big impact on the estimated accuracy of the entire tree. However, this cannot be estimated by cross-

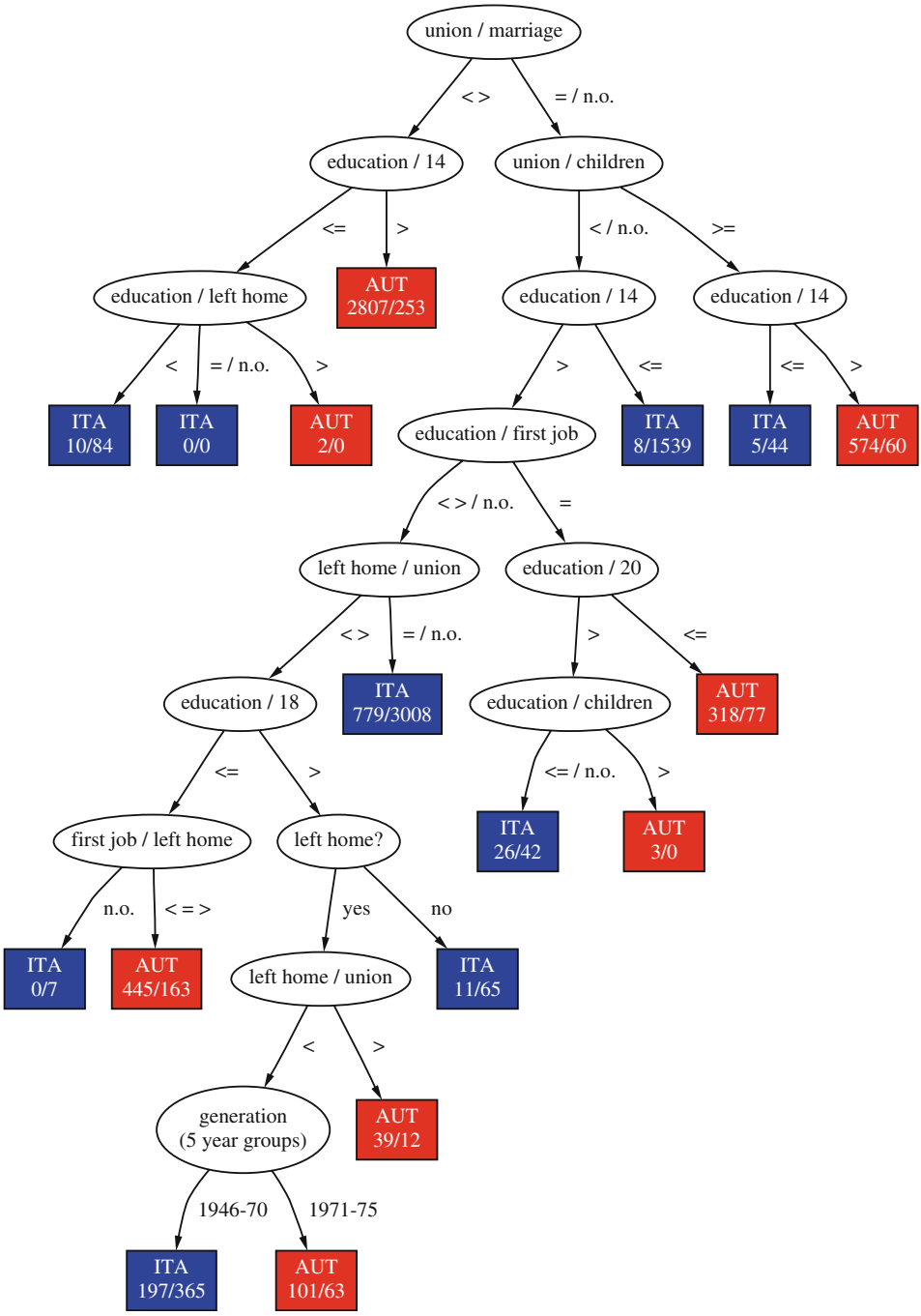

Figure 6. A decision tree applied to data on the transition to adulthood in Austria and Italy. The numbers indicated in each leaf are always ordered such that the first number refers to Austrians and the second number refers to Italians that are covered by this rule.

validation because this procedure depends on the fact that the desired tree can be produced with a certain parameterization of the algorithm, which is only possible as a general trend (smaller vs. bigger tree), but not to the fine details of every individual branch.

Figure 6 presents a simplified tree that uses only 32 nodes. Its estimated error rate is about 15.5%. Its most important attribute is the sequencing of union formation and marriage. Following the right branch that originates at this node and summing over all entries in each leaf implies that 5,445 Italians vs. 2,506 Austrians are covered by this path. From these results we can already deduce a first comparative proposition:

> An important characteristic that identifies the Italian pattern of transition to adulthood vs. the Austrian one is the fact that union formation and marriage are more likely to be synchronized. That is, Italians are much more prone than Austrians to marry directly rather than to start living together before marriage.

This result is in accordance with the literature for the cohorts under study. It only involves the interaction of a single independent variable with the dependent variable and could thus be found with straight-forward contingency analysis. Nevertheless, in the following, as we move deeper down the tree, we will encounter rules that are conditioned on multiple attributes. Such multi-variable dependencies could not be found with conventional tables, which illustrates the strengths of tree and rule-based methods.

If there is no synchronization of marriage and union formation (i.e., if we follow the left branch that originates at the first node), the age at which education is finished (i.e., the timing of the end of education) is chosen as the next most important attribute. Those who finished their education after the age of 14 are then most likely Austrians (2,807 Austrians vs. 253 Italians) – this is to be connected with the different institutional setting concerning compulsory education. If education is finished before the age of 14, a further attribute is added that compares the sequencing and quantum of education and leaving home. However, the total number of cases captured by this branch is negligible, and it might in part be due to the differences in the educational system of Austria and Italy (see Section 3). We will return to this issue later in this section.

If we follow the right branch of the decision tree, that is, in the case either of synchronization of marriage and union formation or of no experience of these events, the discrimination rules are not as straightforward. However, adding the birth of the first child as the third event and comparing the sequencing and quantum of the date of union formation and the birth of the first child essentially helps to distinguish between Italians and Austrians, consistently with the literature stressing the presence of out-of-union child-

bearing in Austria. Following the right-hand branch starting at the second node we are led to a second proposition:

> If the date of union formation and marriage coincides (or if neither event has yet occurred) Austrians are more likely than Italians to have had a child before this union.

Though the timing of education is added as a further attribute to this branch, the number of cases included in the final leaf that identifies Italians is negligible.

The classification becomes more complicated if neither event (union formation nor birth of first child) has occurred or if union formation precedes the birth of the first child. This branch includes 5,341 Italians and 1,927 Austrians. The timing of education becomes again important (as we expected) for further discriminating between Austrians and Italians. Those who completed their education before the age of 15 are most likely Italians (1,539 Italians vs. 8 Austrians). Nevertheless, 3,802 Italians and 1,919 Austrians are not yet distinguished in the case that education is completed after the age of 15. A fourth attribute, which refers to sequencing and quantum of the end of education and start of the first job, adds further information. Those for whom the end of education and the start of the first job are synchronized are most likely Austrians (119 Italians vs. 347 Austrians) – indicating an easier school-to-work transition in Austria. If the sequencing between end of education and start of the first job is indeterminate, or if neither event has yet occurred, a fifth attribute is added: the age at leaving home. Those for whom leaving home is synchronized with first union formation, or for whom neither event has yet occurred, are most likely Italians (3,008 Italians vs. 779 Austrians) – we know that such synchronization is typical of the Mediterranean pattern of transition to adulthood (Billari et al., 2002). If leaving home and first union formation are not synchronized (or their sequencing is not yet known), a more complicated decision rule is proposed, which includes the timing of education once again. It also adds the sequencing and quantum of leaving home, union formation, and the start of the first job. Finally, it includes information about cohort membership at the final node. These results lead us to suggest a third proposition:

> In the case of more traditional patterns in the transition to adulthood, with first union formation synchronized with first marriage (or neither of these two events has yet occurred), and with the birth of the first child – if it has occurred – experienced after union formation, the length of education and the sequencing of the end of education and the first job are two further important discriminating factors.

In particular, those who finished education before the age of 15 are, again, most likely Italians. Among those who finished education after the age of 15, Austrians are discerned from Italians by the fact that the end of education and the first job are synchronized or that leaving home and union formation are synchronized (or have not yet been experienced).

As already indicated in Section 3, the timing of education in Austria and Italy reflects institutional differences in the educational system. To test the importance of such institutional differences (as opposed to less formalized differences in the transition to adulthood) we applied the decision tree algorithm to the same data set, except that we excluded the length of education as an attribute. The resulting decision tree, with an error rate that is 2.2% higher than the preceding one, is presented in Figure 7. A comparison of both trees (with and without the length of education) supports our first two propositions. The sequencing and quantum of the events union formation, marriage, and birth of first child are major attributes for distinguishing between Austrians and Italians, as is the synchronization of leaving home and first union. If we follow the first two rightmost branches in the decision tree in Figure 7, the timing of events is added as a further attribute. The length of education is now replaced by the age at the start of the first job as the most important attribute – both events are likely to be close if the labour market is entered at all. In comparison to educational trajectories, however, job histories are less strictly regulated by institutional settings.

To obtain a more compact representation of rules that best discriminate between Austrians and Italians we apply the rule learning algorithm Ripper to the same dataset. Again, we used a setting that favoured simplicity over accuracy in order to optimise comprehensibility. The resulting rule set is shown in Figure 8. Its error rate is about 16.5%. One of the rules is the sequencing and synchronization between leaving home and marriage. They are synchronized for 2,851 Italians as compared to 592 Austrians, while leaving home precedes marriage for 3,476 Austrians vs. 976 Italians. The fact that the rule algorithm chooses leaving home and marriage as opposed to union and marriage (which was chosen as the most important attribute in the decision tree algorithm) is not a contradiction. To understand it, we need to recall how both algorithms are designed.

The basic difference between both approaches is that the decision tree approach evaluates attributes according to their average discriminatory power over *all possible outcomes* of an attribute. In our case, conditioning on each possible outcome of the attribute *union/marriage* ($<$, $>$, $=$, n.o.) skews the distribution of Austrians and Italians towards one or the other of the two groups, thereby improving the overall discrimination between these two groups.

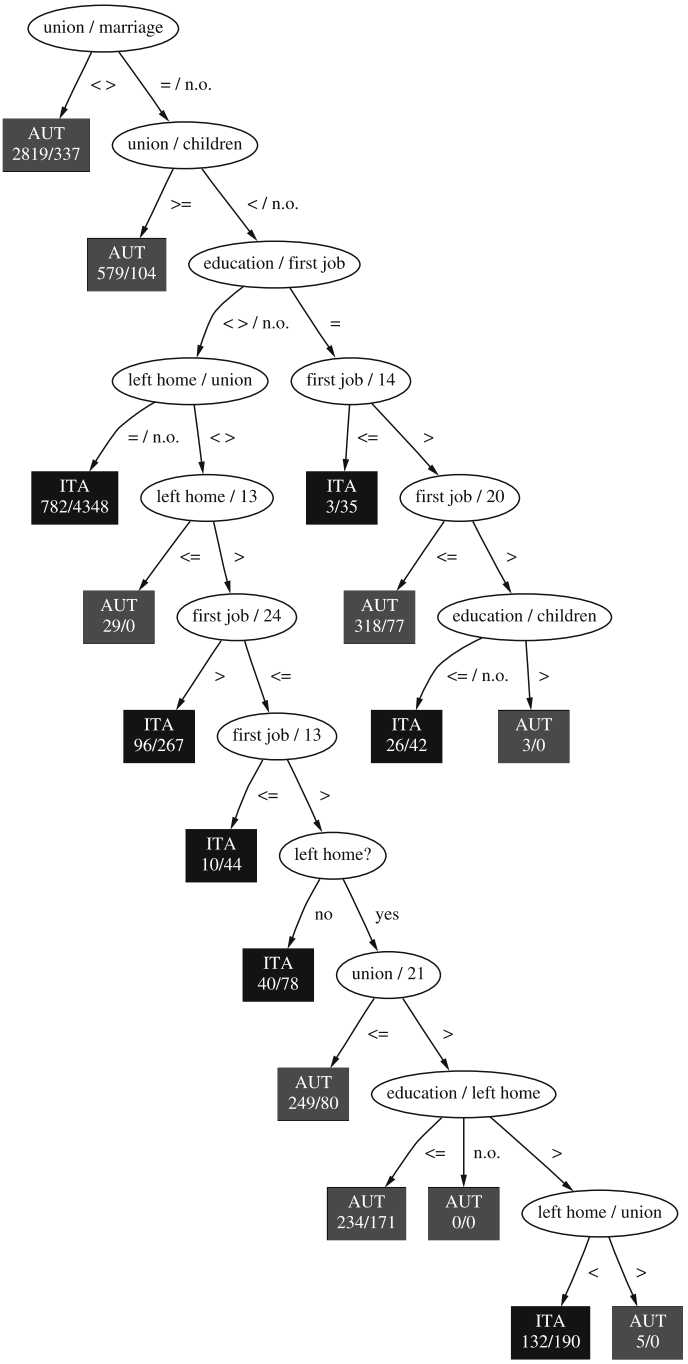

Figure 7. A decision tree applied to data on the transition to adulthood in Austria and Italy excluding the timing of education as an attribute. The numbers indicated in each leaf are always ordered such that the first number refers to Austrians and the second to Italians covered by this rule.

IF left home = marriage
THEN *ITA* (592/2851)
IF left home = n.o. **AND** union = n.o. **AND** child = n.o.
THEN *ITA* (465/1692)
IF union = marriage **AND** education ≤ 14
THEN *ITA* (9/1308)
IF education 22 **AND** union = marriage **AND** union ≥ 24
THEN *ITA* (64/541)

IF left home < marriage
THEN *AUT* (3476/976)
IF left home > marriage **AND** left home? = yes **AND** education ≥ 15
THEN *AUT* (533/46)
IF education ≥ 15 **AND** rst job ≤ 18 **AND** education ≤ 18 **AND** union ≤ 21
THEN *AUT* (1468/215)

DEFAULT *AUT* (197/105)

Figure 8. A rule set describing the data on the transition to adulthood in Austria and Italy. The numbers indicated at the end of each rule are always ordered such that the first number refers to Austrians and the second to Italians covered by the rule.

On the other hand, the rule learning algorithm focuses on *particular values* of the attribute that are most indicative of a given target group. In our case, the synchronicity of leaving home and marriage is determined to be a very good indicator for Italians, much more so as the synchronicity of union formation and marriage, which is used later in the rule set, and only in conjunction with other conditions. Thus, this condition is selected for starting the rule set, although as a whole, *union/marriage* may result in a better discrimination over *all* its possible outcomes.

Besides the important role of synchronization of events like leaving home, marriage, and union formation, Italians also differ from Austrians in that more of them have not yet (i.e., until the interview date) experienced either of these events. As the second rule states, those who have not yet left home, not yet started a union, and not yet had a child are more likely Italians. This is in accordance with the idea that the Italian pattern of transition to adulthood is a "latest late" one (Billari et al., 2002). Austrians experience many of these events at much earlier ages, as is best represented by the last rule.

For the sake of completeness we have added the results of the rule-based algorithm as applied to the data set where we exclude the timing of education as an attribute (see Figure 9).

IF home = marriage
THEN *ITA* (592/2851)
IF home = n.o. **AND** child = n.o.
THEN *ITA* (529/1728)
IF union = marriage **AND** job ≥ 22
THEN *ITA* (215/1419)

IF home < marriage
THEN *AUT* (3476/976)
IF home > marriage **AND** home = yes
THEN *AUT* (534/80)
IF job < 20 **AND** education = job **AND** job ≥ 17
THEN *AUT* (824/57)
IF home > union **AND** union < marriage
THEN *AUT* (458/45)

DEFAULT *AUT* (99/79)

Figure 9. A rule set describing the data on the transition to adulthood in Austria and Italy excluding the timing of education as an attribute. The numbers indicated at the end of each rule are always ordered such that the first number refers to Austrians and the second to Italians covered by this rule.

Though the models learned by the decision tree algorithm differ slightly from those learned by the rule learning algorithm, the results of both also support the proposition that *the sequencing of events in the transition to adulthood is more important than the timing and quantum of these events in order to discriminate between Austrians and Italians*. Our results may also be contrasted with the findings in Ravanera et al. (2004), who show that over recent cohorts in Canada age homogeneity decreases for family life events (like first union, first marriage, birth of first child, etc.) and increases for more "formal" life course transitions like education and work. Our result (Section 2) that family life events are more important factors to discriminate between Austrians and Italians may reflect a stronger heterogeneity as compared to more formal life course transitions, in line with results on Canadian cohorts.

5. Discussion and Perspectives

The methodological contribution of this paper is 2-fold: first, we illustrate how life course data can be analysed with conventional classification algorithms by explicitly and simultaneously encoding quantum, timing, and sequencing information. Second, we applied techniques developed in machine learning for the analysis of life course data from a comparative

perspective. These techniques allow for a high degree of flexibility in the use of data and for problem-specific representations of the available information.

The literature on the transition to adulthood acknowledges a key role to the timing, sequencing, and quantum of events such as leaving home, union formation, marriage, birth of the first child, completion of education, and start of the first job. To disentangle the complex relationship between the timing, sequencing, and quantum of events we proposed a novel representation of life course data that captures this information. We then applied machine learning techniques to data about the transition to adulthood from Austrian and Italian Fertility and Family Surveys. More specifically, we built decision trees and rule sets that allowed to shed light on the key differences in the patterns of transition to adulthood between Austria and Italy.

Our main theoretical result is that we have established the key role of the sequencing of events for distinguishing between Austrian and Italian patterns of transition to adulthood. Information on the timing of events could be regarded as the next best group of features, while information on the quantum of events turned out to be the least important feature set. In terms of the sequencing of events, our findings showed that the synchronization between events such as leaving home and first marriage and first marriage and union formation is the most important feature for distinguishing the pathways of Italian youth from those of Austrians. Information on the timing of events was only of importance as it regards the length of formal education and the age at the start of the first job and only in the case that the algorithm needed to distinguish between Austrians and Italians along more traditional life course patterns. Information on the quantum of events (i.e., whether an event has occurred) was mostly confined to highlighting a well-known characteristic of Italian patterns in the transition to adulthood, namely, the fact that Italians are often the latest-late as regards events such as leaving home, union formation, marriage, and birth of the first child. However, we should stress that by focusing on the transition to adulthood and on non-repeatable events, we have unavoidably underscored the importance of the quantum of events – we expect that in other potential applications to demographic life courses quantum may play a major role.

The adoption of the machine learning approach we proposed allows one to look at life courses from a holistic perspective. This perspective becomes even more important in the case of comparative studies, i.e., if one tries to differentiate between two groups of individuals. Moreover, the results of decision trees and rule sets provide a non-technical audience with a clear representation of results. This is not possible with the techniques currently in use (such as event history analysis), either because they do not start from a holistic perspective or because they do not give results which are clearly interpretable. We thus foresee many applications for the techniques discussed

here in life course research and more generally in demography and sociology. In general, such techniques can be applied to various kinds of comparative research. The comparison of cohorts, and gender comparisons within a society are also envisageable as potential areas of application. The public availability of software for these techniques is definitely a great advantage. Finally, the representation of life course events in terms of timing, sequencing, and quantum we have adopted can be regarded as a further novelty of this paper. This representation can possibly also be adopted in analyses using different techniques based on other statistical models.

Acknowledgements

The authors acknowledge support from the Max Planck Institute for Demographic Research, the Austrian Research Institute for Artificial Intelligence (supported by the Austrian Federal Ministry of Education, Science and Culture), the Vienna Institute of Demography, and Università Bocconi. The experimental work in this paper was greatly facilitated by a set of tools developed within the ESPRIT long-term research project METAL (project no. 26357). We wish to thank their author, Johann Petrak, for invaluable help in using these tools. We wish to thank two anonymous referees for precious suggestions, Karl Brehmer for comments, and the Advisory Group of the FFS programme of comparative research for its permission, granted under identification number 75, to use the FFS data on which this study is based.

Notes

[1] In the statistical literature (cf., e.g., Breiman et al., 1984), decision trees are also known as *classification trees*. Related techniques for predicting numerical class values are known as *regression trees*. Such techniques are also used for predictive purposes in survival analysis. An interesting application of survival trees (trees which assign a survival curve to each node) is described in De Rose and Pallara (1997).

[2] This preference for simple models is a heuristic criterion known as *Occam's Razor*, which appears to work fairly well in practice. It is often recalled in the statistical literature on model selection, but it is still the subject of ardent debates within the machine learning community (Domingos, 1999).

[3] Actually, C4.5 per default evaluates attributes with a slightly more complex measure called gain ratio, which normalises the gain with the goal of countering its tendency to prefer attributes with a larger set of possible values. Other attribute selection measures, which do not conform to gain framework laid out in the text, are also possible, such as CHAID's evaluation with a χ^2 test statistic (Kass, 1980).

[4] Typically, one uses a Laplace-corrected estimate for the prediction accuracy, i.e., $(p+1)/(p+n+2)$, if p is the number of covered examples that support the prediction of the rule and n is the number of covered examples that contradict it. This has the effect that rules that cover only few examples will be corrected towards a prior probability of 1/2.

[5] Extended descriptions and general analyses of the transition to adulthood in Austria and Italy are provided in Nowak and Pfeiffer (1998) and in Billari (2000), respectively.

[6] This approach to making sequence information available to the learner – encoding the additional relations in derived variables – is loosely based on the Linus approach to relational learning (cf., e.g., Lavrač et al., 1993, where learning performance was improved in a medical application by augmenting patient data with additional domain-specific background knowledge that highlighted characteristic combinations of the original measurements).

[7] Other scholars argue that, since decision-making occurs on a fuzzy time scale, larger intervals for synchronization should be considered, e.g., a yearly interval (Courgeau and Lelièvre, 1992). The method we propose would also allow for different synchronization intervals.

[8] C4.5 can be obtained by buying the companion book (Quinlan, 1993), or it can be downloaded for research purposes at http://www.cse.unsw.edu.au/quinlan/. C5.0, its commercial successor, is available from http://www.rulequest.com/. Ripper is available upon request at <http://www.research.att.com/diane/ripper.html>.

[9] The algorithms actually implement a more efficient version of this technique, which sorts the values first and tests only values whose successor has a different class value. It can be shown that this procedure produces the same result as testing all values (Fayyad and Irani, 1992).

[10] In contrast to the error rate, the size of the models learned could have been measured directly by using the entire set of examples for training. However, we report the average model sizes in the 10-folds of the cross-validation procedure, which came for free as a by-product of the cross-validation procedure. The size of the models learned from the complete training set might be somewhat larger, but their relative order can be expected to be the same.

[11] C4.5 employs *error-based pruning*, which uses a heuristic estimate of the confidence intervals for the accuracy of the class probability estimates at each node (Quinlan, 1993). Ripper's pruning is based on the *incremental reduced error pruning technique* (Fürnkranz, 1997), which internally splits the available training data into a learning set and a pruning set and uses the latter to fine-tune the rules learned on the learning set.

References

Billari, F. C., 2000. *L'analisi delle biografie e la transizione allo stato adulto. Aspetti metodologici e applicazioni ai dati della Seconda Indagine sulla Fecondità in Italia.* Cleup Editrice, Padova.

Billari, F. C., 2003. 'Life course analysis', in P. Demeny and G. McNicoll (eds), *Encyclopedia of Population* Vol. 2. New York: Macmillan Reference USA, 588–590.

Billari, F. C., Castiglioni, M., Castro Martin, T., Michielin, F. and Ongaro, F., 2002. 'Household and union formation in a Mediterranean Fashion: Italy and Spain', in E. Klijzing and M. Corijn (eds), *Fertility and Partnership in Europe: Findings and Lessons from Comparative Research* Vol. 2. New York/Geneva: United Nations, 17–41.

Billari, F. C., Philipov, D. and Baizán, P., 2001. Leaving home in Europe: the experience of cohorts born around 1960, *International Journal of Population Geography* 7: 311–338.

Billari, F. C. and Piccarreta, R., 2005. Analysing demographic life courses through sequence analysis, *Mathematical Population Studies* 12(2): 81–106.

Breiman, L., 2001. Statistical Modeling: The Two Cultures, *Statistical Science* 16: 199–215.

Breiman, L., Friedman, J., Olshen, R. and Stone, C., 1984. *Classification and Regression Trees.* Wadsworth & Brooks, Pacific Grove, CA.

Cohen, W. W., 1995. 'Fast effective rule induction', in A. Prieditis and S. Russell (eds), *Proceedings of the 12th International Conference on Machine Learning (ML-95).* Lake Tahoe, CA: Morgan Kaufmann, 115–123.

Corijn, M., 1999. Transitions to adulthood in Europe for the 1950s and 1970s cohorts, *CBGS-Werkdocument*, Vol. 4, Brussels.

Corijn M. and Klijzing E. (eds.), 2001. *Transitions to Adulthood in Europe*. Kluwer Academic Publishers, Dordrecht.

Courgeau, D. and Lelièvre, É., 1992. *Event History Analysis in Demography*. Clarendon Press, Oxford.

De Rose, A. and Pallara, A., 1997. Survival trees: an alternative non-parametric multivariate technique for life history analysis, *European Journal of Population* 13: 223–241.

Domingos, P., 1999. The role of Occam's Razor in knowledge discovery, *Data Mining and Knowledge Discovery* 3(4): 409–425.

Dourleijn, E., Liefbroer, A. C. and Beets, G. C. N., 2002. 'Comparing the 1988 International Standard Classification of Education (ISCED) with retrospective information from educational histories', in E. Klijzing and M. Corijn (eds), *Fertility and Partnership in Europe: Findings and Lessons from Comparative Research* Vol. 2. New York/Geneva: United Nations, 157–172.

Fayyad, U. M. and Irani, K. B., 1992. 'On the handling of continuous-valued attributes in decision tree induction', *Machine Learning* 8: 87–102 (Technical Note).

Fayyad U. M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. (eds.), 1995. *Advances in Knowledge Discovery and Data Mining*. AAAI Press, Menlo Park.

Friedman, J. H. and Fisher, N. I., 1999. Bump hunting in high-dimensional data, *Statistics and Computing Archive* 9(2): 123–143.

Fürnkranz, J., 1997. Pruning algorithms for rule learning, *Machine Learning* 27(2): 139–171.

Fürnkranz, J., 1999. Separate-and-conquer rule learning, *Artificial Intelligence Review* 13(1): 3–54.

Hastie, T., Tibshirani, R. and Friedman, J. H., 2001. *The Elements of Statistical Learning*. Springer-Verlag, New York.

Hogan, D. P., 1978. The variable order of events in the life course, *American Sociological Review* 43: 573–586.

Kass, G. V., 1980. An exploratory technique for categorical data, *Applied Statistics* 29: 119–127.

Kiernan, K., 1999. Childbearing outside marriage in Western Europe, *Population Trends* 98: 11–20.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection, in *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95)*. Montreal, Canada: Morgan Kaufmann, 1137–1143.

Lavrač, N., Džeroski, S., Pirnat, V. and Križman, V., 1993. The utility of background knowledge in learning medical diagnostic rules, *Applied Artificial Intelligence* 7: 273–293.

Lillard, L. A., 1993. Simultaneous equations for hazards. Marriage duration and fertility timing, *Journal of Econometrics* 56: 189–217.

Marini, M. M., 1987. Measuring the process of role change during the transition to adulthood, *Social Science Research* 16: 1–38.

Michalski, R. S., Bratko, I., Kubat M. (eds), 1998. *Machine Learning and Data Mining: Methods and Applications*. John Wiley & Sons.

Mitchell, T.M., 1997. *Machine Learning*, McGraw-Hill.

Modell, J., Furstenberg, F. F. Jr. and Hershberg, T., 1976. Social change and transitions to adulthood in historical perspective, *Journal of Family History* 38: 7–32.

Mulder, C. and Wagner, M., 1993. Migration and marriage in the life course: a method for studying synchronized events, *European Journal of Population* 9(1): 55–76.

Murthy, S. K., 1998. Automatic construction of decision trees from data: a multi-disciplinary survey, *Data Mining and Knowledge Discovery* 2(4): 345–389.

Nowak, V., Pfeiffer, C., 1998. Transition into Adulthood, *Working Paper 8*. Austrian Institute for Family Studies.

Quinlan, J. R., 1986. Induction of decision trees, *Machine Learning* 1: 81–106.

Quinlan, J. R., 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.

Ravanera, Z. R., Rajulton, F. and Burch, T. K., 2004. Patterns of age variability in life course transitions, *Canadian Journal of Sociology* 29(4): 527–542.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions, *Journal of the Royal Statistical Society B* 36: 111–147.

van Wissen L. J. G., and Dykstra P. A. (eds.), 1999. *Population Issues: An Interdisciplinary Focus*. Kluwer Academic/Plenum Publishers, New York.

Witten, I. H., Frank, E., 2000. *Data Mining – Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers.