A Multistrategy Approach to the Classification of Phases in Business Cycles

Katharina Morik, Stefan Rüping
Univ. Dortmund, Computer Science Department, LS VIII
morik@ls8.informatik.uni-dortmund.de

Abstract The classification of business cycles is currently performed using either macro-economic equations or linear discriminant analysis. It is a hard and important problem to classify in which economic phase we are in. Government as well as business decisions rely on the assessment of the current business cycle. In this paper, we investigate how economists can be better supported by a combination of machine learning techniques. We have successfully applied Inductive Logic Programming (ILP). The application of ILP requires pre-processing in order to establish time and value intervals. To this end, top-down induction of decision trees is used. The rule sets learned from different experiments were analysed with respect to correlations in order to find a concept drift or shift.

1 Introduction

The ups and downs of business activities have been observed since a long time [1]. It is, however, hard to capture the phenomenon by a clear definition. The National Bureau of Economic Research (NBER) defines business cycles as "*recurrent sequences of altering phases of expansion and contraction in the levels of a large number of economic and financial time series.*" This definition points at the multi-variate nature of business cycles. It does not specify many of the modeling decisions to be made. There is still room for a variety of concepts.

- What are the indices that form a phase of the cycle? Production, employment, sales, personal income, and transfer payments are valuable indicators for cyclic economic behavior. Are there others that should be included?
- What is the appropriate number of phases in a cycle? The number of phases in a cycle varies in economic models from two to nine. The NBER model indicates two alternating phases. The transition from one phase to the next is given by the turning points *trough* and *peak*. In the RWI model, a cycle consists of a lower turning point, an upswing, an upper turning point, and a downswing. Here, the turning points are phases that cover several months.
- Are all cycles following the same underlying rules or has there been a drift of the rules?

[1]Amstad reports the first definition from Clement Juglar in 1860 [3]. She investigates several models of the business cycle and discusses their distinctions with respect to dating turning points of the business cycle.

All modeling decisions are to be (comparatively) validated with respect to economic theory and to business data. One approach to validation is the formalization by macro-economic equations. A model of business activities is calculated *ex post* and the deviation of the results of the equations from the observed values assesses the model. For instance, the business cycle model of the Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI) only deviated 1.2 per cent for the spring 2000 state of affairs in Germany [5]. The main focus here lies on the prediction of level or growth of business activities. We do not contribute to this approach.

The other approach is an empirical one, in which statistical methods are adjusted to business data and used for prognoses. Again, the statistical models are validated on past data. We are concerned with the development and comparison of methods for the empirical modeling of business cycles. Empirical methods are particularly demanded for the task of dating turning points or phases of the business cycle. This task is less clearly defined than the task of predicting business activities, because business cycles themselves are basically a theoretical model to explain the variation in business data. In this paper, we tackle the dating problem:

Dating: Given current (and past) business measurements, in which phase is the economy currently? In other words, the current measurements are to be classified into the phases of a business cycle.

Linear discriminant analysis has been proposed as the baseline of empirical models [2]. Univariate rules were learned that used threshold values for separating phases. The accuracy of the 18 learned rules was 54% in cross validation. Using this result as the baseline means that the success of any other method has to be shown in comparison to this accuracy. It has been investigated how the classification can be enhanced by the use of monthly data [7]. More sophisticated statistical models have been developed and achieved 63% accuracy [11]. However, even this substantial enhancement still reflects how hard it is to classify business phases correctly.

In this paper, we investigate the applicability of inductive logic programming to the problem of dating phases of a business cycle. We were given quarterly data for 13 indicators concerning the German business cycle from 1955 to 1994, where each quarter had been classified as being a member of one of four phases [6]. The indicators are:

[2] Claus Weihs at a workshop on business cycles at the "Rheinisch-Westfälisches Institut für Wirtschaftsforschung" in January 2002

IE	real investment in equipment (growth rate)
C	real private consumption (growth rate)
Y	real gross national product (growth rate)
PC	consumer price index (growth rate)
PYD	real gross national product deflator (growth rate)
IC	real investment in construction (growth rate)
LC	unit labour cost (growth rate)
L	wage and salary earners (growth rate)
Mon1	money supply M1
RLD	real long term interest rate
RS	nominal short term interest rate
GD	government deficit
X	net exports

We experimented with different discretizations of the indicator values (see Section 2.1). The discretization into ranges (levels) of values was also used in order to form time intervals. A sequence of measurements within the same range is summarized into a time interval. Relations between the different time intervals express precedence or domination of one indicator's level to another ones level. We also compared the two phase with the four phase business cycle. In summary, the following three models were inspected:

- business cycle with four phases, without time intervals, (Section 2.2)
- business cycle with four phases, time intervals, (Section 2.3).
- business cycle with two phases, without time intervals (Section 2.4).

Particular attention was directed towards the appropriate sample size for the dating problem. The homogeneity of the data set of business cycles with two phases was investigated (Section 2.5).

2 Experiments on German Business Cycle Data

Our leading question was whether ILP can support economists in developing models for dating phases of the business cycle. Given the quarterly data for 13 indicators concerning the German business cycles from 1955 to 1994 where each quarter is classified as member of one of four phases, we used all but one cycle for learning rules and tested the rules on the left-out cycle. The leave-one-cycle-out test assesses the *accuracy* (how many of the predicted classifications of quarters corresponded to the given classification) and the *coverage* (how many of the quarters received a classification by the learned rules).

For ILP learning, we applied RDT [8] with the following rule schemata:

m1 (Index1, Value, Phase):
$Index1(T,V), Value(V) \rightarrow Phase(T)$

m2 (Index1,Value,Index2, Phase):
$Index1(T,V), Value(V), Index2(T,V) \rightarrow Phase(T)$

m3 (Index1, Value1,Index2,Value2,Phase):

$Index1(T, V1), Value1(V1), Index2(T, V2), Value2(V2), opposite(V1, V2) \rightarrow Phase(T)$

The predicates that fit to instantiate the predicate variable *Index* are the 13 indicators of the economy (see above). The predicates that fit to instantiate express the discretization of the real values of the indicators. The phase variable can be instantiated by *down*, *ltp*, *up*, *utp* for four phases or by *down*, *up* for two phases of the business cycle.

2.1 Discretization

Before ILP can be applied, the originally real-valued time series of indicator values have to be transferred into discrete-valued temporal facts about this indicators. The goal of discretization is to provide the learning algorithm with data from which it can generalize maximally. This means, the discretization must be general enough such that rules learned from one situation can be transferred to another situation but specific enough such that non-trivial rules can be found. An example for a too specific discretization is to assign different values to every observation, an example for a too general discretization is to assign the same value to every observation. We use the number of generated facts to judge the quality of a discretization.

Actually, the task of discretization consists of two different subtasks:

Discretization of Values: split the continuous range of possible values into finitely many discrete values, e.g. by using equidistant thresholds or calculating suitable quantiles. For example, a gross national product of 4.93 in the fifth quarter could be expressed as the fact $y(5, high)$.

Interval segmentation: for a given time series, find a segmentation of the time points into maximal sub-intervals, such that the values of the series in this interval share a common pattern, e.g. by approximating the time series by piecewise constant or piecewise linear functions. For example, the time series of gross national products $Y = (10.53, 10.10, 9.21, 5.17, 4.93)$ could be described as the temporal facts $y(1, 3, high), y(4, 5, medium)$, but can also be described as $y(1, 5, decreasing)$.

Interval segmentation can be viewed as discretization of the temporal values, therefore in this chapter we will use the name discretization as a generic term for both discretization of values and interval segmentation.

The two subtasks are closely intertwined: Discretized data can be very easily segmented by joining consecutive time points with identical discretization. Also, segmented data can be discretized by building a discretization based on the patterns that lead to the segmentation. In this work, we chose the first approach to discretize the data, first because it is simpler and secondly because the indicators are already given free of trends (growth rates etc.), so it can assumed the relevant information lies in the value of the indicator alone.

To improve the quality of the discretization, we can also use the information that is given by the class of the examples [13]. In this case, we used C4.5 [9],

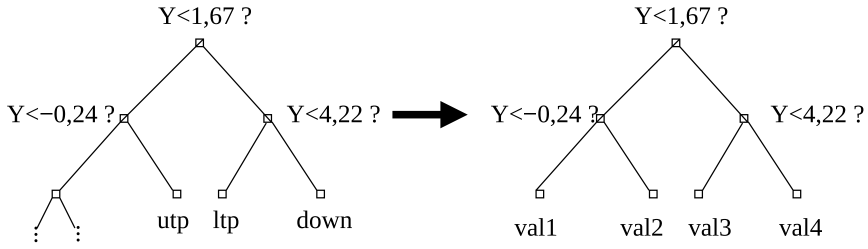

Figure 1: Decision tree and its induced discretization into $val1 \ldots val4$.

Cycle	Accuracy	Coverage	No.of learned rules
LOO1	0.125	0.25	13 upswing
LOO2	0.5	1.0	12 upswing
LOO3	0.462	0.462	10 upswing, 2 downswing
LOO4	0.375	1.0	11 upswing
LOO5	0.696	0.696	10 uspwsing, 1 downswing
LOO6	1.0	0.36	1 upswing
Average	0.526	0.628	total: 60

Figure 2: Results in the four phase model using time points

a decision tree learner, to induce decision trees about the cycle phase based on only one indicator. The resulting trees were cut off at a given level and the decisions in this resulting tree were used as discretization thresholds (see Figure 1). Decision trees of depth 2, i.e. using 4 discrete values, proved to build a suitable number of facts.

A closer look at the resulting discretization showed that in certain cases, the indicators had a very high variation, which leads to many intervals that contained only one time point. In this case, the relevant observation may not be the value of the indicator, but the fact that this indicator was highly variating, i.e. that no definite value can be assigned to it. This can be expressed by a new fact $indicator(T1, T2, unsteady)$, which replaces the facts $indicator(T1, T1 + 1, value_1), indicator(T1 + 1, T1 + 2, value_2), \ldots, indicator(T2 - 1, T2, value_n)$.

2.2 Modeling Four Phases Without Time Intervals

The data correspond to six complete business cycles, made of four phases each. For the upper and lower turning point phases, no rule could be learned. Only for the upswing, each learning run delivered rules. For the downswing, only two learning runs, namely leaving out cycle 3 and leaving out cycle 5, delivered rules. Misclassifications at the turning points are strikingly more frequent than in other phases. Figure 2 shows the results.

The results miss even the baseline of 54% in the average. Leaving out the fifth cycle (from 1974 until 1982) delivers the best result where both, accuracy

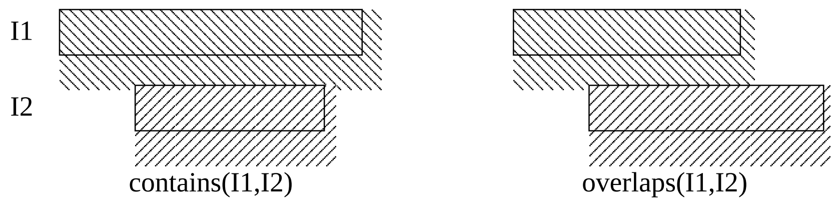

Figure 3: The temporal relations *contains* and *overlaps*

and coverage, happen to approach 70%. This might be due to its length (32 quarters), since also in the other experiment dealing with four phases the prediction of upper turning point and upswing is best, when leaving out the fifth cycle. Since the sixth cycle is even longer (45 quarters), we would expect best results in LOO6 which is true for the accuracy in this experiment. In the other experiment with four phases, the accuracy is best for upswing in LOO6 and second best for it in LOO5.

2.3 Modeling Four Phases With Time Intervals

Let us now see, whether time intervals can improve the learning results. We have used the discretization of the indicator values for the construction of time intervals. As long as the indicator value stays within the predefined level, the time interval is continued. As soon as the indicator value exhibits a level change, the current time interval is closed and the next one is started. We end up with facts of the form `Index(I,Range)`, and for each time point within the time interval I a fact stating that this time point T (i.e. quarter) lies in the time interval I: `covers(I, T)`.

We then described the relations between different time intervals by means of Allen's temporal logic [2]. From the 13 possible relationships between time intervals, we chose `contains` and `overlaps`. The relation $contains(I1, I2)$ denotes a larger interval $I1$ in which somewhere the interval $I2$ starts and ends. `contains(I1, I2)` is true for each time point within the larger interval $I1$. `overlaps(I1, I2)` is true for each time point of the interval $I1$ which starts before $I2$ is starting (see Figure 3). We left out the other possible relations, because they were either too general or too specific to be used in a classification rule or would violate the constraint, that only information about past events can be used in the classification [3]. The time intervals were calculated before the training started. The rule schemata were defined such that they link two indicators with there corresponding time intervals. One rule schema is more special in that it requires the time intervals of the two indicators to either overlap or include each other. This more specific rule schema was intended to find rules for the turning phases, where no rules were learned in the previous

[3]For example, a relation that would require that the end point of one interval was identical to the starting point of another interval would be too specific and a relation that would only require that an interval would happen before another interval, regardless of the amount of time in between, would be too general.

Cycle	Phase	Accuracy	Coverage	No. learned rules
LOO1	upswing	0.167	1	73
	downswing	-	0	1
	utp	-	0	0
	ltp	-	0	2
LOO2	upswing	-	0	103
	downswing	-	0	3
	utp	-	0	2
	ltp	-	0	0
LOO3	upswing	0.461	1	87
	downswing	1	0.200	2
	utp	0	0	2
	ltp	-	0	2
LOO4	upswing	0.167	1	59
	downswing	0.333	1	7
	utp	-	0	0
	ltp	-	0	4
LOO5	upswing	0.481	1	88
	downswing	0	0	3
	utp	-	0	0
	ltp	0.75	0.857	4
LOO6	upswing	0.667	0.296	6
	downswing	0.243	1	2
	utp	-	0	0
	ltp	-	0	0
Average	upswing	0.388	0,716	69.3
	downswing	0.194	0.500	3
	utp	0	0	0.667
	ltp	0.75	0.143	2

Figure 4: Results in the four phase model using time intervals

experiment. In fact, rules for the upper turning point, upswing, and downswing were learned, but no rules could be learned for the upper turning point.

Another intention behind the time interval modeling was to increase the accuracy of the learned rules. Indeed, rules for the upper turning point could be learned with the average accuracy of 75% in the leave-one-cycle-out runs. However, the accuracy for upswing decreased to 34% in the average. Hence, overall the time interval model did not enhance the results of the time point model in as much as we expected (see Table 4).

2.4 Modeling Two Phases

Theis and Weihs [10] have shown, that in clustering analyses of German macro-economic data at most three clusters can be identified. The first two clusters

Cycle	Accuracy	Coverage	No. learned rules
LOO1	0,8125	0,795	9 up, 69 down
LOO2	0,588	1,0	17 up, 35 down
LOO3	0,823	0,571	2 up, 15 down
LOO4	0,8	0,35	6 up, 8 down
LOO5	0,869	0,8	10 up, 39 down
LOO6	1,0	0,701	6 up, 41 down
Average	0,815	0,703	total 50 up, 207 down

Figure 5: Results in the two phase model using time points

roughly correspond to the cycle phases of upswing and downswing and the eventual third cluster corresponds to a time period around 1971. This suggests, that two phases instead of four may be more suited for the description of business data. It also points at a concept drift (see Section 2.5). In our third experiment we mapped all time points classified as upper turning point to upswing and all quarters of a year classified as lower turning point to downswing. We then applied the rule schemata of the first experiment. An example of the learned rules is:

$ie(T, V1), low(V1), c(T, V2), high(V2) \rightarrow down(T)$

stating that a low investment into equipment together with high private consumption indicates a downswing.

Again, leaving out the fifth or the sixth cycle gives the best results in the leave-one-cycle-out test. Accuracy and coverage are quite well balanced (see Table 5).

These learning results are promising. They support the hypothesis that a two phase model is of advantage for the dating task. Concerning the selection of indicators, the learning results show that all indicators contribute to the dating of the phase. However, the short term interest rate does not occur in three of the rule sets. Consumption (both the real value and the index), net exports, money supply, government deficit, and long term interest rate are missing in at least one of the learned rule sets. For the last four cycles, i.e. leaving out cycle 1 or cycle 2, some indicators predict the upswing without further conditions: high or medium number of salary earners (l), high or medium investment in equipment (ie), high or medium investment in construction (ic), medium consumption (c), and the real gross national product (y). It is interesting to note, that a medium or high real gross national product alone classifies data into the upswing phase only when leaving out cycle 1,2, or 4. Since RDT performs a complete search, we can conclude, that in the data of cycle 1 to cycle 4, the gross national product alone does not determine the upswing phase. Further indicators are necessary there, for instance money supply ($mon1$) or consumer price index (pc).

2.5 Concept shift

Starting from the two-phase model, we analyzed the homogeneity of the business cycle data. The learning results from different leave-one-cycle-out experiments were inspected with respect to their correlation. If the same rule is learned in all experiments, this means that the underlying principle did not change over time. If, however, rules co-occur only in the first cycles or in the last cycle, we hypothesize a concept drift in business cycles. We used the correlation analysis of the APRIORI algorithm [1], [12].

We want to know whether there are rules that are learned in all training sets, or, at least, whether there are rules that are more frequently learned than others. Enumerating all learned rules we get a vector for each training set (corresponding to a transaction in APRIORI) where the learned rule is marked by 1 and the others are set to 0. The frequency of learned rules and their co-occurrence is identified. There is no rule which was learned in all training sets. Eight rules were learned from three training sets. No co-occurrence of learned rules could be found. There is one rule, which was learned in four training sets, namely leaving out cycle 1, cycle 4, cycle 5, or cycle 6:

$rld(T,V), l(T,V), low(V) \rightarrow down(T)$

stating that the real long term interest rate and the number wage and salary earners being low indicates a downswing.

We now turn around the question and ask: which training sets share rules? For answering this question, a vector for each learned rule is formed where those training sets are marked by 1 which delivered the rule.

- Eighteen rules were shared in the training sets leaving out cycle 5 and leaving out cycle 6. Four of the rules predict an upswing, fourteen rules predict a downswing. This means, that cycles 1 to 4 have the most rules in common. The data from the last quarter of 1958 until the third quarter of 1974 are more homogeneous than all the data from 1958 until 1994.

- When leaving out cycle 1 or cycle 2, eleven rules occur in both learning results. This means, that cycles 3 to 6 have second most rules in common. The data from the second quarter of 1967 until the end of 1994 are more homogeneous than all data together.

- When leaving out cycle 2 or cycle 3, four rules occur in both learned rule sets.

- Larger item sets (frequently co-occuring rules) were rarely found: two rules were shared by learving out cycle 1 or cycle 2 or cycle 4, the one rule shown above is shared by the training sets leaving out cycle 1, cycle 4, cycle 5, or cycle 6.

The rule set analysis shows that cycles 1 to 4 (1958 – 1974) and cycles 3 to 6 (1967 - 1994) are more homogeneous than the overall data set. We wonder what happened in cycles 3 and 4. The first oil crisis happened at the end of cycle 4 (November 1973 – March 1974). This explains the first finding well. It shows that our rule set analysis can indeed detect concept drift, where we know that a drift occured. However, the oil crisis cannot explain why cycles 3 to 6 share so many rules. The second oil crises occured within cycle 5 (1979 – 1980). We assume that the actual underlying rules of business cycles may have changed over time. The concept drift seems to start in cycle 3. The periods of cycles 1 and 2 (1958 – 1967) are characterized by the reconstrucion after the world war. Investment in construction (*ic*) and in equipment (*ie*) is not indicative in this period, since it is rather high, anyway. A low number of earners (l) together with a medium range of the gross national product deflator (*pyd*) best characterizes the downswing in cycles 1 to 3 – this rule has been found when leaving out cycles 4 or 5 or 6. Since the unemployment rate was low after the war, it is particularly expressive for dating a phase in that period. This explains the second finding of our rule set analysis.

3 Conclusion and Further Work

Machine learning techniques in concert have answered the questions that have been our starting point(see Section 1).

- ILP offers opportunities for the analysis of business cycle data. It is easy to interpret the results so that the learned rules can be inspected by economists easily. The multi-variate nature of ILP and the automatic selection of most relevant indicators fits the needs of dating problem. Its performance was at least comparable to statistical methods.
- Decision tree learning could effectively find appropriate ranges that could be used for discretization. Furthermore, the value ranges could be used to determine time intervals.
- The two-phase model of the business cycle clearly outperformed the four-phase model. Where the best average accuracy in the four-phase model was 53%, the average accuracy of the two-phase model was 82%.
- Rule set analysis in terms of correlations between training set results shows that cycles 1 – 4 (1958 - 1974), i.e. leaving out cycle five or cycle six, had more rules in common than other cycles. The second most rules in common were found when leaving out the first or the second cycle, that is when training on cycles 3 – 6 (1967 - 1994). Both findings can be explained in economical terms.

The results could well be further enhanced. We used discretization in a straightforward manner by creating the interval segmentation based on the discretization of values. This can be extended by using piecewise constant or piecewise linear regression to get the interval segmentation directly. However, in this approach it is unclear, how the slope of an approximating linear function can be interpreted. For our application understandability is a main goal. The discretization might also consist of more complex patterns like peaks or valleys or patterns with outliers. Algorithms that find these patterns [4] can be used as to preprocess the time series.

The partitioning into two phases was very simple. A more sophisticated split within the upper and the lower turning phase, respectively, should lead to enhanced accuracy.

Finally, the concept drift could be the reason for not reaching the level of accuracy that we are used in other domains. Hence, training seperately cycles 4 to 6 and restricting the leave-one-cycle-out testing to these cycles could also enhance the learning results.

Acknowledgment This work has been partially sponsored by the Deutsche Forschungsgemeinschaft (DFG) collaborative research center 475 "Reduction of Complexity for Multivariate Data Structures". The authors thank Ullrich Heilemann, vice president of the Rheinisch-Westfälische Institut für Wirtschaftsforschung, for data of high quality and just as valuable suggestions. We also thank Claus Weihs and Ursula Sondhauss for raising our interest in the task and providing insight in its statistical nature.

References

[1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large data bases. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94)*, pages 478–499, Santiago, Chile, sep 1994.

[2] J. F. Allen. Towards a general theory of action and time. *Artificial Intelligence*, 23:123–154, 1984.

[3] Marlene Amstad. *Konjunkturelle Wendepunkte: Datierung und Prognose.* St.Gallen, 2000.

[4] Donald J. Berndt and James Clifford. Finding patterns in time series: A dynamic programming approach. In Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, chapter 3, pages 229–248. AAAI Press/The MIT Press, Menlo Park, California, 1996.

[5] Rheinisch-Westfälisches Institut für Wirtschaftsforschung. *Arbeitsbericht 2000.* Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Essen, Germany, 2000.

[6] U. Heilemann and H.J. Münch. *West German Business Cycles 1963-1994: A Multivariate Discriminant Analysis.* CIRET-Conference in Singapore, CIRET-Studien 50, 1996.

[7] U. Heilemann and H.J. Münch. *Classification of German Business Cycles Using Monthly Data.* SFB-475 Technical Reports 8/2001. Universitaet Dortmund, 2001.

[8] Jörg-Uwe Kietz and Stefan Wrobel. Controlling the complexity of learning in logic through syntactic and task-oriented models. Arbeitspapiere der GMD 503, GMD, mar 1991.

[9] John Ross Quinlan. *C4.5: Programs for Machine Learning.* Machine Learning. Morgan Kaufmann, San Mateo, CA, 1993.

[10] Winfried Theis and Claus Weihs. Clustering techniques for the detection of business cycles. SFB475 Technical Report 40, Universität Dortmund, 1999.

[11] Claus Weihs Ursula Sondhauß. Using labeled and unlabeled data to learn drifting concepts. In Miroslav Kubat and Katharina Morik, editors, *Workshop notes of the IJCAI-01 Workshop on* Learning from Temporal and Spatial Data, pages 38–44, Menlo Park, CA, USA, 2001. IJCAI, AAAI Press. Held in conjunction with the International Joint Conference on Artificial Intelligence (IJCAI).

[12] Ian Witten and Eibe Frank. *Data Mining // Practical Machine Learning Tools and Techniques with JAVA Implementations.* Morgan Kaufmann, 2000.

[13] D.A. Zighed, S. Rabaseda, R. Rakotomalala, and Feschet F. Discretization methods in supervised learning. In *Encyclopedia of Computer Science and Technology*, volume 40, pages 35–50. Marcel Dekker Inc., 1999.