ИКОНОМИКА И УПРАВЛЕНИЕ, ГОД. VI, №1

гл.ас. д-р ЕЛЕНА КАРАЩРАНОВА
ЮГОЗПАДАЕН УНИВЕРСИТЕТ „НЕОФИТ РИЛСКИ", БЛАГОЕВГРАД

ст. ас. ИРЕНА АТАНАСОВА
ЮГОЗПАДАЕН УНИВЕРСИТЕТ „НЕОФИТ РИЛСКИ", БЛАГОЕВГРАД

МОДЕЛ ЗА ПРИЛАГАНЕ НА ЗАВИСИМОСТТА МЕЖДУ СЛУЧАЙНИТЕ СЪБИТИЯ В БАЗИТЕ ОТ ЗНАНИЯ

A MODEL FOR IMPLEMENTATION OF THE DEPENDENCE BETWEEN RANDOM EVENTS IN KNOWLEDGE BASES

Chief Assistant Prof. Dr. ELENA KARASHTRANOVA
SOUTH-WEST UNIVERSITY "NEOFIT RILSKI" - BLAGOEVGRAD

Senior Assistant Prof. IRENA ATANASOVA
SOUTH-WEST UNIVERSITY "NEOFIT RILSKI" - BLAGOEVGRAD

Abstract: The measures of the dependence between random events can be used for practical purposes with the reasonoble interpretations and explanations. In this paper we will try to build an ontology in the business analysis domain (forecasting and assessment subdomain) in order to develop an easy and correct way to find the needed information in a knowledge base. We will illustrate that by representing a conceptual model of a subontology for forecasting of the ontology Onto-BAn.
Keywords: empirical estimation, knowledge base, ontology, forecasting, assessment.

1. Introduction

Ontology represents common semantics of the domain [Guarino 1997]. It provides a shared understanding and consequently an adequate communication protocol among systems. Since the nineties, ontologies have become one of the hottest issues of research among different communities - Artificial Intelligence, Databases, Knowledge bases, etc. Nowadays, ontologies have become the backbone of many enterprises and governmental institutions. Research in ontologies has been conducted in many different areas.

Our work is motivated by research in other domains as well as by our intention to exploit repositories of information in a specific domain - business analysis. Despite the extensive research in various areas of ontologies, there are still many open problems of research:

- engineering methodologies of integrating and merging a great number of existing ontologies in various domains;
- constructing ontologies by exploring new domains will not only assist in solving the semantic heterogeneity in such domains but will also allow for evaluating and optimising the existing methodologies of development.

In [Vet, Mars 1998] Vet and Mars say: "It is profitable to gain practical experience with ontologies for nontrivial domains, this gives the ontology builder a strt..."

In our area of interest, domain and its applications appear to be an area of challenge in constructing ontologies for some subdomain.

Objective: in this paper, we intend to describe a building ontology in the domain of business analysis (especially the subdomain of time series, forecasting, assessments).

Outlile of the paper: The rest of this paper is organized as follows. Section 2 outlines some ontology definitions and answers the question: why do we prefer to create a new ontology instead of integration of ontology. Section 3 presents the submodel of the model of business analysis for a company – model of forecasting. In section 4 we discuss the methods of forecasting, which are the classes of our ontology. Section 5 concludes this paper.

2. Ontologies

In the past research into ontologies was rather confined to the philosophical sphere. Currently it is widespread in research fields as diverse as knowledge representation, knowledge engineering, database, information system, knowledge management and organization [Giunchiglia et al. 2003]. Ontology can be seen as meta data that explicitly represents the semantics of data in a machine processable way. Ontologies could help people and computers to access the information they need by making the link between the information form and content explicit. Moreover, ontology is now recognized as powerful tool that enables sharing knowledge [Sure, Corcho 2003].

An ontology defines a common vocabulary for researchers who need to share information in a domain. A domain ontology corresponds to an organized set of domain generic terms that can be used to describe a particular domain by providing machine-interpretable definitions of basic concepts in the domain and the relationships between them [Noy, McGuinness 2001].

An ontology of a specific domain is useful in two aspects:

- to make the understanding of the process in a specific domain easier;
- to obtain a standart representation that can be shared and reused in other tools. It is important to highlight that different tools have been developed by several designers and there is no common vocabulary, so ontologies seem to be an appropriate mechanism for integration.

However, recent research has experimentally proved that ontologies are not enough to guarantee semantic interoperability. Four main problems have been detected [Correa et al. 2002]:

1. Reusing ontologies to engineer new ontologies is not straightforward. Guarino observes that ontologies developed from a bottom-up approach based on multiple local ontologies may not work because they focus on conceptual relations in a specific context [Guarino 1998]. Therefore, there is no guarantee that two systems with the same vocabulary have the same conceptualisation. This is what he calls the ontology integration problem. In order to deal with this problem, several authors argue in favour of mapping mechanisms between ontologies [Schorlemmer, Kalfoglou 2003], while others [Guarino 1998] propose using different kinds of ontologies. Guarino distinguishes between top-level, domain, task and application ontologies

2. .

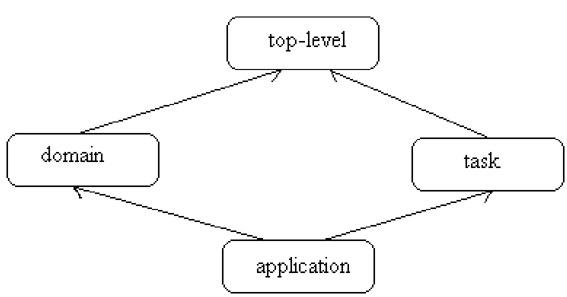

Figure 1. Kinds of ontologies [Guarino 1998]

Domain level ontologies describe the vocabulary related to a generic domain.

3. Ontologies do not provide adequate information when sharing inferences.

4. When reasoning under uncertainty, additional semantic links regarding inference are required.

5. Sharing group knowledge should be appropriately studied in a large scale system.

3. A model of ontology of the business analysis domain (subdomain forecasting and assessment)

In this paper we propose a domain ontology for business analysis. In order to elaborate an ontology for that domain we have used terms proposed in [Karlberg 2003] and [Newbold 1984] for the business analysis of a company.

Onto-BAn ontology

The main terms are organized in:

- Report - Nature and analysis of financial report ;
- Planning and control - Financial planning and control;
- Investment decisions;
- Sales and marketing.

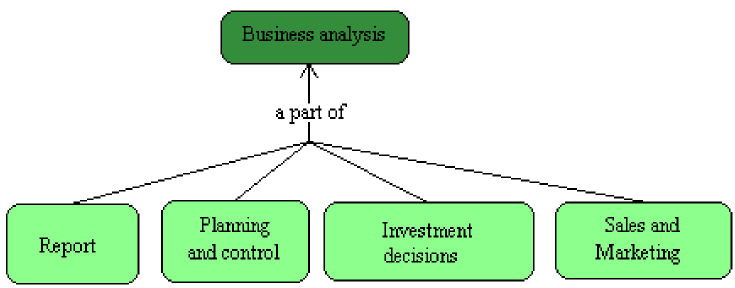

Figure 2. General diagram of the ontology Onto-BAn (superclass and classes)

Each term is defined in properties and relations, generating a complex network of classes, subclasses, instance and slots. The ontology Onto-BAn is designed as a reflection of the views of specialists in the areas of business analysis, statistics, and knowledge engineering. The model of subontology of Forecasting and evaluations can be used as a starting point for the elaboration of a general model of the Business analysis and models of subdomains of business analysis. For the ontology graphical presentation we will use objects with different shapes and colors and arrows with different colors.

	Dark green background	Superclass
	Bright green background	Class
	Orange background	Subclass of class
	Gray background	Instance
	Black arrow	Relation of type „is”, „a part of”
	Blue arrow	Relation between two classes
	Brown arrow	„Instance”

Figure 3. Objects and arrows

In the following we describe the class “Forecasting and evaluations” of the class “Planning and control”.

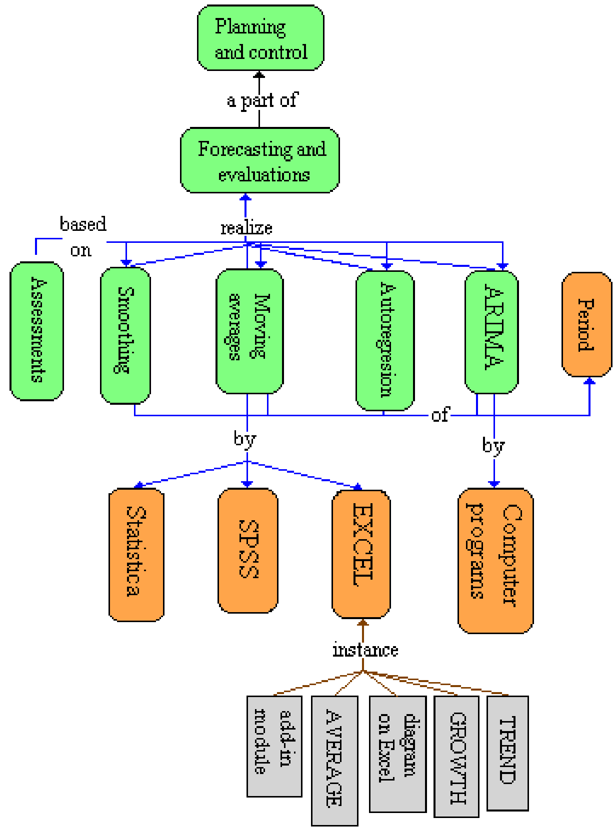

Figure 4. Class “Forecasting and evaluations”

Description of class “Forecasting and evaluations”

The class “Forecasting and evaluations” (see figure 4) is designed to describe the basic methods of forecasting and assessments. For the purpose of that investigation the concepts are grouped into four classes:

- Moving averages;
- Smoothing;
- Autoregression;
- ARIMA.

Since forecasting and assessments are very important for a company they are connected to other classes of the ontology Onto-BAn. The class ”Forecasting” and subclass “Assessments” correspond to the major issues directly, which makes them interesting for analysts – what is planned, what, how and where it is transacted, which are the results and the effects, what is the feedback to the corporate leadership.

The relations in class “Forecasting and evaluations” are of the type “a part of”, “realize”, “of”, “base on”, “by”, and “instance”.

For clarity, only some instances of subclass “Excel” are shown (the most popular). When we realize the ontology Onto-BAn in Protégé, we will show some (the most important) instances of subclasses “SPSS” and “Statistica”.

4. Time series and forecasting

In this section we will deal with some of the issues involved in analyzing a special type of data set. Specifically, we are interested in measurements through time on a particular variable. For examples: monthly product sales, quarterly corporate earning, and daily closing prices for shares of common stock.

Time series are likely to be characterized by certain types of dependence. Thus, an important assumption underlying the great majority of the statistical procedures will very probably not hold for time series data. It is the case that this assumption is typically rather crucial, so that the analysis of a time series as if it consisted of independent measurements can produce seriously misleading conclusions.

We have discussed a negative aspect of the typical kinds of dependency patterns likely to be present in time series data. Certainly these patterns do create problems, necessitating the development of special techniques of data analysis. However, inherent in this same phenomenon lies an opportunity. It is often possible to exploit any dependencies revealed in the past to produce forecasts of future values of a time series.

We will let the series of interest be denoted X_1, X_2, …, X_n, so that at time t, the observed value of a series is represented by X_t.

One way of thinking about the behavior of an actual observed series is to regard it as being made up of various components. Traditionally, four possible components are considered, with the notion that any or all might be present in any particular series. These components are as follows:

(i) Trend component
(ii) Seasonality component
(iii) Cyclical component
(iv) Irregular component

Many time series met in practice exhibit a tendency either to grow or to decrease fairly steadily over time, and this pattern is identified as trend.

Many business and economic time series met in practice consist of quarterly or monthly observations. It is often the case that such series exhibit the phenomenon of seasonality, such that patterns are repeated from year to year.

How one approaches the phenomenon of seasonality depends on the objectives. In some applications, such as routine sales forecasting for the purposes of inventory control, it is important to obtain as good an assessment as possible of the likely outcome in each future month. In that case, it is clear that any pronounced seasonal pattern, which might reasonably, be expected to recur in the future, will provide an important constituent in forecast derivation.

For some purposes, seasonality is rather a nuisance. In many applications, the analyst requires an assessment of overall movements in a time series, uncontaminated by the influence of seasonal factors.

Seasonal patterns in a time series constitute one form of regular, oscillatory behavior. In addition, many business and economic time series met in practice appear to exhibit oscillatory, or cyclical, patterns unconnected with seasonal behavior. These patterns might, for example, mirror business cycles in the economy at large. They are not necessarily regular, but follow rather smooth patterns of upswings and downswings.

We have discussed three sources of variability in a time series. If the only components of a series were trend, seasonality, and cycle, we would expect the time plot of that series to be very smooth and rather easily projected forward to produce forecasts. However, actual data do not behave in this way. In addition to the components already considered, there will be an irregular element, induced by the multitude of factors influencing the behavior of any actual series and whose pattern looks rather unpredictable on the basis of past experience. We might think of this component in much the same way as the error term in a regression model.

The conceptual breakdown of a time series into trend, seasonal, cyclical, and irregular components provides us with a very useful vocabulary for describing its behavior. It is often convenient to go beyond verbal description and to think in terms of a more formal model. Let X_t denote the value of a series at time t. Then we might think of this series as the sum of its components, through the additive model:

$$X_t=T_t+S_t+C_t+I_t$$

Where:
T_t=Trend component
S_t=Seasonal component
C_t=Cyclical component
I_t=Irregular component.

Alternatively, in some circumstance it might be more appropriate to view a series as the product of its constituent components, through the multiplicative model:

$$X_t=T_tS_tC_tI_t$$

In fact it is not necessary to restrict attention to just these two models. Under certain circumstances it may be convenient to treat some factors as additive and others as multiplicative.

- Moving averages: The irregular component in some time series may be so large that it obscures any underlying regularities, thus rendering difficult any visual interpretation of the time plot. Under these circumstances, the actual plot will appear rather jagged, and we may want to smooth it to achieve a clearer picture. This smoothing can be achieved through the method of moving averages, which is based on the idea that any large irregular component at any point in time will exert a smaller effect if the observation at that point is averaged with its immediate neighbors. The simplest technique of this kind is called a simple centered (2m+1)-point moving average. The idea here is to replace each actual observation X_t by the average of itself and its m neighbors. From some points of view, this component (strong seasonal component) is rather a nuisance, and the analyst often wants to remove it from the series, to obtain a keener appreciation of the behavior through time of other components. To be specific, suppose that we have a quarterly time series with a seasonal component. Now, suppose we produce a series of moving averages, whose first term is the average of the first four values of the original series, whose second term is the average of the second through fifth values of the original series, and so on. Then each member of the series of moving averages will be constituted from single observation from each of the four quarters. The series formed in this way should, therefore, be free from strong seasonal patterns. We will mention here a seasonal adjustment approach that is based on an implicit assumption of a very stable seasonal pattern through time. This is known as the seasonal index method. Essentially, the assumption is that for any given month or quarter, in each year, the effect of seasonality is to raise or lower the observation by a constant proportionate amount, compared with what it would have been in the absence of seasonal influences.
- Smoothing: Now we introduce a simple forecasting procedure which is in itself often valuable, and which forms the basis of some more elaborate methods, one of which is known as simple exponential smoothing, is appropriate when the series to be predicted is nonseasonal and has no consistent upward or downward trend. In the absence of trend and seasonality, the objective is to estimate the current level of the time series. This estimate is then used as the forecast of all future values. Many business forecasting procedures in common use are elaborations of the simple exponential smoothing approach. One of such method is the Holt-Winters model.
- Autoregression: A rather different approach to time series forecasting involves using the available data to

construct a model that might have generated the series of interest. In this section we will consider a very useful class of such models. The idea is to regard a time series as a series of random variable. For practical purposes, we might often be prepared to assume that these random variables all have the same means and variances. However, it would be rash, to say the least, to assume that they were independent of one another. Consider, for example, a series of annual values of product sales. We might suspect that the level of sales in the current period would be related to the levels in the immediately preceding years. Thus, we might expect to find a pattern of correlation through time in our series. Correlation patterns of this kind are sometimes referred to as autocorrelation. Now, in principle, any number of autocorrelation patterns is possible.

- Autoregressive integrated moving average models - ARIMA: When the possibility of differencing the data before fitting an autoregressive moving average model is considered, the resulting class of models is called Autoregressive integrated moving average or ARIMA. These models are now often used in business forecasting. Their popularity was stimulated, to a very large extent, by [Box, Jenkins 1970]. They set out a practical methodology for building such forecasting models. For that reason, this approach to forecasting is sometimes referred to as the Box-Jenkins approach. Here, we will be able to present only a brief outline of ARIMA model building methods [Nelson 1973]. Box and Jenkins describe an iterative three-part model building strategy for fitting an appropriate ARIMA model to a particular set of time series data. The three stages are follows:
  - Model selection;
  - Parameter estimation;
  - Model checking.

At the initial stage, a specific model from the general ARIMA class is chosen, based on statistics calculated from the data. That is, a decision is made as to whether to work with the original series or one of its differences, and specific values are chosen for p and q, the orders of the autoregressive moving average model to be fitted. Although the data should contain useful information as to what might be an appropriate model, we will not be able to fix with certainty the right choice. The procedures used require a good deal of judgment, and thus are inexact. For that reason, the analyst is not irrevocably committed to the original model chosen. At a later stage in the analysis this model might be abandoned in favor of some alternative if the available evidence suggests the desirability of such a course. This approach to time series, which has become popular in business forecasting, has several advantages. In concluding this brief discussion, we note that computer programs for carrying out an analysis through ARIMA models are now widely available.

5. Conclusions

Ontology building process is characterized by its very high cost and elaborate overlapping activities of development. Researchers have proposed many approaches namely bottom-up, top-down, and middle-out. In [Vet, Mars 1998] Van der Vet sees that a bottom-up approach is very attractive for many scientific and engineering fields. The approach focuses on building complex concepts from their primitive (basic) concepts and a list of construction rules. We are going to use this approach to design, engineer, and create our ontology.

In knowledge engineering, a number of alignment tools are provided such as Protege [Grosso et al. 1999], Chimeara [McGuinness et al. 2000], PROMPT [Noy, Musen 2000], but we prefer to use Protege to build our ontology. In this paper we present a subdomain ontology for subdomain forecasting. We provide the basic conceptualisation and we will make implementation of our subontology

“Forecasting and assessment” with Protege.

6. Future research

In the future we intend to realize the subontology for forecasting and assessment and the ontology Onto-BAn for business analysis domain by using the bottom-up approach and integrated environment Protégé 3.4.1.

References:

[Box, Jenkins 1970] Time Series analysis, Forecasting, and Control, San Francisco: Holden-Day, 1970.

[Giunchiglia et al. 2003] F. Giunchiglia, A. Gomez-Perez, H. Stuckenschmidt, A. Pease, Y. Sure, S. Willmott. 1 st International workshop on Ontologies and Distributed Systems, (ODS 2003), 2003.

[Grosso et al. 1999] W. Grosso, H. Eriksson, R. Fergerson, J. Gannari, S. Tu, M. Musen. Knowledge modeling at the milenium (the design and evolution of Protege 2000), In proceeding of the Twelfth BannfWorkshop on Knowledge Aquisition, Modeling, and Management, Voyager inn, Bannf, Alberta, Canada, 16-21, 1999.

[Guarino 1997] N. Guarino. Understanding, building and using ontologies., International Journal of Human and Computer Studies, 46(2/3):293-310, 1994.

[Guarino 1998] N. Guarino. Formal Ontology and Information Systems, Proceedings of FIOS'98, IOS Press, 3-15, 1998.

[Karlberg 2003] K. Karlberg. Business analysis with Microsoft Excel, SoftPress Ltd., 2003.

[McGuinness et al. 2000] D. McGuinness, R. Fikes, J. Rice, S. Wilde An environment for merging and testing large ontologies, in Proc. of KR-00, 2000.

[Nelson 1973] C. R. Nelson, Applied Time Series for Managerial Forecasting, San Francisco: Holden-Day, 1973.

[Newbold 1984] P. Newbold, Statistics for business and economics, Prentice-Hall, New Jersey, 1984.

[Noy, McGuinness 2001] N. Noy, D. McGuinness. Ontology Development 101: A Guide to Creating Your First Ontology - Stanford Knowledge Systems Laboratory Technical Report, KSL-01-05, 2001.

[Noy, Musen 2000] N. Noy, M. Musen. Prompt: Algorithm and tool for automated ontology merging and alignment., In proceedings of the 17th national Conference on Artificial Intellligence (AAAI - 2000), Texas, 2000.

[Schorlemmer, kalfoglou 2003] M. Schorlemmer, y. Kalfoglou. Using Information-Flow Theory to Enable Semantic Interoperability, In: Artificial intelligence Research and Development, I. Aguilo et al. (Eds), IOS PRESS, 421-431, 2003.

[Sure, Corcho 2003] Y. Sure, O. Corcho eds. 2 nd International Workshop on Evaluation of Ontology based Tools, (EON 2003), 2003.

[Vet, Mars 1998] P. Vet, N. Mars. Bottom-up construction of ontologies, In IEEE Transaction on knowledge and Data Engineering, vol. 10, no 4, 513-526, 1998.