Statistics and Probability Letters 136 (2018) 165–169

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

On the role of latent variable models in the era of big data

Francesco Bartolucci [a,*], Silvia Bacci [a], Antonietta Mira [b,c]

[a] Department of Economics, University of Perugia, Via A. Pascoli 20, 06123 Perugia, Italy
[b] Institute of Computational Science, Università della Svizzera italiana, Via Buffi 13, Lugano, Switzerland
[c] Università dell'Insubria, Como, Italy

ARTICLE INFO

Article history:
Available online 21 February 2018

MSC:
00-01
99-00

Keywords:
Bayesian inference
Complex data
Maximum likelihood estimation
Parallel computing
Selection bias

ABSTRACT

We discuss how latent variable models are useful to deal with the complexities of big data from different perspectives: simplification of data structure; flexible representation of dependence between variables; reduction of selection bias. Problems involved in parameter estimation are also discussed.

1. Introduction

Big data present different types of complexity that go well beyond their mere size and make their use problematic, possibly leading to biased conclusions on the main investigated effects. We discuss the role of latent variable models (LVMs; Skrondal and Rabe-Hesketh, 2004) in dealing with these complexities, since they offer a very flexible framework to formulate probabilistic assumptions that lend themselves to a meaningful interpretation. In particular, LVMs are useful from at least three different perspectives: (*i*) simplification of data structure thanks to clustering patterns of the response in homogeneous and interpretable groups; (*ii*) flexible representation of the dependence between the response variables, by formulating in a simple way the distribution of these variables given the latent variables and assuming suitable conditional independence relations; (*iii*) reduction of the distortions due to nonrandomness of the data, by accounting for certain features of the data collection process that are typically related to problems of self selection.

As any other model-based approach, parameter estimation for LVMs becomes problematic as the sample size increases. However, certain tools, also coming from machine learning, are becoming popular in statistics to solve these problems with reference to LVMs. In this note we discuss two approaches, based on composite likelihood and variational approximations, which play a role both in frequentist and Bayesian inference.

2. Complex data structures

At least two different types of complexity typically arise in analyzing big data (Dunson, 2016). The first is that the data are referred not only to a huge number of units, but also to a large number of variables of mixed type (categorical and

* Corresponding author.
E-mail addresses: francesco.bartolucci@unipg.it (F. Bartolucci), silvia.bacci@unipg.it (S. Bacci), antonietta.mira@usi.ch (A. Mira).

https://doi.org/10.1016/j.spl.2018.02.023
0167-7152/

quantitative, continuous and discrete) for which defining a joint distribution may be difficult; sometimes, it is also difficult to disentangle covariates from response variables. The second type of complexity, which may arise even if the variables are few, is related to the structure of the sample or to the data collection methodology. In this regard, it is worth recalling that many different complex data structures may be met, such as longitudinal data, in which observations are temporally ordered; multilevel data, in which units are grouped in clusters; or social network data, where the connection between pairs of individuals is of interest. Combinations of these structures are also available, such as longitudinal social network or multilevel data.

Defining a simple and general notation that covers all situations is a difficult task. To simplify, we distinguish two types of problem:

– *Type I*: Given n units of interest, we can associate independent vectors of response variables to each unit. These vectors are denoted by $\boldsymbol{Y}_i$ ($i = 1, \ldots, n$) and are not constrained to have the same number of elements. A vector of covariates $\boldsymbol{X}_i$ is also considered for each unit i. This structure typically arises with cross-sectional or longitudinal data. It may also arise with multilevel data, in which case the label i refers to a single cluster. The collections of all the vectors of covariates and response variables are denoted by $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively.

– *Type II*: This is a setting where it is not possible to reasonably assume any independence between the response variables. This is the typical case in which each sample unit can potentially interact with any other unit, as in a social network. Though we could split the set of all observable variables in unit-specific subvectors, in general there is no real convenience in doing so and thus we simply denote by $\boldsymbol{X}$ and $\boldsymbol{Y}$ the sets of all covariates and responses, respectively, although their structure may be complex.

3. Latent variable models

In the following we provide a definition of LVMs and illustrate their advantages in dealing with big data complexities.

3.1. General formulation

We propose a definition of LVMs that tries to take into account the most important features of these models in connection with the context of interest: *An LVM assumes the existence of a set of unobservable (latent) variables $\boldsymbol{U}$ such that the conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{U}$ and $\boldsymbol{X}$ has a simplified structure*. This means that the distribution of the response variables, given the covariates and the latent variables, follows a simple parametric form and/or certain conditional independence relations among the elements of $\boldsymbol{Y}$ hold (*local independence*). Also note that the distribution of $\boldsymbol{U}$ is usually rather simple and based on assumptions of independence between its elements or subsets of them.

It is worth clarifying that two components must be specified in formulating an LVM: the *structural model* and the *measurement model*. The first refers to the conditional distribution of the latent variables given the covariates, which has a probability or density function denoted by $p(\boldsymbol{u}|\boldsymbol{x})$. The second refers to the conditional distribution of the response variables given the covariates and the latent variables, which corresponds to $p(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{x})$. Consequently, the *manifest distribution* of the response variables is

$$p(\boldsymbol{y}|\boldsymbol{x}) = \int_{\boldsymbol{u}} p(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{x}) p(\boldsymbol{u}|\boldsymbol{x}) d\boldsymbol{u}, \tag{1}$$

whereas the posterior distribution of the latent variables is

$$p(\boldsymbol{u}|\boldsymbol{x}, \boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{x}) p(\boldsymbol{u}|\boldsymbol{x})}{p(\boldsymbol{y}|\boldsymbol{x})}. \tag{2}$$

Note that, if the latent variables are assumed to have a discrete distribution, then the integral in (1) becomes a sum over all possible configurations of $\boldsymbol{U}$.

Here the distinction between Type I and Type II problems plays a relevant role. First of all, in Type I problems the vector of latent variables $\boldsymbol{U}$ may be split in independent vectors $\boldsymbol{U}_i$, $i = 1, \ldots, n$, with each $\boldsymbol{U}_i$ affecting only $\boldsymbol{Y}_i$. Moreover, the manifest distribution $p(\boldsymbol{y}|\boldsymbol{x})$ may be factorized as $\prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{x}_i)$, with $p(\boldsymbol{y}_i|\boldsymbol{x}_i)$ computed by a formula similar to (1) with $\boldsymbol{u}$, $\boldsymbol{x}$, and $\boldsymbol{y}$ replaced by $\boldsymbol{u}_i$, $\boldsymbol{x}_i$, and $\boldsymbol{y}_i$, respectively. Similarly, the posterior distribution may be derived, separately for each vector $\boldsymbol{U}_i$, as in (2). The same simplifications do not hold for Type II problems, with implications that will be discussed in Section 4.

In the following, some examples of LVMs are provided that are useful for our discussion; the first three are Type I problems:

1. *Finite mixture (FM) models* (McLachlan and Peel, 2004): A discrete latent variable U_i is associated to each unit i and the distribution of $\boldsymbol{Y}_i$ given U_i may be arbitrarily defined. The most popular case is when this distribution is a multivariate normal. *Latent class models* (Lazarsfeld and Henry, 1968) are particular FM models for categorical data, in which the elements of $\boldsymbol{Y}_i$ are assumed to be conditionally independent given U_i. In any case, the main aim is to cluster units in separate groups identified by the support points of the distribution of U_i.

2. *Multilevel models* (Goldstein, 2011): These models are used in the presence of multilevel data structures with units nested in clusters, to account for the influence, on the individual responses, of unobservable factors associated to each of these clusters. Every cluster-specific vector of latent variables $\boldsymbol{U}_i$ has typically a continuous distribution and they may also

have a complex structure in the case of hierarchical data with more than two levels of nested units. Moreover, these models are strongly related to *random-effects models* that are commonly used with longitudinal data.

3. *Hidden Markov (HM) models* (Bartolucci et al., 2013): These may be conceived as an extension of FM models for the analysis of longitudinal data. In this case, $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$ are unit specific vectors of time-specific latent variables each of which follows a Markov chain, so that the evolution of the phenomenon of interest may be studied across time. Note that such latent variable vectors are independent each other, but the elements inside each $\boldsymbol{U}_i$ are not independent.

4. *Stochastic block (SB) models* (Nowicki and Snijders, 2001): The response variables are Y_{ij}, with $i, j = 1, \ldots, n$, $i \neq j$, and describe the type of connection between every pair of units (i, j). Moreover, a discrete latent variable U_i is associated to each unit i to identify the cluster to which the unit belongs, with members of the same block having a similar behavior and/or being strongly connected. This is a case of Type II problem in which computing the manifest distribution in (1) has a numerical complexity of order n^k, where k is the number of blocks. As a result, estimation is infeasible already with small samples.

3.2. Dealing with data complexity

LVMs offer a flexibility that is particularly useful to deal with the complexity of big data from different points of view. We articulate our discussion following three main perspectives.

3.2.1. Simplifying the data structure

An effective way to simplify the structure of a (big and complex) dataset is to build clusters of homogeneous response patterns. By using discrete latent variables it is possible to build such clusters in a meaningful and interpretable way. There is a wide statistical literature on this method which is known as *model-based clustering* (Fraley and Raftery, 2002) and is essentially based on FM models. The advantage of using this approach based on latent variables, with respect to using more traditional methods, is that an interpretation in terms of data generation process is possible. Moreover, well principled statistical criteria may be used for model selection and, in particular, to choose the number of clusters and the distribution of the response variables given the cluster. In this regard it is important to recall that the class of FM models is an example of contamination between statistics, engineering, and computer science. In fact, these models are also very popular in pattern recognition and machine learning in general; the same may be stated also for HM models.

3.2.2. Modeling dependence in a flexible way

When a dataset contains many variables and/or has a longitudinal/multilevel structure, it is important to account for the dependence between the response variables in a suitable way. Formulating a multivariate distribution for these variables that accounts for their dependencies may be a very difficult task, in particular when they are of mixed type. On the other hand, introducing latent variables permits to simplify the problem by separately formulating the distribution of each single response variable, or small blocks of response variables, given the latent variables, as is already explicit in our definition of LVMs. Then, the joint distribution of all responses given the latent variables is obtained relying on the local independence assumption. As already discussed in Section 3, this means that the response variables are conditionally independent (individually or blockwise), given the latent variables. Note, however, that this does not imply that the response variables are marginally independent; this is a crucial point connected with the use of LVMs. In other terms, by introducing latent variables we decompose the problem of formulating a distribution for a complex structure in a series of problems of reduced complexity, following a "divide and conquer" strategy. Again, a meaningful model, based on relations between the variables involved in the system, typically results.

3.2.3. Accounting for the data collection process

When dealing with big data, a typical problem is that they cannot be considered on the same footing as data collected on the basis of a properly planned random sampling design. In certain cases these data cover all units in a population, such as that of all Facebook or Twitter users, the selection of which is difficult to clearly identify and, in any case, is very far from a random selection from a larger population that typically corresponds to the entire demographic population. The consequence is a possible strong bias in estimating causal relations (Pearl, 2009), which may be particularly dangerous in studying the efficacy of a certain treatment or policy. Through latent variables we can account for the selection process and, at least partially, correct for this bias. Recent approaches are based on formulating an interpretable model in which a vector of latent variables is associated to each sample unit, and this affects both the unit-specific response variables and an indicator variable for the chosen treatment or for the unit being present in the sample. This approach may be profitably used even with non-ignorable missing data on the basis of the so-called *latent ignorability* assumption (Bacci and Bartolucci, 2015).

4. Estimation of latent variable models

LVMs may be estimated following a frequentist or a Bayesian approach. The main issues involved in both approaches are here briefly summarized.

4.1. Frequentist approach

This approach is normally based on maximum likelihood estimation (MLE), consisting in the maximization of the log-likelihood $\ell(\boldsymbol{\theta}) = \log p(\boldsymbol{y})$, where $\boldsymbol{\theta}$ is the vector of all model parameters. This is performed by a direct maximization algorithm of Newton–Raphson type (e.g., Turner, 2008) or, especially in the presence of discrete latent variables, by the Expectation–Maximization (EM) algorithm (Dempster et al., 1977).

When dealing with big data, the main computational complexity arises in obtaining $p(\boldsymbol{y}|\boldsymbol{x})$ and $\ell(\boldsymbol{\theta})$. For Type I problems, computing $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log p(\boldsymbol{y}_i|\boldsymbol{x}_i)$ has a complexity that increases linearly with n. Therefore, for reasonable values of n, performing MLE of an LVM does not present computational problems, provided that the structure of latent variables underlying each $p(\boldsymbol{y}_i|\boldsymbol{x}_i)$ is sufficiently simple. Thanks to suitable recursions (Baum et al., 1970), this is also the case for HM models, whose estimation does not present relevant problems even with a large number of time occasions per unit.

On the other hand, in the presence of big data, MLE of an LVM presents the typical computational problems of statistical models. In general, it is natural to approach the problem of MLE with very simple solutions, essentially based on using a subsample of the data randomly chosen, or with more sophisticated solutions, based on splitting the overall dataset in small disjoint datasets. The second method, in particular, may exploit parallel computing to obtain separate estimates for each subset that must be then combined in a suitable way, so as to obtain a value close to the exact estimate corresponding to the (ideal) maximum of $\ell(\boldsymbol{\theta})$. Alternatively, parallel computing may be used to obtain this exact estimate by implementing an algorithm that, at each iteration, splits the computation of $\ell(\boldsymbol{\theta})$, and related quantities of interest (e.g., score), among different processors and then the results are suitably joined.

With Type II problems, LVMs may be difficult to estimate due to complexities in computing $p(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{x})$ and, then, $\ell(\boldsymbol{\theta})$, even with moderate sample sizes. The typical case is that of SB models for which obtaining the manifest distribution requires a number of operations that increases exponentially with n. Rarely this may also happen for Type I problems in which the latent structure behind each $p(\boldsymbol{y}_i|\boldsymbol{x}_i)$ is very complex (e.g., Bartolucci and Lupparelli, 2016). In this case, modern frequentist inference relies on different methods that may be seen as an approximation or are somehow related to the MLE and have good inferential properties. In particular, we recall composite likelihood methods and variational approximations. The first ones substitute $\ell(\boldsymbol{\theta})$ with $c\ell(\boldsymbol{\theta}) = \sum_{h=1}^{H} \ell_h(\boldsymbol{\theta})$, where each $\ell_h(\boldsymbol{\theta})$ is the (marginal or conditional) log-likelihood function referred to suitably defined subsets of response variables (Lindsay, 1988); see Varin et al. (2011) for an overview. Under regularity conditions, these methods lead to a consistent estimator although less efficient than that obtained from exact MLE. From a certain point of view, by using a composite likelihood method we are "parallelizing" the estimation process focusing on small subsets of response variables. Therefore, it seems that parallel computing may be implemented in a rather natural way in the composite likelihood setting by separating the computation of the log-likelihoods $\ell_h(\boldsymbol{\theta})$ among different cores, and this may represent an important research field.

Regarding variational approximations, these are typical methods taken from machine learning; for a review see Blei et al. (2017). In the present context, applying these methods amount to approximate the posterior distribution of the latent variables $p(\boldsymbol{u}|\boldsymbol{x}, \boldsymbol{y})$ defined in (2) by a suitable function $q(\boldsymbol{u})$ depending on a vector of parameters $\boldsymbol{\tau}$. Then, the target function to be maximized is defined as $J(\boldsymbol{\theta}, \boldsymbol{\tau}) = \ell(\boldsymbol{\theta}) - K(q(\boldsymbol{u}), p(\boldsymbol{u}|\boldsymbol{x}, \boldsymbol{y}))$, where the second term is the Kullback–Leibler distance between $q(\boldsymbol{u})$ and $p(\boldsymbol{u}|\boldsymbol{x}, \boldsymbol{y})$. For an implementation for SB models see Daudin et al. (2008); for a recent study of the statistical properties of this method we refer to Bickel et al. (2013). Moreover, how to use variational approximations in the presence of big data has been considered, among others, by Vu et al. (2013).

4.2. Bayesian approach

Bayesian inference, based on the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y}) \propto \pi(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x})$, where $\pi(\boldsymbol{\theta})$ denotes the prior on the model parameters, is also very popular for the estimation of LVMs. From a computational point of view, Markov chain Monte Carlo (MCMC) algorithms are used to practically perform Bayesian estimation when the posterior is not available in closed form. Typically, it is convenient to implement an MCMC in a data augmentation framework (Tanner and Wong, 1987; Van Dyk and Meng, 2001), in which latent variable values are sampled together with the model parameters. This scheme has an interpretation, in terms of missing data, similar to that of the EM algorithm and has clear advantages with respect to algorithms directly based on the manifest distribution $p(\boldsymbol{y}|\boldsymbol{x})$.

The general impression is that Bayesian estimation of an LVM is less problematic when compared with MLE. When implementing an MCMC algorithm, the main difficulty is tuning the proposal distributions (and this issue can be well addressed using adaptive MCMC), but at least with moderate values of n, typical problems affecting MLE of certain LVMs do not manifest themselves or are less severe in the Bayesian setting. This is because, relying on data augmentation, it is not necessary to explicitly evaluate $p(\boldsymbol{y}|\boldsymbol{x})$. In any case, for very complex data and models, methods based on composite likelihood or variational approximations are also available for Bayesian inference. On the other hand, with big data we may exploit the estimation methods for general statistical models that are available in the Bayesian literature, which is rather advanced in this regard (Dunson, 2016). For massive data, the point boils down to how efficiently splitting the whole data into disjoint subsets that are used to obtain separate estimates and how to appropriately recombine these estimates. In particular, MCMC algorithms can be separately run on different data subsets and on different machines (in parallel) but then the issue is how to recombine the MCMC samples that do not approximate the posterior distribution obtained conditioning on the entire data but only on the data subset stored on that machine. In this regard, the Bayesian literature is flourishing

and some efficient strategies are, for instance, the consensus Monte Carlo algorithm (Scott et al., 2016) and the use of Gaussian-process approximations (Nemeth and Sherlock, 2017).

A method that is also of interest, in particular to deal with the "velocity" aspect of big data, is that of sequential Bayesian estimation (Oravecz et al., 2017). Here, the posterior distribution on the model parameters obtained on data available at time t becomes the prior distribution for analyzing the next batch of data available at time $t + 1$, and so on. To approximate sequences of posterior distributions that evolve over time, specific efficient MCMC algorithms have been designed that exploit the fact that the posterior at time $t + 1$ typically does not differ much from the one at time t. The most popular is the class of Sequential Monte Carlo (SMC) algorithms where the distributions of interest are approximated by a collection of random samples that evolve over time using sampling and resampling mechanisms. SMC algorithms have been specifically designed also for MLE in LVMs and the method can be easily adapted to Bayesian marginal maximum a posteriori setting.

5. Conclusions

To conclude this note we stress some points. First of all, LVMs represent an important framework to deal with the complexities presented by nowadays big data. Second, the problems in estimating LVMs are those of typical statistical models and may be more severe in particular cases; however, by exploiting certain methods, some of which come from machine learning, it is possible to effectively face these difficulties. Finally, the Bayesian literature on how to deal with big data exploiting efficient computational algorithms seems more advanced not only in general, but also with reference to LVMs.

Our impression is that LVMs represent a very import research field in which statistics and machine learning meet each other within the general framework of *data science*. On one hand, statistics is more advanced than machine learning in formulating appropriate assumptions to account for data complexity and for taking parameter and model uncertainty into proper account, and this may be profitably done exploiting LVMs. On the other hand, methods developed in machine learning can be used for the efficient estimation of these models.

Our "biased" view of statisticians is that the formulation of models is more interesting and exciting than their estimation because, in formulating a model, a researcher touches issues related to the data generation process and to causal relations between variables, which represent essential elements to acquire knowledge on phenomena and to drive decision making processes.

References

Bacci, S., Bartolucci, F., 2015. A multidimensional finite mixture structural equation model for nonignorable missing responses to test items. Struct. Equ. Model. 22, 352–365.

Bartolucci, F., Farcomeni, A., Pennoni, F., 2013. Latent Markov Models for Longitudinal Data. Chapman & Hall/CRC, Boca Raton, FL.

Bartolucci, F., Lupparelli, M., 2016. Pairwise likelihood inference for nested hidden Markov chain models for multilevel longitudinal data. J. Amer. Statist. Assoc. 111, 216–228.

Baum, L.E., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann. Math. Statist. 41, 164–171.

Bickel, P., Choi, D., Chang, X., Zhang, H., 2013. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. Ann. Statist. 41, 1922–1943.

Blei, D.M., Kucukelbir, A., McAuliffe, J.D., 2017. Variational inference: A review for statisticians. J. Amer. Statist. Assoc. 112, 859–877.

Daudin, J.-J., Picard, F., Robin, S., 2008. A mixture model for random graphs. Stat. Comput. 18, 173–183.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 39, 1–38.

Dunson, D., 2016. Probabilistic inference for big & complex data. In: 48th Scientific Meeting of the Italian Statistical Society, June 8–10 2016, Università degli Studi Di Salerno, Italy.

Fraley, C., Raftery, A.E., 2002. Model-based clustering, discriminant analysis, and density estimation. J. Amer. Statist. Assoc. 97, 611–631.

Goldstein, H., 2011. Multilevel Statistical Models. John Wiley & Sons, Chichester, UK.

Lazarsfeld, P.F., Henry, N.W., 1968. Latent Structure Analysis. Houghton Mifflin, Boston, MA.

Lindsay, B.G., 1988. Composite likelihood methods. Contemp. Math. 80, 221–239.

McLachlan, G., Peel, D., 2004. Finite Mixture Models. John Wiley & Sons, New York, NY.

Nemeth, C., Sherlock, C., 2017. Merging MCMC subposteriors through Gaussian-process approximations. Bayesian Anal. 1–24. http://dx.doi.org/10.1214/17-BA1063.

Nowicki, K., Snijders, T.A.B., 2001. Estimation and prediction for stochastic blockstructures. J. Amer. Statist. Assoc. 96, 1077–1087.

Oravecz, Z., Huentelman, M., Vandekerckhove, J., 2017. Sequential Bayesian updating for big data. In: Jones, M. (Ed.), Big Data in Cognitive Science: From Methods to Insights. Psychology Press, Sussex, UK, pp. 13–33.

Pearl, J., 2009. Causality: Models, Reasoning, and Inference. Cambridge University Press, New York, NY.

Scott, S.L., Blocker, A.W., Bonassi, F.V., Chipman, H.A., George, E.I., McCulloch, R.E., 2016. Bayes and big data: The consensus Monte Carlo algorithm. Int. J. Manag. Sci. Eng. Manag. 11, 78–88.

Skrondal, A., Rabe-Hesketh, S., 2004. Generalized Latent Variable Modeling. Multilevel, Longitudinal and Structural Equation Models. Chapman & Hall/CRC, London, UK.

Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. J. Amer. Statist. Assoc. 82, 528–540.

Turner, R., 2008. Direct maximization of the likelihood of a hidden Markov model. Comput. Statist. Data Anal. 52, 4147–4160.

Van Dyk, D.A., Meng, X.-L., 2001. The art of data augmentation. J. Comput. Graph. Statist. 10, 1–50.

Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. Statist. Sinica 21, 5–42.

Vu, D.Q., Hunter, D.R., Schweinberger, M., 2013. Model-based clustering of large networks. Ann. Appl. Stat. 7, 1010–1039.