Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 651–654

Linguistic knowledge about temporal data in Bayesian linear regression model to support forecasting of time series

Katarzyna Kaczmarek
Systems Research Institute, Polish Academy of Sciences,
Newelska 6, 01-447 Warsaw, Poland
Email: K.Kaczmarek@ibspan.waw.pl

Olgierd Hryniewicz
Systems Research Institute, Polish Academy of Sciences,
Newelska 6, 01-447 Warsaw, Poland
Email: Olgierd.Hryniewicz@ibspan.waw.pl

***Abstract*—Experts are able to predict sales based on approximate reasoning and subjective beliefs related to market trends in general but also to imprecise linguistic concepts about time series evolution. Linguistic concepts are linked with demand and supply, but their dependencies are difficult to be captured via traditional methods for crisp data analysis. There are data mining techniques that provide linguistic and easily interpretable knowledge about time series datasets and there is a wealth of mathematical models for forecasting. Nonetheless, the industry is still lacking tools that enable an intelligent combination of those two methodologies for predictive purposes. Within this paper we incorporate the imprecise linguistic knowledge in the forecasting process by means of linear regression. Bayesian inference is performed to estimate its parameters and generate posterior distributions. The approach is illustrated by experiments for real-life sales time series from the pharmaceutical market.**

***Index Terms*—linguistic knowledge, time series analysis, Bayesian linear regression, posterior simulation**

I. Introduction

HUMAN-BEINGS have the unique ability to process imprecise information and solve complex problems based mostly on their intuition and expertise [14]. This ability allows us to easily interpret and describe temporal data in natural language with words and propositions. Such information, often imprecise, is called *temporal linguistic knowledge* within this paper.

As observed in a selected pharmaceutical company experts were able to predict future sales and make related decisions based on approximate reasoning about imprecise information driven from visual inspection of time series data sets. It was observed that experts recognized important dependencies between linguistic temporal knowledge even in situations when analysis of crisp time series datasets showed no significant correlations. We pose the question whether linguistic temporal knowledge may bring information about new correlations useful for the time series forecasting process.

The linguistic knowledge about temporal data is provided by the experts of selected domain or is extracted automatically thanks to the knowledge discovery and data mining techniques. Among the recent developments in the field of intelligent computing there are efficient methods that provide interpretable knowledge from huge datasets.

The problem of time series abstraction and labeling meaningful intervals by means of clustering, machine learning and function approximation methods, statistical test or multiscale methods is addressed e.g. in [2], [10], [13]. The concept of pattern recognition has been widely discussed for example in [7], [8], [9], [11], [12]. One of the goals of data mining research is to provide linguistic and human-consistent description of raw data. Within this paper we take data mining results as the input for the forecasting procedure.

We provide a predictive model to support decision making in the international pharmaceutical sales market. Linguistic knowledge about temporal data is transformed into imprecisely labeled sequences that are incorporated into the probabilistic model as explanatory variables. We adopt Bayesian linear regression model and perform posterior simulation. We operate on parameters for which linguistic concepts are transparent and could easily be interpreted by experts.

The structure of this paper is as follows. Next chapter introduces basic definitions related to temporal linguistic knowledge about time series. Chapter 3 presents the forecasting process with temporal linguistic knowledge incorporated into the regression model. The description of the experiments and results for time series from the pharmaceutical industry are gathered in chapter 4. Paper concludes with general remarks and further research opportunities.

II. Linguistic concepts about temporal data

In this section we define formal language for temporal linguistic concepts that we consider most appealing for predictive purposes.

Let $O = \{o_1, o_2 ..., o_q\}$ denote a finite set of objects in a considered domain. The properties of objects are measured by observables. Let $M = \{m_1, m_2 ..., m_r\}$ denote a finite set of observables in the considered domain.

Definition 1: ***Object's property***
A pair (o, m) such that $o \in O$ and $m \in M$ is called object's property.

The sequence of measurements for object's property is treated as discrete time series.

TABLE I
ILLUSTRATIVE EXAMPLES FOR IMPRECISE LABEL, OBSERVABLE AND OBJECT.

Domain	Object	Observable	Imprecise label
Pharmaceutical market	Product 1	sales	high
	Product 1	supply	high
	Product 1	sales	increasing
	Europe	inflation	increasing
Mood tracking	Patient A	anxiety	high
	Patient A	weight	constant
	Patient A	hours slept	constant
	Patient B	medication	increasing

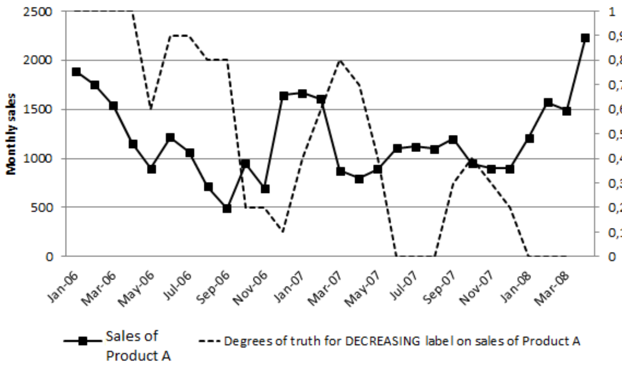

Fig. 1. Example of time series and imprecise labeled sequence.

Definition 2: ***Discrete time series***
Discrete time series $\{y_t\}_{t=1}^n \in \Psi_n$ *is a sequence of observations of given object's property* (o, m) *such that* $o \in O$ *and* $m \in M$ *measured at successive* $t \in T = \{1, ..., n\}$ *moments and at uniform time intervals. For each* $t \in T$ *the observation* y_t *is a realization of the random variable* Y_t*. Random variables* Y_t *are defined on the probability space* (Ω, A, P)*, where* Ω *is the set of all possible outcomes of the random experiment,* A *is a* σ*-field of subsets of* Ω*, and* P *is a probability measure associated with* (A, P)*.*

As stated in [1] during visual inspection people perceive and process shapes rather than single data points. We describe the evolution of time series with adjectives like *high, medium, low, light, heavy, interesting, increasing, constant, decreasing, interesting, long, short, strong, weak, slight, etc.* Such adjectives refer to imprecise values, trends, judgments or features and are called *imprecise labels* within this paper.

Let $S = \{s_1, s_2..., s_l\}$ denote a finite set of imprecise labels referring to either qualitative or quantitative measurements for observables applicable in the considered domain. Depending on the context, values for the imprecise label are assigned subjectively by experts or are calculated based on the fuzzy numbers and membership functions. For basic definitions related to the fuzzy sets theory see e.g. [6].

In real-life situations understanding and interpretation of imprecise labels depends on context and may change in time. Within this approach we assume one interpretation that is constant in time.

As presented in Table I for application *sales, supply* and *inflation* are observables when considering application in *pharmaceutical market*. If the problem of *mood tracking to support medical diagnosis* is considered, the observables measure the *anxiety* or *weight* of a patient. However, the general idea of visual inspection and processing trends is the same regardless of the practical context.

Definition 3: ***Imprecise labeled sequence***
Let $f : \Psi_n \times S \times T \to [0, 1]$ *denote function assigning the degree of truth that label* $s \in S$ *applies at the moment* $t \in T$ *for object's property measured by time series* $y \in \Psi_n$*. Imprecise labeled sequence* $\{x_t^{s,y}\}_{t=1}^n$ *is calculated from* $x_t^{s,y} := f(y, s, t)$*.*

Fig. 1 presents an illustrative example of sales time series and its imprecise labeled sequence. In this context, for each observation of time series representing sales of Product A, the expert subjectively assigned the degree of truth for *decreasing* trend. Imprecise labeled sequences are processed within the presented approach.

III. Bayesian regression with linguistic knowledge

The forecasting procedure consists of the phase of processing temporal data and the posterior simulation. As outlined in Fig. 2 the input for the model are discrete time series and definitions of the linguistic concepts. As a result of the model, the forecast and its regressive components are provided.

A. Processing temporal data

Let $Y_n^k = \{\{y_t^1\}_{t=1}^n, ..., \{y_t^k\}_{t=1}^n\}$ denote k-vector of multivariate discrete time series. Let $\{y_t^1\}_{t=1}^n$ denote a time series of object's property to be predicted. For clarity reasons, we limit considerations to the one-step-ahead forecast and the vector of interest ω contains one element $\omega = \{y_{n+1}^1\}$. Predictions for longer horizons are iterated by repeating the procedure.

For $s \in S$ and $k-1$ time series $y \in \{\{y_t^2\}_{t=1}^n, ..., \{y_t^k\}_{t=1}^n\}$, imprecise labeled sequences $\{x_t^{s,y}\}_{t=1}^n$ are created based on data mining techniques or as a result of subjective expertise. Sequences $\{x_t^{s,y}\}_{t=1}^n$ interpreted as degree of truth that the imprecise label is valid at each moment for given object's property, represent the linguistic knowledge for the regression model.

B. Posterior simulation

Imprecise labeled sequences $\{x_t^{s,y}\}_{t=1}^n$ are included into the linear regression as explanatory variables. We adopt the multiple normal linear regression model which can be written as:

$$y = X\beta + \epsilon, \ \epsilon \sim N(0, \sigma^2 I_n)$$

where X is the $n \times ((k-1) \times l)$ matrix of explanatory variables, y is the $n \times 1$ vector of dependent variables and ϵ is an $n \times 1$ vector of independent identically distributed normal

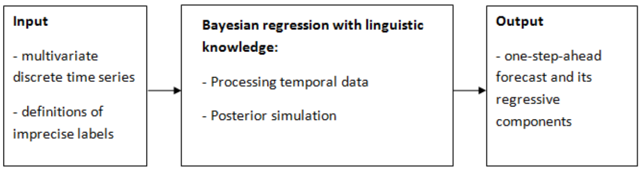

Fig. 2. Overview of the forecasting procedure based on Bayesian regression model with temporal linguistic knowledge.

random variables. We perform Bayesian inference to estimate the vector of parameters $\theta = (\beta, \sigma)$.

Following definition of Geweke [4], [5] the complete model A for Bayesian inference consists of:

1) the observables density

$$p(Y_t|\theta_A, A) = \prod_{t=1}^{T} p(y_t|Y_{t-1}, \theta_A, A)$$

in which $\theta_A \in \Theta_A$ is a $k_A \times 1$ vector of unobservables

2) the prior density $p(\theta_A|A)$

3) the vector of interest density (the posterior density)

$$p(\omega|y^o, A) = \int_{\Theta_A} p(\omega|y^o, \theta_A, A)p(\theta_A|y^o, A)\mathrm{d}v(\theta_A)$$

$$p(\theta_A|y^o, A) = \frac{p(\theta_A|A)p(y^o|\theta_A, A)}{p(y^o|A)}$$

The problem statement is to find a decision, known in Bayesian theory as an action a, which minimizes the following equation:

$$E[L(a, \omega)|y^o, A] = \int_{\Theta_A} L(a, \omega)p(\omega|y^o, A)\mathrm{d}\omega$$

Posterior predictive distributions are approached by means of Markov Chain Monte Carlo Methods (MCMC). Posterior simulation yields a pseudo-random sequence of the vector of interest to estimate its posterior moments. MCMC were initially developed in 1940s and gained popularity thanks to their great success in practical applications [5].

For simplicity, we assume that prior for β is independent from prior for σ and we apply Gibbs Sampling which leads to sampling from multivariate probability density. The sampling procedure begins with arbitrary values for β^0 and σ^0, computes mean and variance of β^0 conditional on the initial value σ^0, uses the computed mean and variance to draw a multivariate normal random vector β^1 and uses the β^1 with a random draw to determine σ^1.

Details for posterior density, drawings construction and sampling algorithm are available in [5]. Gelfand and Smith [3] proved that with Gibbs Sampler large sets of draws converge in the limit to the true joint posterior distribution of parameters.

The results for the linear regression model with linguistic knowledge are the predictive distribution for the future observations of the time series of interest and the model parameters. Parameters are easily interpreted in a natural language as they are directly linked with imprecise labels.

IV. Experimental results

The purpose of this experiment is to illustrate the performance of the forecasting method for real-life data at the example of sales time series from the pharmaceutical industry.

Train dataset consists of 6 normalized time series representing monthly sales of different products in the period from Jan'05 to Dec'09. Fig. 3 shows exemplary time series from the train dataset. The test dataset contains 6-month-long sales continuation for each product and is used for evaluation.

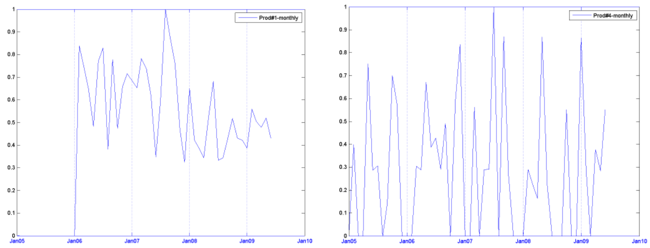

Fig. 3. Exemplary time series from the train dataset representing monthly sales of Product No.1 and Product No. 4.

We consider following 6 imprecise labels in the experiment: *low, medium, high, increasing, constant, decreasing*. Values for imprecise labeled sequences referring to *increasing, constant, decreasing* labels are defined based on experts' subjective beliefs. For labels: *low, medium, high* triangular fuzzy numbers are constructed based on the minimum, average and maximum values calculated from the time series. Then, imprecise labeled sequences are calculated from appropriate membership functions.

We first analyze correlations between the time series to be predicted and the imprecise labeled sequences to verify whether the imprecise linguistic knowledge may bring valuable information in the linear regression model. For each product we compare Pearson correlation coefficient between its sales time series and the imprecise labeled sequences derived for other products.

As demonstrated by the results in Table II correlations between sales time series and imprecise labeled sequences range from 0,10 to 0,14 and are on average by 20% higher than between different sales time series itself. Correlation coefficients are on average higher for labels of imprecise trends than values.

TABLE II
Correlation coefficients between sales time series and imprecise labeled sequences (ILS)

	Mean	Median	StdDev
Sales vs Sales	**0,10**	**0.08**	**0.05**
Sales vs Increasing	0,14	0.14	0.04
Sales vs Constant	0,12	0.13	0.06
Sales vs Decreasing	0,13	0.10	0.07
Sales vs Low	0,11	0.10	0.05
Sales vs Medium	0,12	0.12	0.05
Sales vs High	0,10	0.08	0.04
Sales vs All ILS	**0,12**	**0.11**	**0.05**

Table III provides detailed correlations per product. It is interesting to observe that for example the correlation coefficient (0,21) between sales time series of Product No. 2 and decreasing trends of other products is higher than correlation coefficient (0,17) between Product No. 2 sales time series and other sales time series itself.

TABLE III
CORRELATION COEFFICIENTS PER PRODUCT

	P1	P2	P3	P4	P5	P6
Sales vs Sales	0,16	**0,17**	0,05	0,09	0,04	0,07
Sales vs Increasing	0,17	0,18	0,10	0,11	0,16	0,11
Sales vs Constant	0,13	0,21	0,12	0,13	0,03	0,07
Sales vs Decreasing	0,22	**0,21**	0,09	0,10	0,06	0,09
Sales vs Low	0,19	0,17	0,11	0,07	0,06	0,09
Sales vs Medium	0,19	0,14	0,16	0,05	0,09	0,08
Sales vs High	0,13	0,15	0,06	0,09	0,07	0,08

The second step of the experiment is the comparative analysis of the forecasts' accuracy of the Bayesian regression model with linguistic knowledge (BRLK) and the traditional Vector Autoregression (VAR). Table IV summarizes mean absolute percentage error (MAPE) and deviation (MAPD).

TABLE IV
MEAN ABSOLUTE PERCENTAGE ERROR(MAPE) AND DEVIATION(MAPD) FOR 6- AND 1-STEP-AHEAD FORECAST OF BRLK AND VAR

	h=6 MAPE BRLK	h=6 MAPE VAR	h=6 MAPD BRLK	h=6 MAPD VAR	h=1 APE BRLK	h=1 APE VAR
P1	0,173	0,199	0,101	0,129	0,036	0,129
P2	0,425	0,509	0,193	0,165	0,202	0,317
P3	0,696	0,476	0,210	0,278	0,900	0,745
P4	0,282	0,396	0,371	0,237	0,026	0,326
P5	0,389	0,459	0,317	0,177	0,188	0,364
P6	0,444	0,371	0,281	0,222	0,703	0,559
All	**0,402**	**0,402**	**0,246**	**0,201**	**0,342**	**0,407**

As demonstrated by results in Table IV absolute percentage error for 1-step-ahead forecast is 0,342 and 0,407, respectively for BRLK and VAR. For 6-month-long forecast MAPE is the same and relatively high for both models, and amounts to 0,402. Forecasts generated by VAR are characterized by a lower standard deviation.

We conclude that Bayesian regression model with linguistic knowledge and VAR models are comparable in terms of forecasts' accuracy. The Bayesian regression model with linguistic knowledge delivers forecasts of a higher interpretability than traditional VAR as its components are naturally linked with the linguistic concepts.

V. CONCLUSION

The performed experiment confirmed that the approach with additional linguistic knowledge is adequate to support sales forecasting. Imprecise labeled sequences enable to discover new correlations in the dataset that lead to construction of the linear regression model. Produced forecasts are accurate on a similar level as forecasts provided by Vector Autoregression.

The main advantage of the proposed solution is the easy interpretation of predictions and model parameters required for forecasting process, which is of special importance for experts involved in real-life forecasting for large datasets. The proposed solution is in line with visual pattern recognition capabilities of humans and delivers additional knowledge about dependencies in multivariate time series datasets.

Next experiments for multivariate time series from other domains and on benchmark data are planned in order to analyze further benefits and limitations of the proposed technique. Another topic planned to be explored is the introduction of multiple interpretation for imprecise labels.

Within the approach simple forms of linguistic knowledge are considered. The potential to include advanced forms of linguistic knowledge like imprecise features, frequent temporal patterns, association rules and temporal linguistic summaries remains open for future research.

ACKNOWLEDGMENT

Katarzyna Kaczmarek is supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing. Project financed from the European Union within the Innovative Economy Operational Programme 2007-2013 and European Regional Development Fund.

REFERENCES

[1] F. Attneave, "Some informational aspects of visual inspection," *Psychological Review*, 61(3), 1954.

[2] G. Das, K. Lin, H. Mannila, G. Renganathan, and P. Smyth, "Rule discovery from time series," *Proc. of the 4th Int. Conf. on Knowl. Discovery and Data Mining*, 1998, pp. 16-22.

[3] A. Gelfand and A. Smith, "Sampling-based approaches to calculating marginal densities," In *Journal of the American Statistical Association*, Vol. 85, 1990, pp. 398-409.

[4] J. Geweke, *Contemporary Bayesian econometrics and statistics*. Wiley series in probability and statistics. John Wiley, 2005.

[5] J. Geweke and C. Whiteman. *Handbook of Economic Forecasting*, volume 1, chapter Bayesian Forecasting, Elsevier, 2006, pp. 3-80.

[6] M. Gil and O. Hryniewicz. "Statistics with imprecise data," In *Encyclopedia of Complexity and Systems Science*, Springer, 2009, pp. 8679-8690.

[7] F. Höppner. *Knowledge Discovery from Sequential Data*. PhD thesis, 2002.

[8] A. Wilbik and J. Kacprzyk. "Temporal sequence related protoform in liguistic summarization of time series," *Proc. of WConSC*, 2011.

[9] S. Kempe, J. Hipp, C. Lanquillon, and R. Kruse, "Mining frequent temporal patterns in interval sequences," *Fuzziness and Knowledge-Based Systems* in International Journal of Uncertainty, Vol. 16, No. 5, 2008, pp. 645-661.

[10] F. Klawonn and R. Kruse, "Derivation of fuzzy classifcation rules from multidimensional data," *Advances in Intelligent Data Analysis* Windsor, Ontario., 1995, pp. 90-94.

[11] F. Mörchen, I. Batal, D. Fradkin, J. Harrison and M. Hauskrecht "Mining recent temporal patterns for event detection in multivariate time series data," *KDD*, 2012, pp. 280-288.

[12] S. Schockaert and M. De Cock "Temporal reasoning about fuzzy intervals," *Artificial Intelligence* 172, 2008, pp. 1158-1193.

[13] A. P. Witkin, "Scale space filtering," *Proc. of the 8th Int. Joint Conf. on Artifficial Intelligence*, Karlsruhe, Germany, 1983, pp. 1019-1022.

[14] L. A. Zadeh, "From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions," In *Intelligent Systems and Soft Computing*, volume 1804 of *Lecture Notes in Computer Science*, Springer, 2000, pp. 3-40.