

Segmenting Time Series for Weather Forecasting

Somayajulu G. Sripada, Ehud Reiter, Jim Hunter, and Jin Yu
Dept of Computing Science
University of Aberdeen
Aberdeen, UK
{ssripada, ereiter, jhunter, jyu}@csd.abdn.ac.uk

Abstract

We are investigating techniques for producing textual summaries of time series data. Deep reasoning techniques have proven impractical because we lack perfect knowledge about users and their tasks. Data analysis techniques such as segmentation are more attractive, but they have been developed for data mining, not for communication. We examine how segmentation should be modified to make it suitable for generating textual summaries. Our algorithm has been implemented in a weather forecast generation system.

1. Introduction

Textual summaries of time series data, either independently or in conjunction with graphical presentation, aid human understanding of the underlying data sets. SUMTIME is a research project aiming to develop a generic model for summarisation of time series data. Initially, we have applied standard knowledge acquisition (KA) techniques such as think aloud sessions and corpus studies [1] to study how humans carry out the task. These KA studies helped us in specifying the requirements for data summarisation.

We found three alternative ways in which we could model data summarisation. One approach is based on deep reasoning methods suggested by [2]. This model explicitly reasons with user tasks and knowledge to determine information to include in the text summary. But in practice this is not attractive for two reasons. The first reason is that deep reasoning methods are computationally intensive. The second is that deep reasoning methods require precise knowledge of the user's tasks and goals. However, our KA studies have shown that it is not possible to get such knowledge. For example, meteorologists might know that the user wants to carry out supply boat operations alongside the rig structure, but would only know broadly the operational procedures followed by the oil company staff to carry out this task, or the precise operational characteristics of specific supply boats.

Alternatively, a model of data summarisation using existing models of data analysis is feasible. However, as explained in the paper, such a model does not agree with the observations made in our KA studies of humans summarising data. Therefore, in SUMTIME we have tried to adapt existing data analysis models so that the

adapted model agrees with our KA observations. The result is a model, which has some sensitivity to end-user needs but does not reason from first principles.

There are two perspectives to the current work. From the perspective of communicating summaries to end users, this work offers the alternative approach of adapting existing data analysis techniques to the approach of reasoning from first principles. From the perspective of data analysis, the work addresses the issues involved in adapting the existing techniques for use in a communicative context.

2. Background

SUMTIME is an ongoing research project aiming to develop generic techniques for producing textual summaries of time series data. We focus our study on time series data derived from the domains of meteorology and gas turbines. In the domain of meteorology, time series data produced by numerical weather prediction (NWP) models is summarised as weather forecast texts. In the domain of gas turbines, sensor data from an operational gas turbine is summarised for the maintenance engineers. More details on SUMTIME have been described in [3].

day	hour	wind dir	wind speed (Knots)
20-1-01	6	S	4
20-1-01	9	S	6
20-1-01	12	S	7
20-1-01	15	S	10
20-1-01	18	S	12
20-1-01	21	S	16
21-1-01	0	S	18

Table 1: Wind predictions from numerical model on 20-01-01

Table 1. shows an example time series from the domain of meteorology. It shows the wind speed and wind direction data predicted by the numerical model for the forecast period 0600-2400 hours on 20 Jan 2001. It is a time series sampled every three hours.

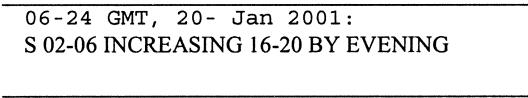


Figure 1. Wind texts from human written forecast for 20-01-01

Figure 1. shows the human written summary for the data shown in Table 1. The wind text is part of a marine forecast issued by our collaborating organisation, WNI/Oceanroutes, Aberdeen for offshore oil company staff. There are a number of tasks performed by the oil company staff, such as flaring the excess gas and carrying out supply boat operations that depend upon the condition of the weather. Marine forecasts are required to keep the oil company staff informed about the