A New Architecture for Summarising Time Series Data

Jin Yu, Ehud Reiter, Jim Hunter, and Somayajulu G. Sripada
Department of Computing Science
University of Aberdeen
Aberdeen, AB24 3UE, UK
{jyu, ereiter, jhunter, ssripada}@csd.abdn.ac.uk

Abstract

This paper presents a new architecture for summarising complex time series data, in which four main components together with a knowledge base and a database are integrated. Based on the architecture, a knowledge-based text generation system has been implemented and its main functions are briefly explained in the context of a sample of data. Evaluation of the system has been done and some conclusions are obtained.

1 Introduction

Although there are some natural language systems that produce English textual summaries of time series data such as FoG [1], ANA [2], StockReporter[1] , etc., it is still a challenge to summarise complex time series data. SumTime-Turbine (one of the systems developed in the SumTime Project[2]) summarises sensor data from gas turbines. This is challenging because of the large amount of data being summarised; a typical gas turbine has 250 analogue data channels that are sampled once per second. SumTime-Turbine uses a new architecture, which integrates pattern recognition, knowledge-based temporal abstraction (KBTA) [3], data mining and natural language generation techniques to produce summaries of such data. A short extract from SumTime-Turbine's input data is shown in Table 1. Typical output from the system is shown in Figure 1. This includes *background information,overview information* and *most significant patterns*.

[1]http://www.ics.mq.edu.au/ ltgdemo/StockReporter/ On Web November 2003
[2]More details in http://www.csd.abdn.ac.uk/research/sumtime/

Table 1: Part of a sample of time series data input to SUMTIME-TURBINE

The first column is the date in the format of dd/mm/yy, the second column is the time in the format of hh/mm/ss, and the other columns are channel names and their values (unit: Celsius). TTXD-1 – TTXD-6: exhaust temperature thermocouple # 1 – # 6.

Date	Time	TTXD-1	TTXD-2	TTXD-3	TTXD-4	TTXD-5	TTXD-6
27/11/99	12:02:19	430.56	450.88	429.27	481.40	452.13	463.38
27/11/99	12:02:20	430.98	451.15	429.27	481.68	451.99	463.93
27/11/99	12:02:21	430.84	451.29	429.59	481.09	452.27	463.80
27/11/99	12:02:22	431.12	451.15	429.27	481.40	452.27	464.21
27/11/99	12:02:23	431.25	451.85	429.59	481.68	452.27	464.35
27/11/99	12:02:24	431.53	452.40	429.00	481.40	451.85	464.07
27/11/99	12:02:25	431.39	451.57	429.27	481.40	451.71	463.66
27/11/99	12:02:26	431.25	451.57	429.41	481.26	451.99	463.52
27/11/99	12:02:27	430.28	449.73	429.14	480.53	451.85	464.07
27/11/99	12:02:28	430.00	449.59	429.73	480.81	452.54	464.07
27/11/99	12:02:29	430.56	450.74	429.59	480.95	452.54	464.21
27/11/99	12:02:30	430.56	450.74	429.41	481.68	452.40	463.80
27/11/99	12:02:31	430.56	451.15	429.27	481.09	452.40	463.38
27/11/99	12:02:32	431.25	451.85	429.41	481.82	452.27	463.52
27/11/99	12:02:33	431.39	452.40	430.14	481.82	452.27	463.66
27/11/99	12:02:34	431.25	452.13	430.28	481.82	452.82	464.80
27/11/99	12:02:35	430.98	450.88	429.59	481.96	451.99	464.21
27/11/99	12:02:36	430.28	450.15	429.00	481.82	452.13	463.38
...	...	....					

```
[Background information]
Gas turbine:aylesford
Subsystem:  Exhaust temperature
Monitoring channels:  TTXD-1, TTXD-2, TTXD-3, TTXD-4, TTXD-5 and TTXD-6
Turbine running state:  part load
Time interval of these channels:  from 12 to 15 on 27 Nov 99

[Overview information]
There were large erratic spikes in all channels at 12:59, 13:01,
13:41 and 14:40.

[Most significant patterns]
At 12:59, there were large erratic spikes in TTXD-1, TTXD-2, TTXD-3,
TTXD-4, TTXD-5 and TTXD-6.  These patterns violated the pairs and
follows check.  In more detail, there were dips with oscillatory
recoveries in TTXD-3 and TTXD-4, followed 1s later by dips with
oscillatory recoveries in TTXD-1, TTXD-2, TTXD-5 and TTXD-6.  This
occurred between 12:59:17 and 12:59:54.
```

Figure 1: A summary of the sample data generated by the system

2 Architecture of SumTime-Turbine

From temporal data to textual summaries, there are four steps that correspond to four components in the architecture of the system. Each component contains different modules as described in Figure 2.

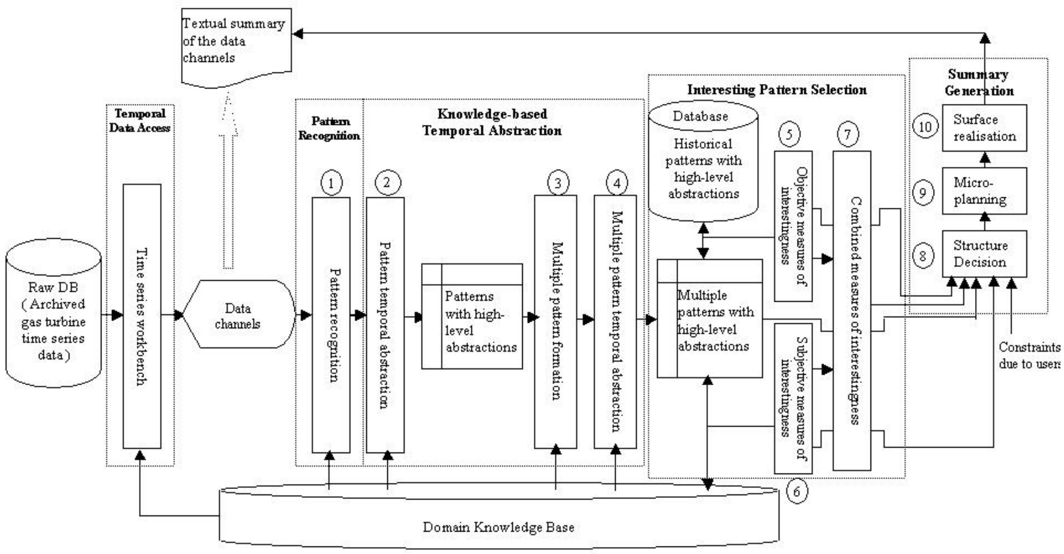

Figure 2: Architecture of SumTime-Turbine

The first step is *pattern recognition.* Patterns from time series data are located and classified by using a context sensitive landmark model and a shape definition language.

The second step is *KBTA adaptation.* Through adapting KBTA, high-level abstractions of the patterns or multiple-patterns in different contexts are obtained.

The third step is *interesting pattern selection.* In order to ensure the most important information to be communicated, two measures of interestingness, subjective and objective measures, are applied to choose the most significant patterns from the data.

The last step is *summary generation.* Through summary structuring, lexicalisation and aggregation, a concise summary with complex sentence structures is produced.

3 Conclusions

SumTime-Turbine has been evaluated at both the component and system levels, in a series of experiments conducted with domain experts at Intelligent Applications (IA). The evaluation results were encouraging, and IA felt that the system could be useful to them if it was further developed.

Summarising time series data involves many different, challenging tasks, only a few of which could be investigated in the system. Some limitations of the current system and future work have been drawn out as follows.

The current system generates summaries that only describe patterns occurring in the data. We would like to extend the system to generate summaries that interpret as well as describe patterns.

SumTime-Turbine is currently used to generate summaries of three-hour periods. It is desirable to generate summaries of time periods of any length as selected by the user.

There are lots of applications for description and/or interpretation of time series data in other domains such as medicine and finance. We believe that the architecture used in SumTime-Turbine could be applied to other domains.

More information on SumTime-Turbine is in [5].

References

[1] Goldberg, E., Driedger, N., Kittredge, R.,: Using Natural-Language Processing to Produce Weather Forecasts, IEEE Expert, **9** (1994) 45–53.

[2] Kukich, K.,: Knowledge-Based Report Generation: A Knowledge Engineering Approach to Natural Language Report Generation. Ph.D. Thesis, Information Science Department, University of Pittsburgh, 1983.

[3] Shahar, Y.,: A Framework for Knowledge-Based Temporal Interpolation. Artificial Intelligence, **90** (1997) 79–133.

[4] Milne, R., Nicol, C., Massuyes, L. T.,: TIGER with Model Based Diagnosis: Initial Deployment. Knowledge Based Systems, Elsevier, **14** (2001) 213–222.

[5] Yu, J.,: SumTime-Turbine: a Knowledge Based System to Generate English Textual Summaries of Gas Turbine Time Series Data. Ph.D. Thesis, Department of Computing Science, University of Aberdeen, 2004.