

A New Architecture for Summarising Complex Time Series Data

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

Natural Language Generation Group
 @ Department of Computing Science

Introduction

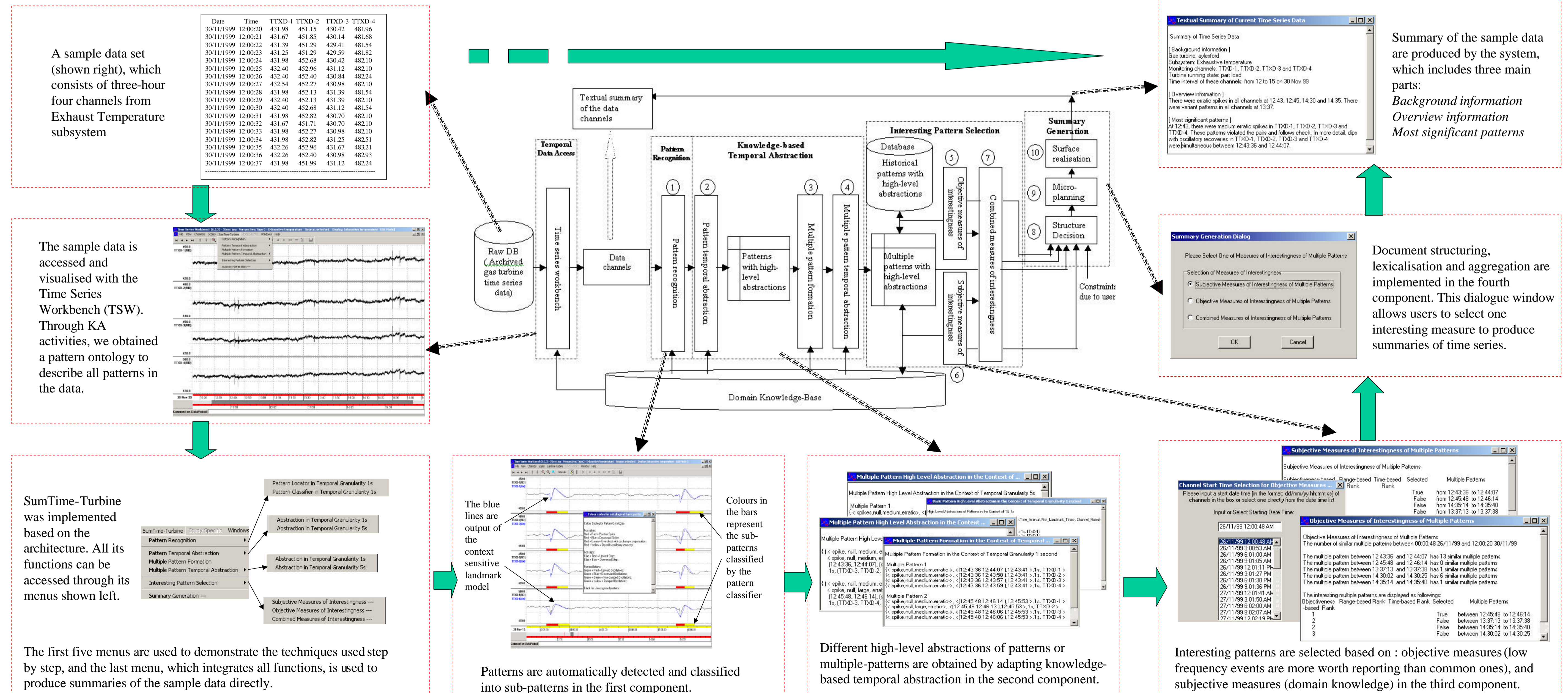
In the SumTime project, we developed generic techniques to produce English textual summaries of time series data, taking into account contextual information, by adapting recent work in AI and NLG. The project investigated three different domains, one of which is sensor data from gas turbines. A knowledge-based text generation system named SumTime-Turbine was implemented in the gas turbine domain based on the architecture presented in the poster.

From temporal data to textual summaries, there are four steps that correspond to the four components in the architecture: *pattern recognition*, *knowledge-based temporal abstraction*, *interesting pattern selection*, and *summary generation*. Each component contains different modules as shown in the figure.

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. But 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.

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 helpful to other researchers who are investigating how to generate textual summaries of complex temporal data in other domains.



A sample data set (shown right), which consists of three-hour four channels from Exhaust Temperature subsystem

The sample data is accessed and visualised with the Time Series Workbench (TSW). Through KA activities, we obtained a pattern ontology to describe all patterns in the data.

SumTime-Turbine was implemented based on the architecture. All its functions can be accessed through its menus shown left.

The first five menus are used to demonstrate the techniques used step by step, and the last menu, which integrates all functions, is used to produce summaries of the sample data directly.

Summary of the sample data, which includes three main parts:
 Background information
 Overview information
 Most significant patterns

Document structuring, lexicalisation and aggregation are implemented in the fourth component. This dialogue window allows users to select one interesting measure to produce summaries of time series.

The blue lines are output of the context sensitive landmark model
 Colours in the bars represent the sub-patterns classified by the pattern classifier

Patterns are automatically detected and classified into sub-patterns in the first component.

Different high-level abstractions of patterns or multiple-patterns are obtained by adapting knowledge-based temporal abstraction in the second component.

Interesting patterns are selected based on : objective measures (low frequency events are more worth reporting than common ones), and subjective measures (domain knowledge) in the third component.