Summarizing Time-Varying Data

Sarah Boyd
Microsoft Research Institute, Macquarie University
Sydney, NSW, Australia 2109
sarahb@mpce.mq.edu.au

In generating textual summaries of data, the content determination problem is even more complicated when summarizing time-varying data, such as in weather or stockmarket report generation. As well as the maximum, minimum and mean, what is of interest is the behaviour of the variable over time; e.g. dramatic changes, trends and degree of variability in the data. For example, in the graph of temperature shown in Fig. 1, in addition to the max, min and mean of this data we would be interested in commenting on dramatic temperature changes such as the drop between the 25th and the 28th. The techniques of wavelet analysis and

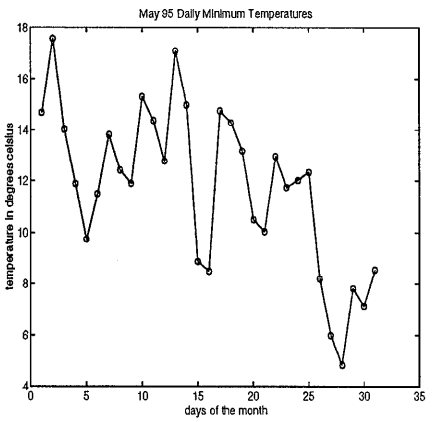

Figure 1: **May Daily Minimum Temperature**

scale space filtering used by Bakshi and Stephanopoulos (1995), in chemical process modelling, have been adapted in this work and applied to the content determination task in summarizing time-varying data. Wavelets are a signal processing tool that provide a localisation of a signal's features in both time and frequency and can be used to detect features at different frequency scales. Scale Space filtering is a technique for selecting dominant features in a signal (Bakshi and Stephanopoulos 1995). These techniques can be used for content determination in summary generation systems which do such content determination as weather forecasts (Goldberg et al. 1994) and basketball game summaries (Robin and McKeown 1996); neither of these systems focus on content determination.

Implemented so far is a system which extracts the four most dramatic changes in a time-varying signal using wavelets and scale space filtering. The system was applied to ten sets of monthly temperature data. For the data in Figure 1, the four dramatic changes selected for description were: the combined fall and rise beginning on the 13th and ending on the 17th and the two temperature drops from the 2nd to the 5th and from the 25th to the 28th. The features chosen were then evaluated against descriptions of the same data written by a weather expert. Eighteen of the nineteen dramatic changes mentioned by the expert were detected, although forty features overall were chosen.

Further work involves identifying other features besides dramatic changes such as trends and degree of variability in the data. In addition, integration with sentence planning and linguistic realization components will be necessary to provide a wide range of natural language output. In tandem with system development an evaluation of these techniques is planned including a corpus analysis and comparisons with human-generated descriptions.

References

Bakshi, B. R. and G. Stephanopoulos 1995. Reasoning in Time *Advances in Chemical Engineering* Vol. 22. Academic Press, New York.

Goldberg, E.; Driedger, N.; and Kittredge, R. 1994. Using Natural-Language Processing to Produce Weather Forecasts. In *IEEE Expert* Vol 9. No. 2.

Robin J. and K. McKeown 1996. Empirically Designing and Evaluating a New Revision-Based Model for Summary Generation In *Artificial Intelligence* Vol 85.