Overview of Original Software

Activity is central to thought and cognition. Through interaction autonomous agents build working representations of the environment they inhabit (Trajkovski 2007).

TamTam is a software demo based on interactionist principles (Bickhard 1980) based on unsupervised learning (Buisson 2006). This applet was developed to recognize and anticipate rhythmic patterns entered via a computer keyboard. TamTam always starts with a basic set of rhythms (e.g., four full-notes), and then uses a sophisticated algorithm for generating more and more complicated child patterns, based on the previously recognized patterns input by the user.

Methods

One of the issues when working with the code was that it was written as a java applet (see [Buisson 2003] for details), with all of the classes residing in one file, and a subsequent reliance on global variables. Based on the TamTam code, we developed PatternDiviner, a software suite for pattern mining in gene sequences. The core of the product is the pattern recognition engine (PRE), that relies on a multitude of other independent classes and interfaces, including Note.java, PatternBuilder.java,
PatternDisplay.java, Sequence.java, SequenceCanvas.java, Stroke.java, TamTamPanel.java, and TouchPanel.java. PatternDisplay is an interface, used by TamTamPanel, which implements the method `updatePatternDisplay(Sequence)`. This method is used to explicitly display the results of the PRE. In the case of the original TamTam applet, TamTamPanel displays a graphical sequence of notes, based on the pattern being fed to it by the user, as output in the applet. The core pattern recognition code was placed in PatternBuilder, which is the PRE of this system. See Figure 2 for the UML diagram of the essential classes.

Figure 2: UML Diagram

A separate class, DNAPatterns.java, implements PatternDisplay, which loads up the sample dataset and feeds it to the Sequence class. We used the following conversion between notes representation in TamTam and the base nucleotides as follows: adenine (abbreviated A, equivalent to a full note), cytosine (C, half note), guanine (G, 1/8 note) and thymine (T, 1/8 note). We then iterated through the gene sequence in set sizes of 4 genes through 22 genes, as customary in bioinformatics data mining when studying nucleotide sequence interaction.

For example, given a partial gene sequence of ‘AGGGTGCGCA AATTGGCGCA …’, the first round of sequences would be ‘AGGG’, ‘GGGT’, ‘GGTG’, ‘GTGC’, etc. Any patterns recognized by the PRE are stored by the engine internally.

Work in Progress
The results of running the Salmonella gene sequence (NCBI GeneBank 2006) through the engine were negative. The key sequence lengths we were interested in were between 18 and 22 genes. The PRE was able to recognize patterns of up to five genes in length, but nothing further.

There are several possible reasons for the negative result. The first, and most obvious, is that there may simply be no patterns to be recognized. A further issue might be the particular algorithm for generating the child patterns off of the winning parent patterns. A winning pattern of, for example, ‘atat’ might generate a child pattern of ‘atgat’, which may very well not be a larger pattern within the sequence. Finally, larger gene sequence datasets should be tested against the tool. As well as different types of data (meteorological, traffic-flow patterns in various urban areas, etc.).

This is a work in progress. More research into alternative pattern generation schemes might be worthwhile. Also, it would be interesting to further develop the PRE into a more abstract, and extensible, class. With a little work, a base class could be developed that managed some of the basic pattern recognition, and then let any sub-classes implement the specific pattern recognition algorithms needed by the developer. Using a basic Factory pattern,
TamTam could be used in a variety of environments, easily modifiable and testable.

Acknowledgements
This work was assisted generously by Dr Jean-Christophe Buisson, and was partially funded by the Faculty Research and Development Committee of Towson University.

References

2) TamTam, Retrieved online on June 5, 2006 at http://diabeto.enseeiht.fr/tamtam/, Dr. Jean-Christophe Buisson of L’Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique, d'Hydraulique et des Télécommunications (http://enseeiht.fr/)


