23 Redundant Hash Addressing of Feature Sequences using the Self-Organizing Map

Panu Somervuo

Temporal sequences arise from various kinds of sources in the nature. Sensory elements transform the events into measurements and corresponding feature vectors. The present work addresses the question of how to efficiently process the feature sequences. Applications include retrieval, error correction, and recognition of sequential data. Due to the durational differences in the feature sequences and the variation and noise in the feature vectors, both temporal and spatial fluctuations must be tolerated in the sequence comparison. Dynamic programming (DP) based methods provide solutions for this, but they can be computationally heavy. A different approach is to use local fixed-sized features of the sequence. This facilitates the use of fast associative methods.
The present work combines two methods developed by Teuvo Kohonen. These are Redundant Hash Addressing (RHA) [1, 3] and the Self-Organizing Map (SOM) [3]. The central idea in the RHA is to extract multiple features from the same input item. The comparison of the input item against the reference items is based on these features. In case of character strings, segments of N consecutive letters (N-grams) have been used. The RHA system consists of the N-gram table and the dictionary, see Fig. 39. Multiple features (N-grams) are extracted from the input string and each extracted N-gram associates the input string with the dictionary items.

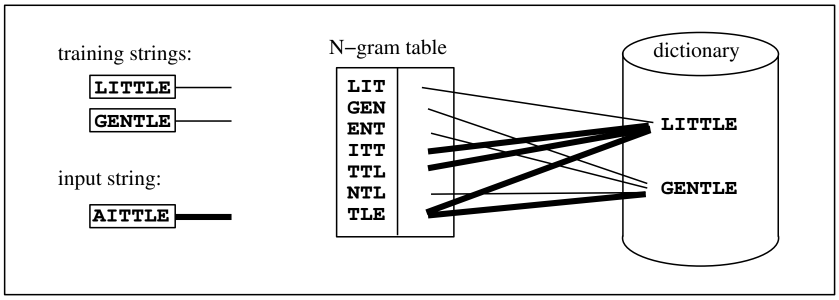

Figure 39: The RHA principle applied to character strings. The N-gram table is constructed by extracting the N-grams from training strings. Here trigrams are used ($N = 3$). From each item in the N-gram table there are associations (pointers) to the dictionary items. Associations activated by the erroneous input string 'AITTLE' are depicted by thick lines.

RHA has mainly been used for correcting textual output from speech recognizers, see e.g. [2]. In these, the recognition result is already in the form of a phoneme string. But in some applications, e.g. music processing, it is more difficult to extract, or even define, the underlying symbol sequence. Nevertheless, the RHA principle can still be used. As the RHA makes use of the N-grams of symbols and therefore the feature vectors must first be quantized, the SOM is used as a codebook to map the input feature vectors into the finite set of prototype vectors. When each SOM node is provided with an index, feature vector sequences can be mapped into symbolic index sequences. Each feature vector is encoded by the index of its best-matching unit (BMU). The node indices of the SOM are thus the alphabet of the system.

Music retrieval and speech recognition experiments were carried out as a demonstration of the method. Mel-scaled spectrum vectors were computed from acoustic piano music samples. A 10-by-10-unit SOM was trained and the RHA table was constructed using 24 training sequences with duration of 10 seconds each. The test material consisted of 24 subsequences of the training sequences with the duration of one second. The beginnings of these subsequences were randomly chosen. As a result, for N=1, there were 2 erroneous retrievals, 4 completely correct retrievals, and 18 retrievals where the correct music piece shared a tie with an incorrect retrieval. For N=2, all retrievals were completely correct. Also for the values N=3, N=4, and N=5, all retrievals were completely correct.

The speech material used in the experiments consisted of 1760 utterances collected from 20 speakers (5 female speakers and 15 male speakers). The vocabulary was 22 Finnish command words. The recognition results are shown in Table 11.

feature	recognition method	error, per cent	time/ms
10-dim cepstrum	DTW $^{1)}$	3.4	78
BMU index	Levenshtein $^{2)}$	3.4	1060
BMU index	RHA N=2 $^{2)}$	6.6	5

Table 11: Multi-speaker speech recognition experiment. Averaged results of four independent runs. Time is the average recognition time for one input sequence. 1) one reference sequence per class, 2) 60 reference sequences per class.

Although the recognition accuracy of the RHA method is not as high as the accuracy using 10-dimensional cepstrum vectors and DTW, the recognition time is an order of a magnitude smaller. The BMU index sequences were matched against reference templates also by using the Levenshtein distance. This result shows that the recognition accuracy using BMU index sequences can be as good as using cepstrum vectors and DTW if multiple reference templates are used for each class, but this increases the recognition time considerably. A remarkable property of the RHA method is that adding new sequence templates to the recognition system does not slow its performance distinctly.

References

[1] T. Kohonen and E. Reuhkala, “A very fast associative method for the recognition and correction of misspelt words, based on redundant hash addressing”, in Proc. 4IJCPR, pp.807-809, Kyoto, Japan, Nov. 7-10, 1978.

[2] T. Kohonen, H. Riittinen, E. Reuhkala, and S. Haltsonen S, “On-line recognition of spoken words from a large vocabulary”, in INFORMATION SCIENCES, Vol. 33, pp. 3-30, 1984.

[3] T. Kohonen, Self-Organizing Maps, Springer Series in Information Sciences, Vol. 30. Springer, Heidelberg, 1995. Second ed. 1997.