CHAPTER V

CONCLUSION AND FUTURE WORK

Pseudometrics or subadditive distances are both theoretically and practically appealing. Asymptotically, pseudometric based k-NN has the chance to err no more than twice that of the Bayes classifier, and several implementations can be used to accelerate pseudometric based k-NN. We provide two frameworks for designing subadditive distance measures for time series, namely the *condensation* framework and the *shortcut distance* framework.

5.1 Condensation

Condensation of a distance is a new distance based on an existing one, the base distance. In the condensation framework one can get a subadditive distance by choosing an appropriate set of morph operations wrt. the base distance. In order to have a subadditive condensation of a distance d wrt. a set $\mathcal{M}$ of morph operations, one need to check that the three following conditions are valid,

1. the base distance d is subadditive,
2. the morphs $\mathcal{M}$ is complete,
3. $\mathcal{M}$ contracts d.

DTW can be regarded as a condensation, but since the morph operations involved are not appropriate, DTW is not subadditive. An existing pseudometric called ERP and its generalization can be constructed in this framework. For norm metric based condensations, one can do interpolation between time series in a way similar to linear interpolation.

By two tools we proposed, Theorem 3 and Lemma 1, we designed our example pseudometrics in a somewhat modular fashion. The condensations were developed and we just checked the three conditions of Theorem 3 without worrying that some expression involving infinity may trouble the possibility of real implementation. By Lemma 1, and maybe its modifications, for some well-behaved morphs we can reduce computations involving infinity to finite ones, and fortunately all the morphs in our example condensations are such well-behaved.

5.2 Shortcut Distance

The second framework based on shortcut distances is more general. Any shortcut of a distance is always subadditive. Moreover, any subadditive distance is an edit distance in disguise. A concrete definition of the shortcut of DTW and its algorithm are currently unknown. A more general form of an existing distance called ERP can also be constructed in the second framework. We fine tuned it by adding a penalty value to prevent too much *morphs* to match another time series. The fine tuned distance has potential to yield better classification results. All of the proposed distances can be computed in $O(mn)$ time like DTW. Numerical results showed that they are useful alternatives to DTW.

5.3 Future Work

Theorem 3 and Lemma 2 are in fact applicable to any space equipped with a pseudometric or a distance function, respectively. In this work we focused on distances of vectors and real valued sequences, and perhaps extension to tensors is the next step to study.

Fagin and Stockmeyer (1998) proposed that the triangle inequality can be weakened and distances satisfying the relaxed condition are still useful in practice. They also proposed a distance measure that does satisfy the *relaxed triangle inequality* although it does not satisfy the triangle inequality.

Given $\lambda \in \mathbb{R}$, a distance d is λ-subadditive if for every x, y, z,

$$d(x,y) \leq \lambda(d(x,z) + d(z,y)) \ .$$

We say that a distance satisfies the *relaxed triangle inequality* if it is λ-subadditive for some λ. Thus every subadditive distance satisfies the relaxed triangle inequality since a subadditive distance is 1-subadditive.

By a little modification to the proof of Theorem 3 we can have that the subadditivity condition of the base distance can be changed to λ-subadditivity and the resulting condensation is guaranteed to also be λ-subadditive.

It would be nice if the existing asymptotic results hold for k-NN with weakly subadditive distances with no major modifications to their proofs. We opine that this is likely the case since this weaker notion subadditivity is only a multiplicative constant away from subadditivity. The sought result may be in the form "Under some regular assumptions, if d is λ-subadditive with $\lambda < \theta$ then the k-NN classifier will do asymptotically good".