WHAT IS CLAIMED IS:

1. A method, comprising:

obtaining, with at least one processor, from at least one database, a plurality of known time series;

for each known time series of the plurality of known time series:

computing, with the at least one processor, a pairwise distance matrix between that known time series and each learned template of a plurality of learned templates to generate a plurality of pairwise distance matrices;

stacking, with the at least one processor, the plurality of pairwise distance matrices together to generate a tensor; and

processing, with the at least one processor, with a residual network, the tensor, wherein the residual network receives, as input, the tensor, and provides, as output, a feature vector for that known time series; and

providing, with the at least one processor, the feature vector for each known time series of the plurality of known time series.

2. The method of claim 1, further comprising:

storing, with the at least one processor, in the at least one database, the feature vector for each known time series of the plurality of known time series.

3. The method of claim 2, further comprising:

obtaining, with the at least one processor, an unknown time series;

computing, with the at least one processor, a pairwise distance matrix between the unknown time series and each learned template of the plurality of learned templates to generate a further plurality of pairwise distance matrices;

stacking, with the at least one processor, the further plurality of pairwise distance matrices together to generate a further tensor;

processing, with the at least one processor, with the residual network, the further tensor, wherein the residual network receives, as input, the further tensor, and provides, as output, a feature vector for the unknown time series;

for each known time series of the plurality of known time series stored in the database, determining, with the at least one processor, based on the stored

feature vector for that known time series and the feature vector for the unknown time series, a distance between that known time series and the unknown time series; and

identifying, with the at least one processor, based on the distance between each known time series and the unknown time series, at least one known time series determined to correspond to the unknown time series.

4. The method of claim 1, wherein the residual network is trained using a loss function defined according to the following Equation:

$$\sum_{(\mathrm{t}_i,\mathrm{t}_{i,+},\mathrm{t}_{i,-})\in\mathcal{B}} -\log\sigma(f_\theta\left(\mathrm{t}_i,\mathrm{t}_{i,+}\right)-f_\theta(\mathrm{t}_i,\mathrm{t}_{i,-}))$$

where $\mathcal{B}$ is a batch of training data $\mathcal{B}=[\mathcal{T}_0,\ldots,\mathcal{T}_m]$, m is a batch size, each sample in the batch is a tuple $\mathcal{T}_i$ including a query time series t_i, a positive time series t_{i+}, and a negative time series t_{i-}, $\sigma(\cdot)$ is a sigmoid function, and $f_\theta(\cdot,\cdot)$ is the residual network.

5. The method of claim 1, wherein the plurality of known time series includes a plurality of known transaction time series associated with a plurality of merchants, and wherein each known time series is associated with metadata associated with a merchant associated with that known time series.

6. The method of claim 1, wherein the plurality of learned templates includes thirty-two learned templates, wherein the plurality of pairwise distance matrices includes thirty-two pairwise distance matrices, wherein the tensor includes an input dimension of thirty-two, and wherein the feature vector for each known time series of the plurality of known time series includes a size sixty-four vector.

7. The method of claim 1, wherein the residual network includes a two-dimensional residual network.

8. A system, comprising:

at least one processor coupled to a memory and configured to:

obtain, from at least one database, a plurality of known time series;

for each known time series of the plurality of known time series:

compute a pairwise distance matrix between that known time series and each learned template of a plurality of learned templates to generate a plurality of pairwise distance matrices;

stack the plurality of pairwise distance matrices together to generate a tensor; and

process, with a residual network, the tensor, wherein the residual network receives, as input, the tensor, and provides, as output, a feature vector for that known time series; and

provide the feature vector for each known time series of the plurality of known time series.

9. The system of claim 8, wherein the at least one processor is further configured to:

store, in the at least one database, the feature vector for each known time series of the plurality of known time series.

10. The system of claim 9, wherein the at least one processor is further configured to:

obtain an unknown time series;

compute a pairwise distance matrix between the unknown time series and each learned template of the plurality of learned templates to generate a further plurality of pairwise distance matrices;

stack the further plurality of pairwise distance matrices together to generate a further tensor;

process, with the residual network, the further tensor, wherein the residual network receives, as input, the further tensor, and provides, as output, a feature vector for the unknown time series;

for each known time series of the plurality of known time series stored in the database, determine, based on the stored feature vector for that known time series and the feature vector for the unknown time series, a distance between that known time series and the unknown time series; and

identify, based on the distance between each known time series and the unknown time series, at least one known time series determined to correspond to the unknown time series.

11. The system of claim 8, wherein the residual network is trained using a loss function defined according to the following Equation:

$$\sum_{(\mathrm{t}_i,\mathrm{t}_{i,+},\mathrm{t}_{i,-})\in\mathcal{B}} -\log\sigma(f_\theta\left(\mathrm{t}_i,\mathrm{t}_{i,+}\right)-f_\theta(\mathrm{t}_i,\mathrm{t}_{i,-}))$$

where $\mathcal{B}$ is a batch of training data $\mathcal{B}=[\mathcal{T}_0,\ldots,\mathcal{T}_m]$, m is a batch size, each sample in the batch is a tuple $\mathcal{T}_i$ including a query time series t_i, a positive time series t_{i+}, and a negative time series t_{i-}, $\sigma(\cdot)$ is a sigmoid function, and $f_\theta(\cdot,\cdot)$ is the residual network.

12. The system of claim 8, wherein the plurality of known time series includes a plurality of known transaction time series associated with a plurality of merchants, and wherein each known time series is associated with metadata associated with a merchant associated with that known time series.

13. The system of claim 8, wherein the plurality of learned templates includes thirty-two learned templates, wherein the plurality of pairwise distance matrices includes thirty-two pairwise distance matrices, wherein the tensor includes an input dimension of thirty-two, and wherein the feature vector for each known time series of the plurality of known time series includes a size sixty-four vector.

14. The system of claim 8, wherein the residual network includes a two-dimensional residual network.

15. A computer program product including a non-transitory computer readable medium including program instructions which, when executed by at least one processor, cause the at least one processor to:

obtain, from at least one database, a plurality of known time series;

for each known time series of the plurality of known time series:

compute a pairwise distance matrix between that known time series and each learned template of a plurality of learned templates to generate a plurality of pairwise distance matrices;

stack the plurality of pairwise distance matrices together to generate a tensor; and

process, with a residual network, the tensor, wherein the residual network receives, as input, the tensor, and provides, as output, a feature vector for that known time series; and

provide the feature vector for each known time series of the plurality of known time series.

16. The computer program product of claim 15, wherein the program instructions, when executed by the at least one processor, further cause the at least one processor to:

store, in the at least one database, the feature vector for each known time series of the plurality of known time series.

17. The computer program product of claim 16, wherein the program instructions, when executed by the at least one processor, further cause the at least one processor to:

obtain an unknown time series;

compute a pairwise distance matrix between the unknown time series and each learned template of the plurality of learned templates to generate a further plurality of pairwise distance matrices;

stack the further plurality of pairwise distance matrices together to generate a further tensor;

process, with the residual network, the further tensor, wherein the residual network receives, as input, the further tensor, and provides, as output, a feature vector for the unknown time series;

for each known time series of the plurality of known time series stored in the database, determine, based on the stored feature vector for that known time series and the feature vector for the unknown time series, a distance between that known time series and the unknown time series; and

identify, based on the distance between each known time series and the unknown time series, at least one known time series determined to correspond to the unknown time series.

18. The computer program product of claim 15, wherein the residual network is trained using a loss function defined according to the following Equation:

$$\sum_{(\mathrm{t}_i,\mathrm{t}_{i,+},\mathrm{t}_{i,-})\in\mathcal{B}} -\log\sigma(f_\theta\left(\mathrm{t}_i,\mathrm{t}_{i,+}\right)-f_\theta(\mathrm{t}_i,\mathrm{t}_{i,-}))$$

where $\mathcal{B}$ is a batch of training data $\mathcal{B} = [\mathcal{T}_0, \ldots, \mathcal{T}_m]$, m is a batch size, each sample in the batch is a tuple $\mathcal{T}_i$ including a query time series t_i, a positive time series t_{i+}, and a negative time series t_{i-}, $\sigma(\cdot)$ is a sigmoid function, and $f_\theta(\cdot,\cdot)$ is the residual network.

19. The computer program product of claim 15, wherein the plurality of known time series includes a plurality of known transaction time series associated with a plurality of merchants, and wherein each known time series is associated with metadata associated with a merchant associated with that known time series.

20. The computer program product of claim 15, wherein the plurality of learned templates includes thirty-two learned templates, wherein the plurality of pairwise distance matrices includes thirty-two pairwise distance matrices, wherein the tensor includes an input dimension of thirty-two, and wherein the feature vector for each known time series of the plurality of known time series includes a size sixty-four vector, and wherein the residual network includes a two-dimensional residual network.