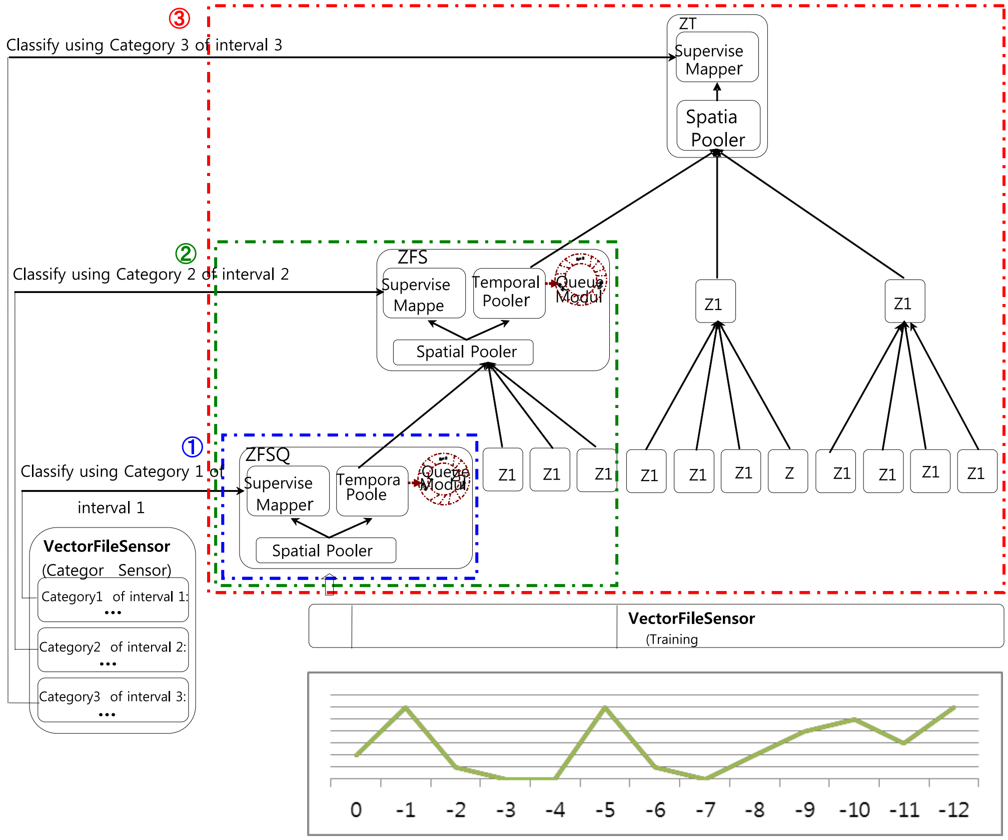

Fig. 6. Real-time integrated hierarchical temporal memory (RIHTM) network for the training stage. Z1=Zeta1Node, ZT=Zeta1TopNode, ZFSQ=Zeta1FirstSpecializedQueueNode.

In the first step of training (①), as shown in Fig. 6, a historical data stream with data items recorded per minute (a data stream with a 1-minute interval) is input into the network. Only the ZFSQNode at level 1 is enabled to receive data items from the VectorFileSensor and gets into the learning mode. After all the learning data has been represented, the structures of the spatial pooler and the temporal pooler in the ZFSQNode are copied to the Zeta1Nodes at the same level. Entering into Step 2 (②) shown in Fig. 6, the ZFSQNode at level 2 gets into the learning mode. Four child nodes (ZFSQ and 3 Zeta1Nodes) at level 1 get into the inference mode. The ZFSQNode at level 2 receives the learning data stream from its four child nodes at a 4-mimute interval. After the ZFSQNode at level 2 has learned all of the data, the data structures are copied to the remaining Zeta1Nodes at level 2 as with the ZFSQNode at level 1. This procedure is repeated throughout the hierarchy until all of the nodes in the network are enabled. As shown in Fig. 6, the Zeta1TopNode at the top level is put into the learning mode (③). Its 3 child nodes at level 2 and 12 grandchild nodes at level 1 perform inference. The inference outputs of the nodes at level 1 are sent to the nodes at level 2, and then the inference outputs of the nodes at level 2 are sent to the Zeta1TopNode as learning data. The Zeta1TopNode is trained using the data stream at a 12-minute interval. During the whole training stage, the sQUEUE is disabled as it is used only in the prediction stage.

4.2 Prediction Stage

During the prediction stage, the RIHTM network uses an advanced prediction technique for RCMIP. Fig. 7 shows the structure of the RIHTM network for the prediction stage. It is obvious that the RITHM network has a simpler structure than the CIHTM network (depicted in Fig. 2) and that it is faster and easier to operate.

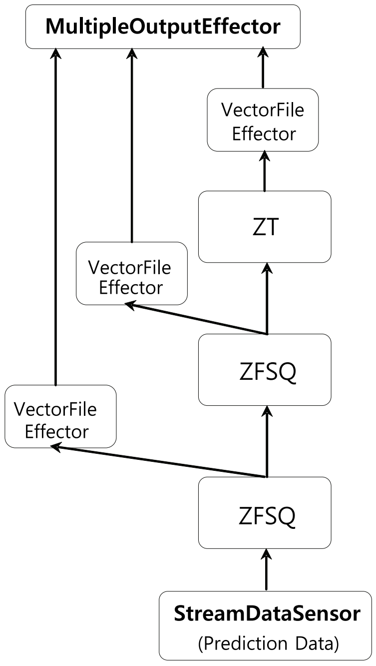

Fig. 7. The real-time integrated hierarchical temporal memory (RIHTM) network for the prediction stage. ZT=Zeta1TopNode, ZFSQ=Zeta1FirstSpecializedQueueNode.

During the prediction stage, the data that is input into the trained network is a dynamic sequence of real-time data. To generate prediction results in the real-time mode, we used a StreamDataSensor to process real-time data streams in a timed sequence. StreamDataSensor plays the same role as the VectorFileSensor in the training stage. The only difference between them is that the StreamDataSensor contains only one sensor for real-time input data streams with a fixed interval (1 minute).

The ZFSQNode with its sQUEUE in each level and the Zeta1TopNode at the top level are enabled to expose new data with 1-minute, 4-minute, and 12-minute intervals, respectively, and to produce the prediction results for these three intervals simultaneously. The prediction results of these three intervals are received by the MultipleOutputEffector to determine whether these outputs are valid, and the MultipleOutputEffector generates the multi-interval prediction results of the network.

Fig. 8 shows different steps of the RIHTM network during the prediction stage.

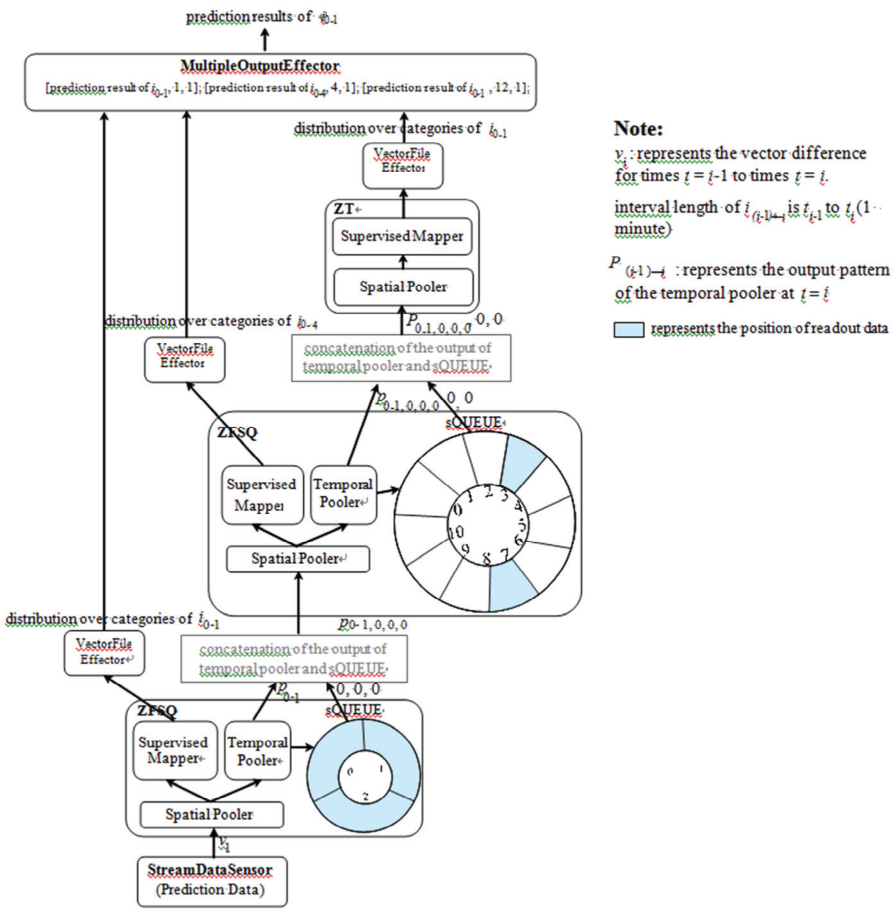

Fig. 8. Predictions by the real-time integrated hierarchical temporal memory (RIHTM) network in Step 1. ZT=Zeta1TopNode, ZFSQ=Zeta1FirstSpecializedQueueNode, sQUEUE=specialized circular queue.

In Fig. 8, the ZFSQNode at level 1 receives the sequence of input data with the 1-minute interval from StreamDataSensor. The input data v_i at t_i is the value difference from t_{i-1} to t_i. For instance, if the value is 8.01 at t_{i-1} and the value is 8.12 at t_i, then the difference value in the interval (i-1)-i from t_{i-1} to t_i, v_i is +0.11. The spatial pooler of the ZFSQNode compares this input vector to each coincidence that was learned during the training stage and computes a belief vector for its input vector. This belief vector is sent to the temporal pooler and supervised mapper simultaneously. The supervised mapper of the ZFSQNode receives this belief vector and produces a distribution over the categories as the 1-minute interval prediction result of level 1. Meanwhile, the temporal pooler receives the belief vector from the spatial pooler and calculates belief distribution over groups of intervals (i-1)-i, which is marked as $p_{(i-1)-i}$. The $p_{(i-1)-i}$ is a vector that is the same size as the number of temporal groups. For example, the node has 12 coincidences within its spatial pooler and four temporal groups within its temporal pooler. The present input matches the fourth coincidence, which belongs to temporal group 2. Therefore, the output of the temporal pooler $p_{(i-1)-i}$ is a vector of four groups with one match corresponding to group 2 and is represented as [0100]. The output $p_{(i-1)-i}$ of the temporal pooler is concatenated with the outputs of sQUEUE. The Size, Readoutnum, and Position of sQUEUE at level 1 in Fig. 8 is 3, 3, [0, 1, 2], respectively. If there is no data in the sQUEUE, each output of the queue is set to the default value 0. Therefore, the concatenation of the output from the temporal pooler and sQUEUE is p_{0-1}, 0, 0, 0, which is the input of the ZFSQNode at level 2.

This procedure is repeated up the hierarchy until the outputs of all the levels are valid. These multiple prediction results are used to predict the real-time data stream trends at the current time according to different intervals.

5. Performance Analysis

The RIHTM network is suitable for RCMIP. It is effective in tremendously reducing memory and time consumption when the number of prediction intervals is increased. In the following section, the performance analysis of memory and time consumption is discussed.

5.1 Analytical Expression

In this section we will describe the analytical expression for the estimation of memory and time consumption in the CITHM network and RIHTM network at the prediction stage.

5.1.1 Estimation of memory consumption

The spatial pooler, temporal pooler, and the supervised mapper are the key substructures of nodes that perform learning and inference. The spatial pooler analyzes the stream of input patterns and generates a coincidence matrix with a row for each coincidence and one column for each element in the input pattern. The memory consumption of spatial pooler is $m_{SP}(n_{coincs}, m_{patterns}) = n_{coincs} \times m_{pattern}$, where n_{coincs} is the number of coincidences generated by the spatial pooler and $m_{pattern}$ denotes the memory of an input pattern. The temporal pooler receives coincidence indices that are sent by the spatial pooler and forms non-overlapping groups of coincidences using a time-adjacency matrix with n_{coincs} rows and n_{coincs} columns. The memory consumption of temporal pooler is $m_{TP}(n_{coincs}) = n_{coincs} \times n_{coincs} + n_{coincs}$. The supervised mapper does not form groups, but instead forms a matrix with n_{coincs} rows and one column for each category that it encountered from the category sensor. The memory consumption of the supervised mapper is $m_{SM}(n_{categs}, n_{coincs}) = n_{categs} \times n_{coincs}$, where the n_{categs} denotes the number of categories.

A Zeta1Node consists of a spatial pooler and a temporal pooler. Therefore, the memory consumption of a Zeta1Node is $m_{Z1} = m_{SP} + m_{TP} = (n_{coincs} \times m_{pattern}) + (n_{coincs} \times n_{coincs} + n_{coincs})$. A Zeta1FirstNode consists of a spatial pooler, a temporal pooler, and a supervised mapper. Therefore, the memory consumption of a Zeta1FirstNode is $m_{ZF} = m_{SP} + m_{TP} + m_{SM} = (n_{coincs} \times m_{pattern}) + (n_{coincs} \times n_{coincs} \times m_{index} + n_{coincs} \times m_{index}) + (n_{categs} \times n_{coincs} \times m_{index})$. A Zeta1TopNode consists of a spatial pooler and a supervised mapper. The memory consumption of a Zeta1TopNode is $m_{ZT} = m_{SP} + m_{SM} = (n_{coincs} \times m_{pattern}) + (n_{categs} \times n_{coincs} \times m_{index})$.

The CIHTM network is constructed with all of these three types of nodes at the prediction stage. In the network, there is only one Zeta1FirstNode in each level and it always located in the first position and the remaining locations use the Zeta1Nodes. However, there is an exception where the top level has only one Zeta1TopNode. Therefore, the memory consumption of the CIHTM network with n intervals prediction at the prediction stage can be calculated by the following formula:

$$M_C = m_{ZT} + \sum_{i=1}^{n-1} m_{ZF(i)} + \sum_{i=1}^{n-1} (n_{Z1(i)} \times m_{Z1(i)}) \tag{4}$$

where m_{ZT} is the memory consumption of Zeta1TopNode; $m_{ZF(i)}$ is the memory consumption of Zeta1FirstNode at level i; $m_{Z1(i)}$ is the memory consumption of Zeta1Node at level i; and $n_{Z1(i)}$ denotes the number of Zeta1Nodes at level i in the network.

The RIHTM network is constructed with two types of nodes, where there is only one ZFSQNode at each lower level and one Zeta1TopNode at the top level of the prediction stage. Therefore, the memory consumption of the RIHTM network with n intervals prediction at the prediction stage can be calculated by the following formula:

$$M_R = m_{ZT} + \sum_{i=1}^{n-1} m_{ZFSQ(i)} \tag{5}$$

where $m_{ZFSQ(i)}$ is the memory consumption of ZFSQNode at level i. The ZFSQNode is an extension of the Zeta1FirstNode with an additional specialized circular queue. The specialized circular queue as a memory stores the output of the temporal pooler for the previous data items in each unit cell. The output of the temporal pooler is a vector that is the same size as the number of temporal groups. Therefore, the memory consumption of the queue is $m_Q = n_{groups} \times s_Q$, where s_Q is the size of the queue. The memory consumption of the ZFSQNode is $m_{ZFSQ} = m_{ZF} + m_Q = m_{ZF} + (n_{groups} \times s_Q)$. Therefore, the memory consumption of the RIHTM network with n intervals prediction at the prediction stage can also be shown with the following formula:

$$M_R = m_{ZT} + \sum_{i=1}^{n-1} (m_{ZF(i)} + m_{Q(i)}) = m_{ZT} + \sum_{i=1}^{n-1} m_{ZF(i)} + \sum_{i=1}^{n-1} m_{Q(i)} \tag{6}$$

5.1.2 Estimation of time consumption

The time consumption of a Zeta1FirstNode with a spatial pooler, a temporal pooler, and a supervised mapper is $t_{ZF}(t_{SP}, t_{TP}, t_{SM}) = t_{SP} + t_{TP} + t_{SM}$. The time consumption of a Zeta1TopNode with a spatial pooler and a supervised mapper is $t_{ZT}(t_{SP}, t_{SM}) = t_{SP} + t_{SM}$. The time consumption of a ZFSQNode with a spatial pooler, a temporal pooler, a supervised mapper, and a queue is $t_{ZFSQ}(t_{SP}, t_{TP}, t_{SM}, t_Q) = t_{SP} + t_{TP} + t_{SM} + t_Q$.

The time consumption of the CIHTM network with n intervals prediction at the prediction stage can be calculated by the following formula:

$$T_C = t_{ZT} + \sum_{i=1}^{n-1} t_{ZF(i)} + \sum_{i=1}^{n-1} (n_{Z1(i)} \times t_{Z1(i)}) \tag{7}$$