Probabilistic modeling and numerical simulation of neural circuits for multisensory integration

Liyu Xu[1], Xinsheng Liu[2,*]

School of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing, China

*Corresponding author: liyuxu@nuaa.edu.cn

Abstract. People in real life receive stimulus information through various senses, and the process by which the brain integrates this information is called multisensory integration. Multisensory integration is an important branch of neuroscience, and the research on its neural mechanism holds significant application value to the development of artificial intelligence such as designing intelligent robots. Researches suggests that the brain likely employs Bayesian rules to integrate information and make judgments. In machine learning, neural networks based on Spike-Timing-Dependent Plasticity (STDP) have shown promising results in multimodal emotion recognition. In this paper, we model a neural network based on STDP, try to explain spike events using a probabilistic model, and unify network output with Bayesian calculation. This paper uses numerical simulation to verify the performance of the proposed network in multisensory classification problems. The results show that multisensory integration can improve classification accuracy and is better than the popular supervised learning method.

Keywords: multisensory integration, spiking neural networks, STDP, Bayesian calculation.

1. Introduction

The human brain is an incredibly complex system, which is made up of about 90 billion neurons [1]. In the year 2000, Nobel laureate in Physiology or Medicine, Edleman, said that knowledge of brain science will influence our perception of the world. The complexity and efficiency of the human brain make it highly valuable for study [2], and thus, neuroscience is considered to be the focus of research in the 21st century, and one of the greatest challenges of scientific development today.

The human brain serves as the carrier of intelligence and information. To investigate the processes of the mechanisms of intelligence, it is essential to understand the neural mechanisms of the brain [3–6]. The process of combining cues from different sensory channel is referred to as multisensory integration [2,7–12]. Research into multisensory integration is an important subfield of neuroscience, as it helps determine the relationships between different sensory signals and enhances the detection and recognition of external events. [13,14]. In artificial intelligence, intelligent robots need to accurately locate objects by integrating image data and sensor data. Similarly, in autonomous driving, systems must integrate information from multiple sensors to ensure the accuracy of final decisions. Hence, the study of the neural mechanisms of multisensory integration holds significant practical value in the design of intelligent robots and other research areas related to artificial intelligence.

How the brain integrates inputs from different sensory modalities to produce accurate representation of object is a fundamental question in cognitive neuroscience [5]. Recent psychophysical experiments suggest that neurons not only represent probability distributions but also combine these distributions using methods close to Bayesian rules. Extensive behavioral studies indicate that the brain weights the reliability of different modality cues in an almost optimal statistical manner when integrating cross-modal cues [5,7]. Many researchers have proposed potential neural mechanisms for multisensory integration. For instance, Ma et al. introduced the theory of Probabilistic Population Coding (PPC), which holds that neuronal estimates of stimuli are not just single values but rather encoded by populations of neurons as probability distributions [7]. Fang et al. proposed that the core issue in multisensory integration is probabilistic estimation of stimuli. They introduced a unified probabilistic model that combines multisensory integration and causal inference, and implement it through importance sampling and probabilistic population coding simulations [6].

Recently, SNNs based on Spike-Timing-Dependent Plasticity (STDP) have performed exceptionally well in multisensory tasks. Rathi et al. proposed a multimodal SNNs that combines visual and auditory modes. They show that compare to single-modal networks, the multimodal network suppresses the impact of noise on classification accuracy [15]. Benssassi and colleagues investigated the influence of neural synchrony on multisensory integration in emotion recognition. The model was demonstrated performance comparable to state-of-the-art multimodal fusion techniques such as deep learning fusion [14].

Despite SNNs based on STDP get significant achievements in pattern recognition fields, we still have limited understanding of the connection between SNNs and the neural mechanisms of the brain, as well as the probabilistic interpretation of SNNs. Therefore, this paper attempts to establish a probabilistic model of a multisensory spiking neural network based on STDP. It provides a probabilistic interpretation of neurons and their spikes in the network. Through Bayesian analysis, the paper establishes a connection between the probabilistic generative model and the output spikes of the network. Finally, the classification performance of the multisensory network is verified through simulations on training and testing datasets. The results demonstrate that the classification performance of our network is better than that of single-modal network, and the unsupervised single-modal network exhibits better classification performance than the K-nearest neighbors (K-NN) algorithm.

2. The Network Model

2.1. The structure of spiking neural network model

In this paper, we use visual mode and auditory mode to model the multisensory network, as humans get most of the information from sight and hearing. Then we will describe the structure of the network and some definitions.

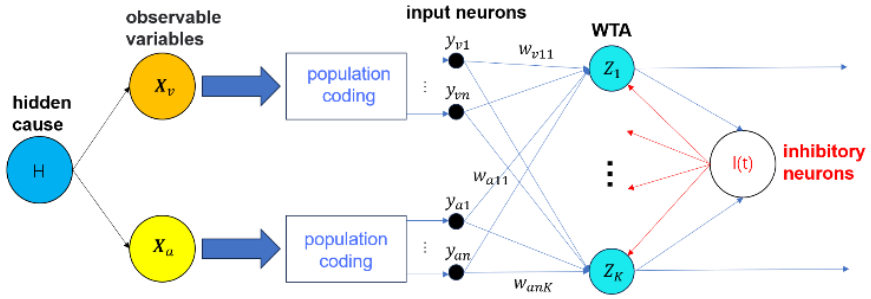

Fig. 1 Schematic representation of the Spike Neural Network model

Our network consists of two parts, one is the probabilistic generation model of a hidden cause, the other is the structure of the neural network model. The generation model in Fig.1 contains a hidden cause H and several observable variables. In a classification task, H represents an unobservable variable which has K different possible values, the visual observable variable $\boldsymbol{X}_v = (x_{v1}, \cdots x_{vm})$ and the auditory observable variable $\boldsymbol{X}_a = (x_{a1}, \cdots x_{am})$ is generated by H ,which contain the feature of H . In emotion recognition tasks, H represents different kinds of emotion, such as happy or sad. If H represents happy, $\boldsymbol{X}_v$ can be the image of a happy face, and $\boldsymbol{X}_a$ can be a happy word. The observable variables are encoded as excitatory neurons $y_1, \cdots y_N$ through population coding, which send information to the neural network.

The neural network is based on spike events, consisting of a set of spiking neurons arranged in a winner-takes-all (WTA) circuit and a set of input neurons. The WTA mechanism is one of the most commonly studied connection patterns in cortical microcircuits [16]. The input neurons consist of two types of excitatory neurons, the visual ones and the auditory ones. Each input neuron is connected to excitatory neurons in the WTA circuit through synapse, which represent different values of. Each spike neuron in WTA circuit is modeled as a simplified Spike Response Model (SRM) in which the

membrane potential is determined by the difference between excitatory and inhibitory inputs [16]. represents a common inhibitory signal that is fed to all WTA neurons and in turn depends on the activity of these neurons. Experimental results show that neurons in close proximity to each other have a common local inhibitory signal [17], thus competitive mechanism is implemented through. represents the sum of excitatory signals from the input neuron, expressing as

$$u_k(t) = w_{k0} + \sum_{i=1}^{n} w_{vki} \cdot y_{vi}(t) + \sum_{i=1}^{n} w_{aki} \cdot y_{ai}(t) . \quad (1)$$

Where w_{k0} models the intrinsic excitability of the neuron, $w_{vki} \cdot y_{vi}(t)$ and $w_{aki} \cdot y_{ai}(t)$ models the EPSPs of neuron z_k evoked by spikes of the presynaptic neuron y_{vi} and y_{ai}. For simplicity, we assume that y_i is a two-valued function. The value of y_i takes 1 in the time interval δ after it fires, otherwise it takes 0.

2.2. Probabilistic interpretation of spike events

In this part, we will show the probabilistic interpretation of spike events. First, we establish a stochastic firing model for z_k. A hypothesis consistent with experimental data is that the firing behavior of z_k is an independent non-homogeneous Poisson process, and the firing probability is dependent exponentially on its membrane potential [16]. In the WTA circuits, due to the existence of competition mechanisms, if a spike event occurs at time t, the spike is emitted by only one neuron. Since neurons have refractory period [18], it is assumed that only one spike exists in a minimal time interval $[t, t+\varepsilon t]$. Therefore, in $[t, t+\varepsilon t]$, the firing probability for each z_k is approximate as $\lambda_k(t)\varepsilon t$.

Since the firing behavior of z_k is an independent non-homogeneous Poisson process, the sum of the firing activities of K neurons in the WTA circuits is the sum of K independent Poisson processes, and the conditional probability of the spike originated from neuron z_k can be expressed as

$$p_k(t) = \frac{\lambda_k(t)\varepsilon t}{r(t)\varepsilon t} = \frac{e^{u_k(t)-I(t)}}{\sum_{m=1}^{K} e^{u_k(t)-I(t)}} = \frac{e^{u_k(t)}}{\sum_{m=1}^{K} e^{u_m(t)}} . \quad (2)$$

The conditional probability $p_k(t)$ is called the spike instantaneous distribution of z_k. From Eq,(2), we can find that inhibitory signals do not influence the instantaneous distribution. In addition, the instantaneous distribution is a non-homogeneous distribution, and each spike in the WTA circuits can be understood as an independent sampling of the system from the non-homogeneous distribution.

2.3. The method of weight updating

Physiological experiments have shown that for excitatory synaptic connections between neurons, when the presynaptic neuron fires before the postsynaptic neuron, a long-term potentiation (LTP) effect occurs in the synaptic connection between these two neurons, whereas firing in the reverse order results in long-term depression (LTD) of the synaptic connection [16]. STDP is a learning rule frequently used by researchers in addressing multisensory integration problems. In this paper, the updating of synaptic connection weights uses a simple STDP learning rule, and the weight update follows a Hebbian form as described below:

$$w_{ki} = w_{ki} + \eta \cdot \Delta w_{ki} . \quad (3)$$

Where w_{ki} is the synaptic connection weight between input neuron y_i and output neuron z_k, and η is the learning rate. Δw_{ki} is the amount of synaptic weight change, related to the firing state of the

presynaptic neuron y_i and the negative exponential power of the current weight w_{ki} .The condition of weight updating is the postsynaptic neuron firing at the moment t^f .

2.4. Bayesian calculation of instantaneous spike distribution

In this part, we will show that the spike instantaneous distribution $p_k(t)$ in Eq.(2) can be explained as the Bayesian computation results of the generative model based on hidden causes and multiple observed variables. First, we describe the generative model. For simplicity, we use the input neurons $\boldsymbol{y}_v$, $\boldsymbol{y}_a$ to represent the observed variables $\boldsymbol{X}_v$, $\boldsymbol{X}_a$ .For any component of the observed variable x_j ,all possible values of x_j have a corresponding y_i . Note that the possible value of x_j represented by y_i is $v(i)$, and the index set representing all possible values of x_j is G_j ,which consists of all y_i that represent x_j . The spike from the input neuron y_i at time t can then be interpreted as the value of x_j is $v(i)$ at time t . Assume that each component of the observed variable is independent of each other, we have then we can write the generation model of observed variables as

$$p(\boldsymbol{X}_v, \boldsymbol{X}_a \mid \theta) = \sum_{k=1}^{K} \frac{1}{Z} (e^{w_{k0}} \prod_{j=1}^{m} \prod_{i \in G_{vj}} e^{w_{vki} \cdot \{x_{vj} = v(i)\}} \prod_{j=1}^{m} \prod_{i \in G_{aj}} e^{w_{aki} \cdot \{x_{aj} = v(i)\}}) = \frac{1}{Z} \sum_{k=1}^{K} e^{w_{k0} + \sum_{i=1}^{n} e^{w_{vki} \cdot y_{vi}} + \sum_{i=1}^{n} e^{w_{aki} \cdot y_{ai}}} . \quad (4)$$

The last term is relevant to the input vector, the hidden cause and weights, which can be recorded as $p(\boldsymbol{y}_v, \boldsymbol{y}_a \mid \boldsymbol{w})$, where $\boldsymbol{w}$ represents the parameter vector in the network, including the connection weights of visual input neurons and output neurons and the auditory weights. The variable Z is a standardized constant, explicitly calculated as

$$Z = \sum_{k=1}^{K} (e^{w_{k0}} \cdot \prod_{j=1}^{m} Z_{vkj} \prod_{j=1}^{m} Z_{akj}) . \quad (5)$$

Where $Z_{vkj} = \sum_{i \in G_{vj}} e^{w_{vki}}$, $Z_{akj} = \sum_{i \in G_{aj}} e^{w_{aki}}$.

According to the total probability formula, Eq.(4) can be decomposed as

$$p(\boldsymbol{y}_v, \boldsymbol{y}_a \mid \boldsymbol{w}) = \sum_{k=1}^{K} e^{w_{k0}} \cdot \frac{\prod_{j=1}^{m} Z_{vkj} Z_{akj}}{Z} \cdot \prod_{i=1}^{n} (\frac{e^{w_{vki}}}{Z_{vkj}})^{y_{vi}} \prod_{i=1}^{n} (\frac{e^{w_{aki}}}{Z_{akj}})^{y_{ai}} . \quad (6)$$

From Eq.(5) and Eq.(6), we can write the prior as $p(k \mid \boldsymbol{w}) = e^{w_{k0}} \cdot \frac{\prod_{j=1}^{m} Z_{vkj} Z_{akj}}{Z}$ and the likelihood as $p(\boldsymbol{y}_v \mid k, \boldsymbol{w}) = \prod_{i=1}^{n} (\frac{e^{w_{vki}}}{Z_{vkj}})^{y_{vi}}$, $p(\boldsymbol{y}_a \mid k, \boldsymbol{w}) = \prod_{i=1}^{n} (\frac{e^{w_{aki}}}{Z_{akj}})^{y_{ai}}$ under the normalized condition $\sum_{k=1}^{K} e^{w_{k0}} = 1$ and $\forall k, j \quad \sum_{i \in G_j} e^{w_{ki}} = 1$, it is easy to know that $Z_{vkj} = 1, Z_{akj} = 1$,thus $Z = 1$.

Next, we establish the connection between the above probabilistic generation model and the WTA circuit based on spike events. It can be proved that given the input neuron activity $\boldsymbol{y}(t^f)$ at time t^f , the posterior distribution of hidden causes H is equivalent to the firing probability of the output neuron. According to Bayes' formula, from Eq.(4), Eq.(6) and the normalized condition, we have

$$p(k \mid \boldsymbol{y}(t^f), \boldsymbol{w}) = \frac{e^{w_{k0}} e^{\sum_{i=1}^{n} w_{vki} \cdot y_{vi}(t^f)} e^{\sum_{i=1}^{n} w_{aki} \cdot y_{ai}(t^f)}}{\sum_{q=1}^{K} e^{w_{q0} + \sum_{i=1}^{n} e^{w_{vqi} \cdot y_{vi}(t^f)} + \sum_{i=1}^{n} e^{w_{aqi} \cdot y_{ai}(t^f)}}} = \frac{e^{u_k(t^f)}}{\sum_{q=1}^{K} e^{u_q(t^f)}} = p_k(t) . \quad (7)$$

In this way, we have established a connection between the probabilistic generative model and the spike events in the WTA circuit. Consequently, each variable input can be seen as a single sampling from the posterior distribution in the probabilistic generative model, which is equivalent to the instantaneous spike distribution in the WTA circuit at every spike time. This allows for STDP weight learning based on spike timing, modifying the weight parameters in the probabilistic generative model. This process is repeated iteratively until the model completes learning the distribution of hidden variables.

3. Simulation Experiment and Results

In order to explore the performance of our network on the classification task, we designed a numerical simulation, choosing 36 different locations on a 28*28 2-D plane as the centers of different categories. Each integer point in the plane is binary, and a value of 1 indicates that there is a light spot or sound source in that position. The idea of this numerical simulation is inspired by the robot experiments in reference [12]. The samples of the training set and the test set are generated by Gaussian distribution in each center, and each sample point represents the position of the light spot or sound source in the plane. Only variance is different in the generation of visual samples and auditory samples, for simplicity we show a partial sample of the auditory training set.

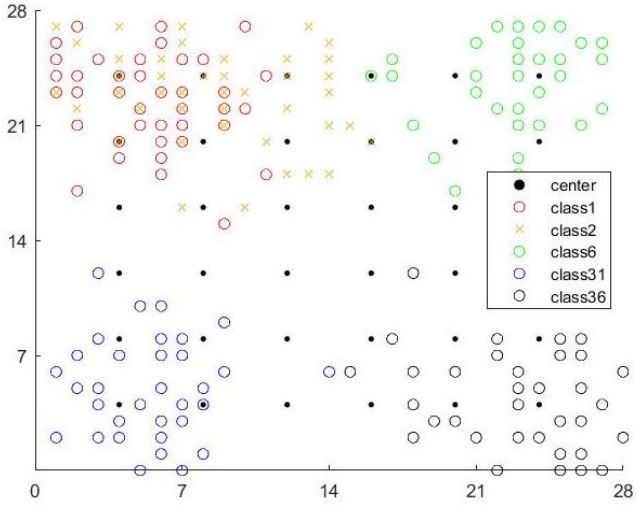

Fig. 2 Schematic diagram of some auditory samples in two-dimensional space

Fig.2 shows the centers of each class and some sample points of different classes of hearing in the plane. As can be seen from Fig.2, categories with similar centers have higher sample coincidence rate, thus there exists the possibility of misclassification. Since our goal is to give the correct category of the sample in the test set through the network, the error rate is selected as the evaluation index of the network performance.

When training, the multisensory network takes a set of visual and auditory samples as input and produces spikes from neurons in the WTA circuit, while the unisensory version takes only one modality. The synaptic weights update according to the timing of spikes. Recording the category of this input set, the values of the hidden causes corresponding to each output neuron are assigned to the class of input that caused that neuron to fire the most during the training process. Although class labels are used when determining the category of output neurons, no class labels are used at all during the entire training process, making the network unsupervised [15]. Once training is completed, the generalization of the model is validated using a test set, with error rate as the criterion for assessing the model's classification accuracy.

In the next part, we compare the classification performance of the model under different sensory input conditions.

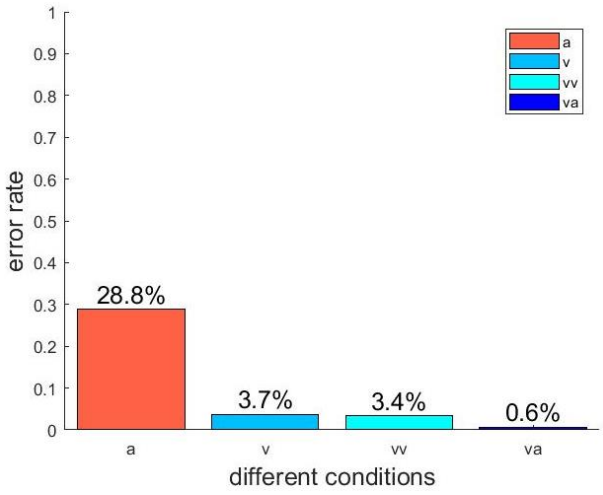

Fig. 3 Schematic representation of classification performance under four sensory input conditions

Fig.3 shows the classification performance of the model under four different sensory input conditions. The four input conditions are single visual (v), single auditory (a), visual and auditory combined (va), and dual visual (vv). Dual visual input refers to a set of input samples consisting of two visual samples. The inclusion of dual visual input is intended to eliminate the potential impact of varying numbers of input neurons between single-sensory and multisensory cases. From Fig.4, it can be observed that after sufficient training, the classification error rate for multisensory audiovisual input is much lower than the error rate for any single sensory input, indicating that the integration of multisensory information effectively improves the network's classification performance. In the case of dual visual input, the classification error rate is only 0.3% higher than that of single visual input, suggesting that the reduction in error rate is not dependent on the number of neurons but rather arises from the integration of information from different sensory inputs.

4. Summary

In the research on multisensory integration, both probability models proposed to explain neurophysiological phenomena and neural network models developed to address multisensory tasks are equally important. The proposed probability models can explain various phenomena observed in neurophysiological experiments, such as the inverse effectiveness principle and spatial principles [8,9]. However, these probability models are rarely applied to real data. On the other hand, researchers have achieved advanced performance in tasks like multimodal emotion recognition [14], while they do not provide a probabilistic interpretation of these models.

This paper primarily focuses on studying the probabilistic model of multisensory integration in neural circuits. It introduces a neural network model based on the WTA mechanism for multisensory integration. The paper also provides probabilistic explanations for spike events within the network through Bayesian analysis and attempts to validate the effectiveness of the model in multisensory classification problems through numerical simulations. This research may provide directions for the future development of more comprehensive and practically valuable probabilistic models. For instance, while the model proposed in this paper was derived and numerically simulated for only two sensory modalities, its framework can be extended to multiple sensory modalities. Furthermore, research suggests that the effectiveness of multisensory integration may be related to the synchrony of neural spike firing [1,4,14]. This paper only considered fully synchronous visual and auditory stimuli, and investigation on asynchronous multisensory stimuli is needed in the future.

Acknowledgements

The authors gratefully acknowledge the financial support in part by National Natural Science Foundation of China (61374183).

References

[1] Taherkhani A, Belatreche A, Li Y, et al. A review of learning in biologically plausible spiking neural networks[J]. Neural Networks, 2020, 122: 253–272.

[2] Kayser C, Shams L. Multisensory Causal Inference in the Brain[J]. PLOS Biology, 2015, 13(2): e1002075.

[3] Gu Y, Angelaki D E, DeAngelis G C. Neural correlates of multisensory cue integration in macaque MSTd[J]. Nature Neuroscience, 2008, 11(10): 1201–1210.

[4] Cuppini C, Shams L, Magosso E, et al. A biologically inspired neurocomputational model for audiovisual integration and causal inference[J]. European Journal of Neuroscience, 2017, 46(9): 2481–2498.

[5] Ursino M, Cuppini C, Magosso E. Multisensory Bayesian Inference Depends on Synapse Maturation during Training: Theoretical Analysis and Neural Modeling Implementation[J]. Neural Computation, 2017, 29(3): 735–782.

[6] Fang Y, Yu Z, Liu J K, et al. A unified neural circuit of causal inference and multisensory integration[J]. Neurocomputing, 2019, 358: 355–368.

[7] Ma W J, Beck J M, Latham P E, et al. Bayesian inference with probabilistic population codes[J]. Nature Neuroscience, 2006, 9(11): 1432–1438.

[8] Ursino M, Cuppini C, Magosso E, et al. Multisensory integration in the superior colliculus: a neural network model[J]. Journal of Computational Neuroscience, 2009, 26(1): 55–73.

[9] Ohshiro T, Angelaki D E, DeAngelis G C. A normalization model of multisensory integration[J]. Nature Neuroscience, 2011, 14(6): 775–782.

[10] Pouget A, Beck J M, Ma W J, et al. Probabilistic brains: knowns and unknowns[J]. Nature Neuroscience, 2013, 16(9): 1170–1178.

[11] Fetsch C R, DeAngelis G C, Angelaki D E. Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons[J]. Nature Reviews Neuroscience, 2013, 14(6): 429–442.

[12] Bauer J, Magg S, Wermter S. Attention modeled as information in learning multisensory integration[J]. Neural Networks, 2015, 65: 44–52.

[13] Körding K P, Beierholm U, Ma W J, et al. Causal Inference in Multisensory Perception[J]. O. Sporns. PLoS ONE, 2007, 2(9): e943.

[14] Benssassi E M, Ye J. Investigating Multisensory Integration in Emotion Recognition Through Bio-Inspired Computational Models[J]. IEEE Transactions on Affective Computing, 2023, 14(2): 906–918.

[15] Rathi N, Roy K. STDP Based Unsupervised Multimodal Learning with Cross-Modal Processing in Spiking Neural Networks[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2021, 5(1): 143–153.

[16] Nessler B, Pfeiffer M, Buesing L, et al. Bayesian computation emerges in generic cortical microcircuits through spike-timing-dependent plasticity[J]. PLoS computational biology, 2013, 9(4): e1003037.

[17] Fino E, Yuste R. Dense Inhibitory Connectivity in Neocortex[J]. Neuron, 2011, 69(6): 1188–1203.

[18] Dayan P, Abbott L F. Theoretical neuroscience: computational and mathematical modeling of neural systems[M]. Cambridge, Mass: Massachusetts Institute of Technology Press, 2001.