

南京师范大学

# 毕业论文(设计)

(2020 届)



题    目： 快同效应的认知机制：  
基于人工神经网络的模拟研究

学    院： 心理学院

专    业： 应用心理学

姓    名： 孙禾嘉

学    号： 19200210

指导教师： 胡传鹏

南京师范大学教务处    制

## 摘 要

理解人类的决策是深入认识人脑的重要窗口。在快速的异同判断中,研究者发现人们对相同刺激对做出“相同”反应的速度快于对不同刺激对做出“不同”反应的速度,这一稳定的效应被称为快同效应。然而,发现这一效应六十年后,其背后的认知机制仍然不清楚。当前主要存在两种对立的观点。其中一种认为,人们加工不同刺激对时与加工相同刺激对时只有单一的加工过程,即单过程理论;另一观点则认为,人们在加工两种刺激对时存在两个不同的加工过程,即双过程理论。但行为的数据难以为这一争议提供满意的证据。本研究通过构建人工神经网络模型得到与人类类似的快同效应并分析神经网络模型的内在特点来帮助理解快同效应的产生机制。研究一中,我们构建并训练 FSNet 模型,使之复现人类被试在异同判断任务上的表现出的快同效应。研究二中,我们采用 PCA 方法对神经网络特点进行分析,发现 FSNet 在对相同刺激对与不同刺激对在编码与反应阶段均有不同,这一结果支持双过程理论,指出快同效应产生的原因在于对不同刺激对与相同刺激对的加工上的差异。运用新兴的人工神经网络对经典的快同效应进行模拟并分析神经网络内在的特点,本研究为理解快同效应的认知机制提供了新视角。本研究也表明了人工神经网络作为研究人类认知机制的潜在价值,为后续研究者提供思路和借鉴。

**关键词:** 知觉决策, 异同判断, 快同效应, 神经网络模型

## **Abstract**

Understanding human decision-making is important in comprehending the brain. In rapid same-different judgments, researchers have found that people respond more quickly to pairs of identical stimuli with a "same" response than they do to pairs of different stimuli with a "different" response. This consistent effect is known as the fast same effect. However, even after sixty years of discovering this effect, its underlying cognitive mechanisms remain unclear. Currently, there are two opposing viewpoints. One holds that there is only a single processing stage when people process different stimulus pairs and the same stimulus pairs, known as the single-process theory. The other viewpoint suggests that there are two distinct processing stages for the two types of stimulus pairs, known as the dual-process theory. Behavioral data, however, has not been able to provide satisfactory evidence for this controversy. This study aims to build an artificial neural network model that replicates the fast same effect found in humans and analyzes the intrinsic characteristics of the neural network model to help understand the mechanism behind the effect. In Study One, we constructed and trained the FSNet model to replicate the fast same effect exhibited by human participants in same-different judgment tasks. In Study Two, we used PCA (Principal Component Analysis) to analyze the characteristics of the neural network, discovering that FSNet showed differences in both the encoding and response stages for identical and different stimulus pairs. This result supports the dual-process theory, suggesting that the fast same effect arises from the differential processing of same and different stimulus pairs. By using emerging artificial neural networks to simulate the fast same effect and analyzing the intrinsic characteristics of the networks, this study offers a new perspective on understanding the cognitive mechanisms behind the effect. It also demonstrates the potential value of artificial neural networks in researching human cognitive mechanisms, providing ideas and references for subsequent researchers.

**Key words:** Perceptual decisions, Same-different judgement task, Fast-same effect, Artificial Neural Network

# 目录

摘 要 .....	2
Abstract .....	3
第 1 章 引言 .....	5
1.1 知觉决策与快同效应 .....	5
1.1.1 异同判断任务中的快同效应 .....	5
1.1.2 快同效应的理论研究 .....	6
1.2 通过 ANN 模型来研究认知过程的内部机制 .....	7
1.3 研究问题与研究假设 .....	10
第 2 章 研究 1: FSNet——人工神经网络的快同效应 .....	11
2.1 研究目的 .....	11
2.2 方法 .....	11
2.2.1 训练数据 .....	11
2.2.2 模型架构 .....	12
2.2.3 训练框架与训练流程 .....	15
2.3 结果 .....	15
2.3.1 反应时分布 .....	15
2.3.2 快同效应 .....	17
2.4 讨论 .....	17
第 3 章 研究 2: 人工神经网络中快同效应的网络特性 .....	18
3.1 研究目的 .....	18
3.2 研究方法 .....	18
3.2.1 神经网络隐藏层的 PCA 方法 .....	18
3.2.2 模型训练和测试方法 .....	19
3.2 结果 .....	19
3.2.1 PCA 聚类结果 .....	19
3.2.2 对因子的解释 .....	21
3.2.2 快同效应的认知机制 .....	22
第 4 章 讨论 .....	24
第 5 章 总结与展望 .....	27
致谢 .....	35
附录: 预实验 .....	36
1. 研究目的 .....	36
2. 模型结构简述 .....	36
2.1 基于 ANN 的模型 .....	36
2.2 ANN 和其他模型混合结构 .....	37
3. 模型训练数据 .....	39
4 结果 .....	39

## 第 1 章 引言

### 1.1 知觉决策与快同效应

#### 1.1.1 异同判断任务中的快同效应

决策(decision-making)指的是个体在不同选项中做出选择的过程(Shadlen & Kiani, 2013)。人们每天都要进行无数的决策,因此决策一直是社会科学家广泛关注的主题。心理学家和经济学家针对决策提出了诸多理论,如经济学中的期望效用理论(expected-utility maximization theory)、行为经济学中的前景理论(prospect theory)(Edwards, 1954)。

知觉决策(Perceptual decision-making)是一类特殊的决策过程。其指将感觉信息转化为判断,信念和行动的决策过程。知觉决策是一个重要的窗口,帮助理解决策背后的认知与神经机制(O'Connell & Kelly, 2021)。知觉决策的过程往往是快速的,无意识的,心理物理学数据显示,大多数感知任务的时限为 200-300ms (Uchida, Kepecs, & Mainen, 2006)。认知神经科学的研究者也采用神经科学的方法(如 fMRI、EEG、颅内电记录)对知觉决策的过程提出了不同的理论模型,通过这些模型,我们能够更好的理解决策的本质以及研究认知系统做出决策的内部过程(刘逸康, 胡传鹏, 2024)。

异同判断(same-different judgment)是一类经典的知觉决策任务,在该实验任务中,被试需要判断呈现的刺激是相同还是不同,并根据判断结果做出不同的反应。异同判断任务是人类认知能力中的最基本的能力。James(1984)在他的书中写道:“对于异同的觉知是我们的思维之支柱”,“是我们心灵结构的最重要特性”。研究异同判断对于研究人知觉过程具有重要的作用,因为这一过程展示了人类信息处理的基本机制(Posner & Mitchell, 1967)。

在异同判断的研究中,研究者发现稳定的快同效应(fast-same effect)或快同现象(fast-same phenomenon),即被试识别两个相同的刺激(如“A-A”)要比识别两个不同的刺激更快(如“A-B”) (Goulet & Cousineau, 2020)。Nickerson(1965)的实验显示,95%的被试在“不同”的判断任务上耗时比“相同”判断更长,反应时的平均相差为 80ms。Krueger(1978)在划线方格的比对实验中,发现不同刺激间隔下相同

判断的反应时都低于不同判断。快同效应在不同的情境下得到了重复验证,并且有研究者还发现了在不同情境下快同效应的变化(Chignell & Krueger, 1984)。快同效应是异同判断任务的一个显著特点,这预示了异同判断任务的某种特殊加工规律,对快同效应进行研究解释,能够帮助揭开异同判断任务的加工机制。

### 1.1.2 快同效应的理论研究

为了解释快同效应产生的原因,有研究者针对异同判断任务提出了理论模型,现有主要有两种理论观点,分别是 Krueger(1983)和 Proctor(1983)等人提出的单过程模型和 Bamber(1969)和 Reed 等(2013)提出的双过程模型。这两种模型都假设了一个慢速比较器和一个同一性指示器,慢速比较器用于对刺激各个特征逐一对比分析,同一性指示器用于快速地给出刺激是否相同的判断。两种模型的区别在于具体调用这两个比较器的加工过程在“相同”的判断上和“不同”的判断上是否是一致的。其中,单过程模型认为,个体在做出“相同”判断和“不同”判断时使用的是同一个认知过程,在加工相同刺激对和不同刺激对时,慢速比较器和同一性指示器的调用是一致的。双过程模型则认为“相同”判断和“不同”判断由分别由不同的认知过程来进行,在相同刺激对和不同刺激对中,慢速比较器和同一性指示器是以不同方式调用的。这个差异也是目前争论的焦点:人脑做出相同的判断和做出不同的判断时是否是基于同一个认知过程(Farell, 1985),即加工过程是单过程的或双过程的,进一步的,研究者还需要回答慢速比较器和同一性比较器这两个比较结构是如何运作的,是同时启动的或是顺序启动的。

现有的关于异同判断任务中的“快同效应”的研究多基于行为实验的数据,少数有研究采用脑电相关的技术,要对异同判断任务的内部机制做进一步研究,需要新的研究方法来提供更深入的证据。对于类似于异同判断任务这样特定的认知过程,传统的理论模型研究往往是通过人工建构(hand-constructed)的,理论模型的具体结构基于对特定认知过程的具体理论。这些基础的认知模型在解释认知任务上取得了进展,例如漂移扩散模型对快同效应在反应偏向上的解释(Goulet, 2020; Harding, 2022)。但是由于实际的认知情况可能会面临复杂的情境,具体的认知过程也可能会受到多种因素的影响而发生改变。因此人工构建的理论模型可能难以解释复杂的认知的全部过程。虽然人工建构的理论具有简洁明了的特性,但是可能无法解释复杂的神经活动和认知过程(Rigotti et al., 2010, 2013)。因此,

对于异同判断这样一个复杂多因素的任务,现有的理论研究工具可能难以提供深入的证据以及构建更全面的模型。我们需要一种新的研究工具帮助我们建构更全面的异同判断模型。

而近年来,人工神经网络(Artificial Neural Networks, ANNs)的发展为心理学带来一种新的研究工具,人工神经网络能够直接建立复杂的,多因素的人类认知过程的模型,并且能够通过多种方法对其内部结构进行分析,甚至进行破坏性的操作(Yang et al., 2019)。许多研究者已经使用人工神经网络来构建心理学模型,并且取得了成效,例如使用 CNN 构建人类视觉的模型来研究“what”通路,通过 CNN 网络的输出和内部状态来预测下颞叶皮层(inferior temporal cortex)的神经活动和 V4 低级视觉区的神经活动(Yamins et al., 2014)。Testolin, De Filippo, 和 Zorzi(2017)通过训练 ANN 进行不同空间参考帧之间的显式映射发现网络结构和学习范式都会影响神经编码过程。人工神经网络在建立人类知觉行为的模型研究上有着独特的优势,能帮助我们解决传统建模难以深入内部机制,以及人工建构无法拟合多种情况的问题。

许多认知过程通过 ANN 这一新工具得到进一步研究,然而目前尚未有研究者将这一方法应用到异同判断任务中来研究快同效应的机制,因此,本研究希望通过 ANN 工具来建立异同判断任务的理论模型,并通过分析这一模型的活动特点和内部结构来研究异同判断任务中的快同效应的认知机制。

### 1.2 通过 ANN 模型来研究认知过程的内部机制

使用 ANN 模型来研究认知过程的内部机制,一个重要的问题是这种方法和传统的行为研究或者脑电研究方法究竟有什么本质上的区别,它们在使用目的和结果上有何区别?要回答这个问题,必须对这些方法应用的对象——“心理学研究”这一过程做解释。心理学研究是我们尝试对人脑进行理解的过程,它本质上是一个对复杂信号处理系统的理解的过程。计算神经科学的奠基人 Marr (1983)在他的著作中提出一个框架用于阐释这样一个理解的过程:对于一个信号处理系统,例如人脑这样一个信息处理系统,存在三个理解的层次,分别是计算理论层,算法表达层和硬件实现层。

计算理论层是一个抽象的理解层面,这个层面关注于这个系统的功能以及选

择这种功能的原因，以超市收银机为例子，超市收银机需要运用加法原理来累加账单，加法原理是一种基本的理论，使用阿拉伯数字，算筹或者二进制来表征，它的运算结果都是不变的。加法原理的可交换律，结合律等等性质在任何一种表征中都是一致的。对于把购买物品的价格合并到账单中这个过程，其遵循的原理与加法是一致的。因此，对于收银机这个信号处理系统来说，加法原则就是其计算理论层面的理解，由于加法与账单累加这个过程从本质来说有着同样的约束规则，加法是适合于这一信号处理系统的原理。

第二个层次是算法表达层，这个层级的关键在于这个信号处理系统是如何工作的，它处理信号的原则是什么？这个层级关注这个系统的输入输出形式是什么，哪些算法被用于原酸过程。收银机需要遵守“ $\Sigma \text{单价} = \text{总价}$ ”这个算法，输入各个物品的单价，输出总价。在人类的视觉系统中，会使用例如梯度信息或透视信息等信息差异来计算深度，这两者都是算法表达层面的理解。

第三个层次是硬件实现层，这个部分主要表示这个信号处理系统是如何在物理上搭建的，例如人的视觉处理系统是基于神经元的链接而形成的生物神经网络，收银机的信号处理则是基于晶体管搭建的集成电路。

这三个层面相互具有联系，但联系并不紧密，同一种算法可能由不同的硬件来实现，不同的表示和算法也可能基于同一种计算理论，一个信号处理系统中每个层次都有广泛的选择。

在心理学研究中，不同的研究对应着不同的解释层面。神经生理学的研究方法，主要关注的是硬件实现层的信息，通过研究神经元以及脑皮层的活动从物理角度解读人脑，推演算法甚至计算理论层面的信息。这一过程是间接的，从硬件层级到算法甚至计算理论层级的猜想和假说可能需要较多证据的支持，并且难以验证。而行为实验的研究方法则关注于算法表达层的输入与输出，通过这些外显的信息以及其他的辅助证据来提出有关算法与表示的理论假设，并根据这些理论来研究计算理论，这一方法难以直接触及到算法表达层的具体实现方法，只能通过控制输入和输出来间接推断。

通过建立 ANN 模型的方法，能够超越生物学的硬件，基于目标信息处理任务，构建一个信息处理系统。如果这个系统能够进行某个信息处理任务，并且与人脑处理类似任务时有着同样的功能和特点，他们在计算理论上就有着相似的结



构。由于 ANN 是基于计算机代码的，所有的运算和数据都被存储在磁盘中，能够直接被观察到，所以比起对人脑的算法层面进行研究，对 ANN 的结构以及运行原理分析就会更加容易。

通过研究和分析 ANN 模型，我们能够得到 ANN 完成某项任务的运作原理，即具体的算法表征层面的信息，这些基本原理反映完成某项任务所需要的一般规律，即计算理论层面的信息。人脑的认知系统在完成同样的认识任务的过程中可能也存在着相似的计算理论。因此，通过建立 ANN 模型，对模型的算法以及理论做出解释，能够帮助我们更好的理解人脑执行任务时的认知机制，这一过程可以用一个形象的比喻来说明这一过程(图 1)。

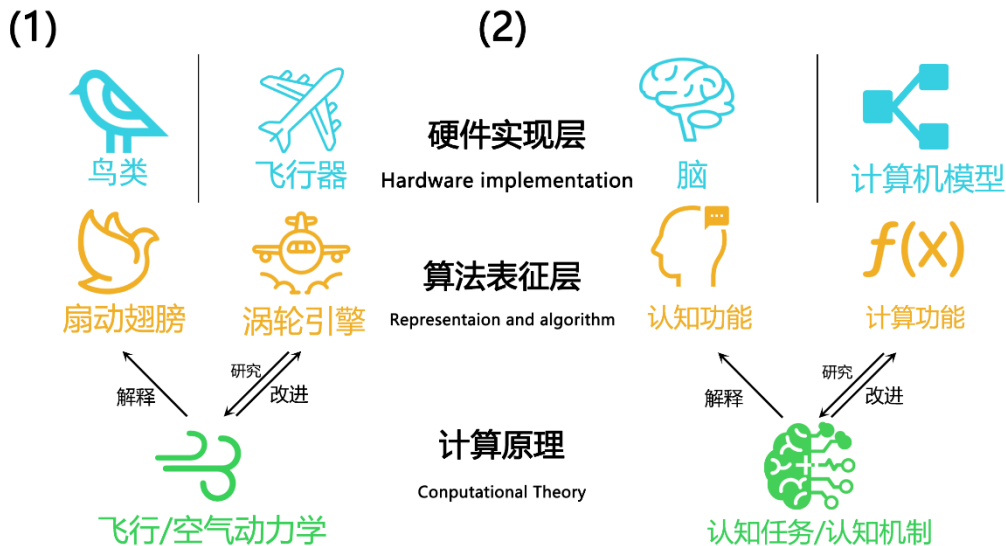


图 1 通过模型来研究脑认知过程的内部机制. (1)鸟类与飞行器有同样的功能，即同样的计算理论，发明家尝试建造飞行器来达成飞行的功能，飞行器要实现功能需要借助某种具体方法，例如涡轮引擎(算法表征)，虽然这与鸟类挥舞翅膀的算法并不相同，但是在涡轮引擎的研制中，发明家发现了空气动力学原理(即计算原理)，通过空气动力学，科学家能够解释鸟类是如何飞行的。(2)ANN 并不是人认知系统的完整复刻，但其功能上的相似性可能预示着在计算理论上的共同点。通过对 ANN 的算法表征以及计算理论进行探究，能够帮助我们理解人脑的认知系统的计算理论。

综上，本文希望通过 ANN 来建立一个与人类有着类似计算原理的计算机模型，使用神经网络分析方法对模型的算法表征进行研究，进一步的，从模型的算法表征中总结出模型的计算原理。结合神经网络的算法表征和计算原理与现有的研究和理论，对异同判断任务的内部机制进行研究，并尝试对异同判断任务中快

同效应做出解释。

### 1.3 研究问题与研究假设

如前所述，异同判断是一个重要的心理过程，对其的研究能够帮助我们理解人脑运作的基本机制。而快同效应是异同判断任务中产生的特殊现象，研究快同效应能够帮助我们解答异同判断的内部机制。

针对异同判断任务，研究者提出单过程模型和双过程模型，这两种模型都假设相同的两种结构，但对于这些结构是如何被组合调用的还尚未有定论。传统的心理学方法通过行为实验和脑电技术对异同判断任务做出一些解释，但仍需要更深入的新方法来提供新的证据。人工神经网络这一新的研究工具在许多心理学问题上取得了成效，而尚未有研究者将其应用于快同效应的研究中。

因此，本研究的研究问题是：通过人工神经网络来研究异同判断任务的认知过程，并对快同效应的认知机制做出解释。

本研究的具体步骤是，基于真人被试进行异同判断任务实验的数据，通过人工神经网络来建立与人类被试有相似反应特点的异同判断任务模型，并复现出快同效应。进一步的，通过神经网络分析方法来对网络的内部结构进行研究，对其进行算法层面和计算理论层面的解释，进一步来揭示快同效应的认知机制。

基于前述的有关异同判断的理论，本文对预期结果做出如下假设：

1. 在异同判断任务中，慢速比较器和同一性指示器是顺序加工的。
2. 在异同判断任务中，对相同刺激的加工和不同刺激的加工使用的是不同的加工过程。

## 第2章 研究1: FSNet——人工神经网络的快同效应

### 2.1 研究目的

本研究的主要目的是通过人工神经网络的工具来建立一个异同判断任务的认知模型,旨在模拟人类被试具有的反应特点,并复现出快同效应。本研究构建的模型称为 FSNet(Fast-same Net)模型。

研究1的主要内容包括以下几个部分:1.研究方法,这一部分主要介绍与评估训练神经网络所用的数据,模型架构与训练框架。2.结果,这一部分将对神经网络的输出与人类被试的反应结果,评估神经网络模型是否能够具有与人类被试相似反应特点,并复现出快同效应。

### 2.2 方法

#### 2.2.1 训练数据

实验数据来自研究组内研究者先前的研究数据,该研究进行(1)图-词异同判断任务(2)即时习得图-词异同判断任务两个实验。本研究将主要基于实验(1)的数据,作为原始实验数据来进行模型训练,该实验采用图词异同判断任务,被试将根据指导语学习图形与其对应的文字(圆形图片对应“圆形”文字)并对呈现的图片和文字异同情况进行判断。该实验使用 jsPsych 7.3.1(Leeuw et al., 2023)及 psychophysics 插件编写实验程序,通过网页呈现实验刺激、收集数据。本研究将主要采用该实验中“同时呈现刺激”组的实验数据,具体的实验流程如下:

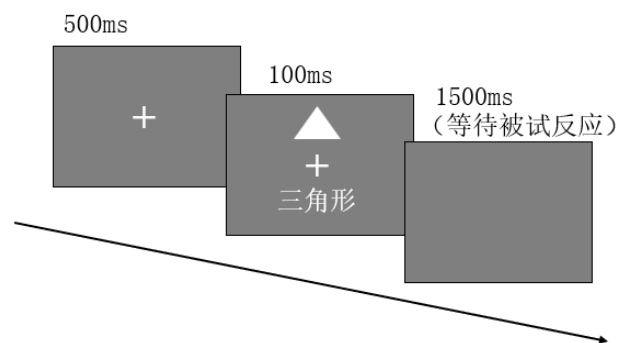


图2 实验流程. 实验有三个阶段,分别是注视点阶段(500ms),刺激呈现阶段(100ms)和反应阶段(1500ms),被试需要在反应阶段作出反应

在每个试次中,首先呈现 500ms 的中央注视点,然后同时呈现一对图形——文字刺激 100ms,随后是 1500ms 的空屏。被试需要在空屏阶段进行按键,如果图形刺激对应文字刺激,就按 m 键;如果不符合就按 n,原始实验的总时长为 2100ms。实验刺激包含叠加噪声的三个简单图形刺激(三角形、圆形、正方形)以及三个文字标签刺激(“三角形”、“圆形”、“方形”)。具体的实验流程图见图 2。

该实验符合条件的数据共包含 40 名被试,共 9544 条反应数据,反应时均值为  $708.69 \pm 154.59\text{ms}$ ,正确率为 0.92,相同刺激对条件和不同刺激对条件下反应时数据的分布形态见图 2,可以从图中看到,相同刺激的反应试次有更快的反应时。数据分析结果显示,相同判断的反应时显著低于不同判断的反应时,  $t(9543) = 21.28, p < .001$ 。这表明,本实验中出现了明显的快同效应。同时数据量也保证了训练的需要,适合基于此数据来建立神经网络模型。

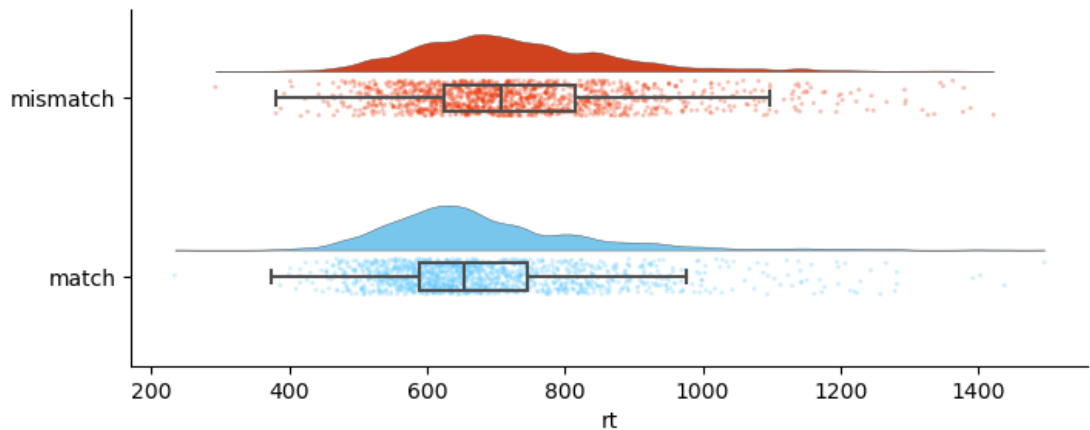


图 3 反应时数据的分布形态

本研究的训练将使用原始数据中随机划分的 70% 的数据,共 6664 条数据作为训练集,训练集将参与网络的训练部分,使用原始数据中随机划分的约 30% 的数据,共 2880 条数据作为测试集,作为检验训练结果的指标。

### 2.2.2 模型架构

本研究通过预实验对 ANN 模型进行初步筛选,以选取最适合本研究目的的模型架构(见附录:预实验)。通过预实验,选择基于 RNN(Recurrent Neural Network, RNN)模型作为基础来构建 FSNet。

FSNet 主要包括三部分(见图 4): 编码网络、决策网络和解码网络。编码网络将实验流程和被试的反应数据进行编码, 转化为时间序列向量, 形成决策网络的输入; 决策网络对输入信息进行决策, 确定“异”或者“同”的决策结果; 解码网络将输出反应时和反应选择。

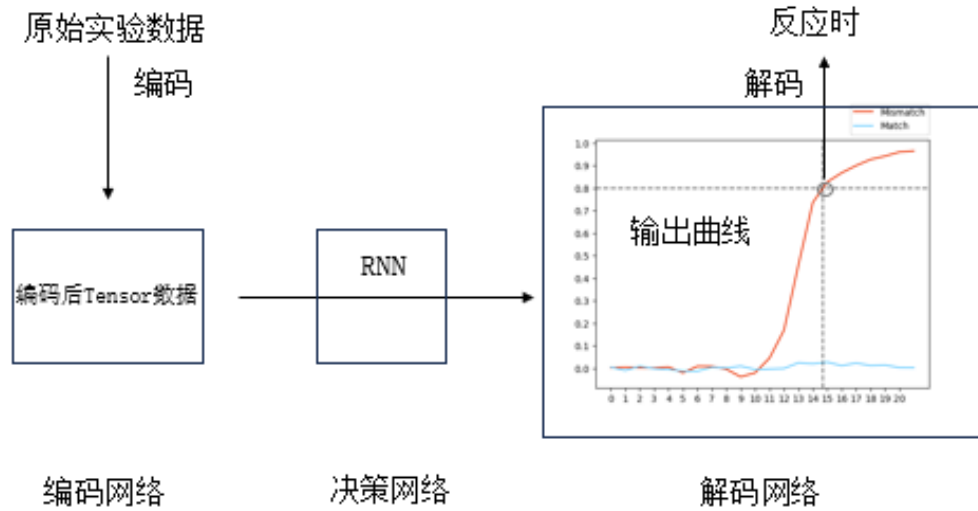


图 4 模型结构示意图

作为模型的第一部分, 编码网络部分是连接人类行为数据与网络行为的关键。由于 FSNet 的决策网络中的 RNN 神经网络的输入和输出必须基于时间序列形式的数据(Sherstinsky, 2020), 因此, 需要对原始实验流程与被试反应数据进行编码, 根据特定的时间点映射在时间轴上, 转换为时间序列形式的数据。这个映射转换过程具体步骤如下:

以 100ms 作为一个最小时间单位, 将原始实验的完整流程编码为长度为 21 的时间序列, 即数组, 其中 0-1 位为注视点阶段, 共两个时间单位, 合 200ms; 2-6 位为刺激阶段, 共五个时间单位, 合 500ms, 7-21 位为反应阶段, 共 15 个时间单位, 合 1500ms。将原始实验中的六种刺激(文字的圆, 方形和三角形; 图形的圆, 方形和三角形)编码为六个不同水平的激活值, 给时间序列中刺激阶段赋予相应的激活值, 即对数组的 2-6 位赋予两个刺激对应的激活值。根据被试的反应时间和反应结果, 给时间序列中反应阶段赋予相应的激活值, 即在数组中的 7-21 位赋予相应的值。为模拟真实人类认知活动的随机性, 输入的刺激激活值被叠加一层噪声。

根据上述转化规则, 对一个实验试次和被试反应的数据进行转化后的时间序列如图 5, 在这一试次中, 呈现一对“不同”的刺激, 被试接受刺激之后, 被试经过 400ms 做出“不同”的反应。神经网络会进行一个相似的过程: 神经网络输入两个“不同”的激活值, 网络输入刺激之后, 经过 4 个时间单位(合 400ms)后, “不同”反应通道的值从 0 变为 1, 即红线在 4 个单位处引来一个拐点, 值上升到了 1, 而绿线值保持为 0, 红线代表对“不同”刺激的判断, 而绿线代表对“相同”刺激的判断。

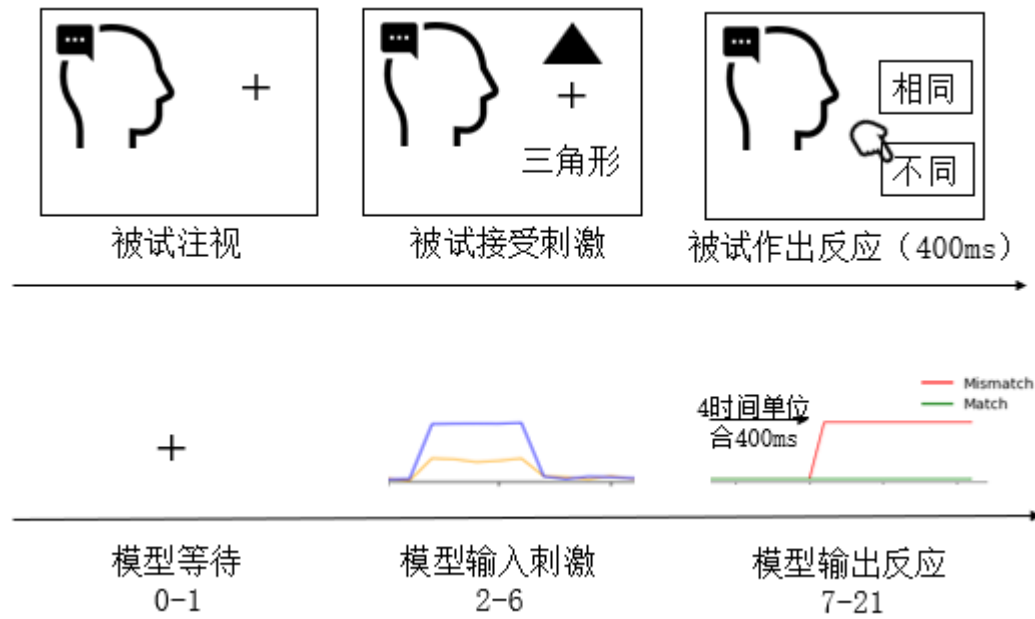


图 5 从实验流程到编码后数据的对应情况。上方为人类被试完成一个实验试次时所经历的流程; 下方为神经网络完成一个试次时的对应流程。

决策网络是 FSNet 的核心部分。本网络基于经典 RNN 结构的神经网络, 包括一个循环层和一个全连接层。其输入为  $23 \times 8 \times 2$  的向量, 该向量经由全连接层处理后输出  $23 \times 2$  的时间序列反应数据。时间序列反应数据将进一步在解码网络中被解码为反应时和选择。训练中, 决策网络将根据人类被试数据对网络参数进行优化, 使其能够模拟人类对异同判断的反应。决策网络输出时间序列反应数据本质上是一组反应曲线。一条代表对“相同”刺激的判断(图 6 中的蓝线), 另一条代表对“不同”刺激的判断(图 6 中的黄线)。

解码网络将对上述的时间序列反应数据进行解码。本部分设定一个反应阈值, 当曲线的值在某一点高于这个反应阈值时, 模型就做出反应, 并根据这一点的横

坐标值得到反应时。如图6所示,蓝色曲线大约在13个时间单位(约为1300ms)处达到0.8的阈值,可认为模型在13个时间单位处做出“相同”反应(蓝色曲线代表对相同的判断),根据第一部分中的数据转化规则解码反应时,可以得到模型最终的输出的反应时是600ms,判断为“相同”。

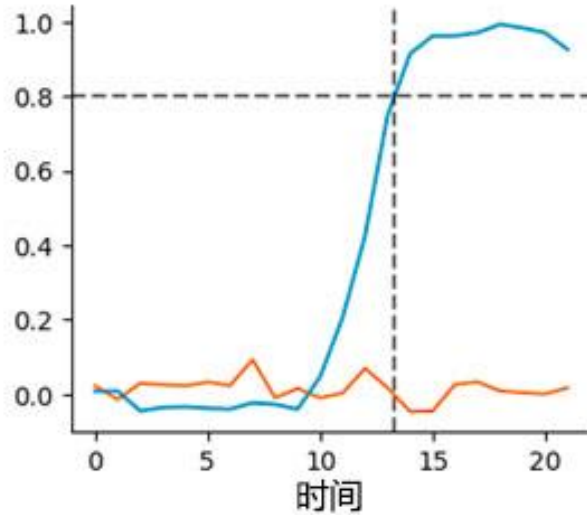


图6 神经网络的输出曲线对

### 2.2.3 训练框架与训练流程

本研究选用 Python 程序语言进行建模环境的搭建,使用 Pytorch 机器学习库来构建具体的模型。为保证训练结果的可重复性和可靠性,使用训练数据集对模型进行 40 次平行训练,对训练的最终模型参数进行保存。

对这 40 个独立模型采用人类被试数据划分出的测试数据集进行测试。每个模型进行 72 次测试,共输出 2880 条反应数据。计算反应时分布偏度,快同效应显著性以及正确率并记录,将结果与人类被试进行比对。

每次训练的具体操作如下:采用 Adam 优化器和 MSELoss 均方损失函数对网络的参数进行优化,设定 loss 阈值为 0.034,达到阈值后停止该次模型的训练。

## 2.3 结果

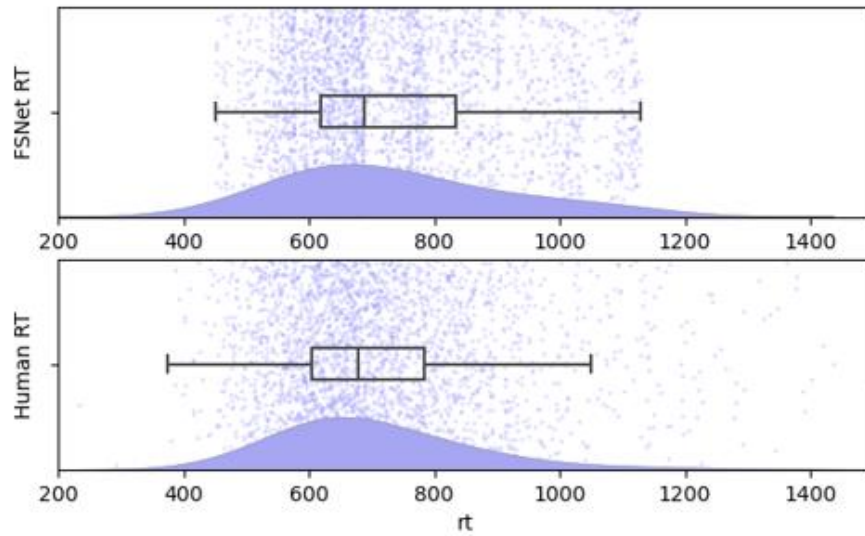
### 2.3.1 反应时分布

将网络的输出绘制提琴图,与人类被试的反应数据对比,见图7上图。在简单的二选择实验中,人类的反应时分布普遍呈现正偏态,并且偏度会随着具体的



试验任务而改变(Forstmann et al., 2016; Ratcliff & McKoon, 2008)。FSNet 的输出结果呈正偏态分布(偏度=0.79)，对每个单独训练的模型偏度分布做显著性检验，所有模型输出的反应时分布的偏度均显著大于 0， $t(2879)=7.20$ ， $p<.001$ 。本研究人类被试的数据中，反应时分布呈现正偏态(偏度=1.21)，神经网络的反应时输出在分布上与人类被试输出相似。

(1)



(2)

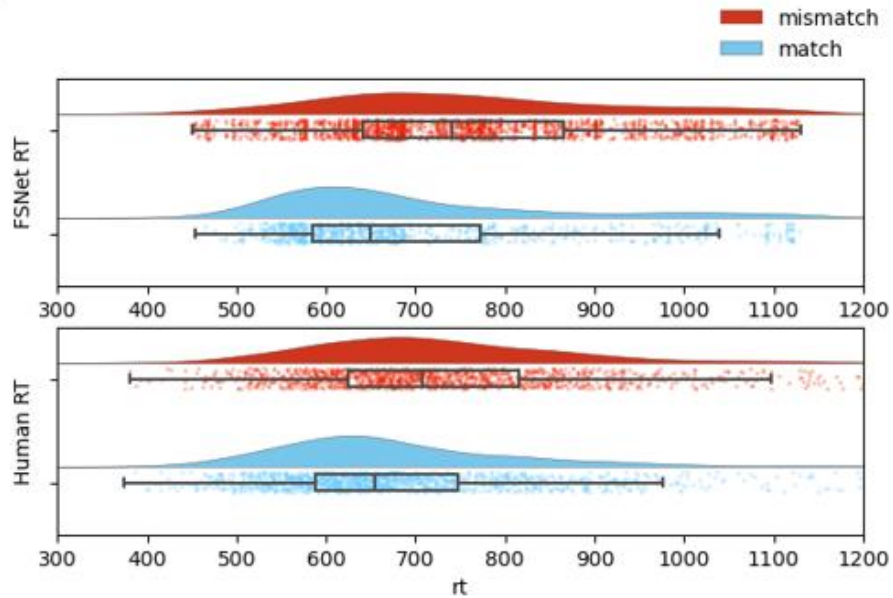


图 7 模型输出的反应时分布与人类被试数据分布



### 2.3.2 快同效应

按照相同和不同刺激试次的条件分别绘制反应时分布,如图 7 下图,描述性统计结果显示,40 个模型的相同判断平均反应时为  $613.34 \pm 40.33\text{ms}$ ,不同判断平均反应时为  $663.64 \pm 40.98\text{ms}$ ,对相同判断和不同判断的反应时进行 t 检验,结果显示相同判断反应时显著低于不同判断反应时,  $t(2879)=23.91$ ,  $p<0.01$ ,这一结果证明在本实验的模型中,出现与人类被试相似的快同效应。

## 2.4 讨论

研究 1 通过构建 FSNet 以捕捉人类在异同任务中的反应特征,分析结果发现,与人类被试的反应数据相比,模型在反应时分布及快同效应的指标上均与人类被试表现一致,证明模型能有效模拟人类的反应特征。通过对实验数据的编解码和人工神经网络模型,本实验构建了与人类相似的知觉决策模型。

## 第3章 研究2：人工神经网络中快同效应的网络特性

### 3.1 研究目的

本研究的目的在于对 FSNet 模型的内部结构进行研究，分析 RNN 网络活动在时间维度上的变化情况。从具体算法表征层面研究网络的信息处理机制，并尝试解释模型的计算理论，为现有的理论提供进一步证据，完善异同判断任务的理论模型，并以此来揭示快同效应的认知机制。本研究的目标是验证本文提出的两个假设：

1. 在异同判断任务中，慢速比较器和同一性指示器是顺序加工的。
2. 在异同判断任务中，对相同刺激的加工和不同刺激的加工使用的是不同的加工过程。

### 3.2 研究方法

#### 3.2.1 神经网络隐藏层的 PCA 方法

在机器学习的研究中，研究者可以通过分析网络内部参数的时间和空间分布来研究网络的内部结构。主成分分析法(principal component analysis, PCA)是一种在机器学习和数据分析中广泛使用的降维方法。PCA 指的是通过将数据点投影到一个超平面，并选取数个主成分，要求这些主成分能够解释尽可能大的方差并互不相关，使数据点在超平面上的投影尽可能分开的一种特征提取算法(刘敬浩, 孙晓伟, 金杰, 2020)。PCA 能够提取影响数据分布形态的最主要特征，并且在尽可能保留原始信息的前提下，使数据精简，便于分析和研究。

心理学研究中，PCA 常被用于抽取统计数据的因子，以此来验证数据的主要维度是否与预期的理论维度一致。而在机器学习的应用中，PCA 可以对神经网络内部的隐藏层进行分析，对隐藏层的向量进行降维，进而来研究网络层中的潜在结构。

在本研究中，将使用 PCA 方法对 FSNet 中的 RNN 网络进行分析，抽取有效因子，并对因子在时间维度上的变化情况进行分析。RNN 网络的运算基于隐藏层的向量，隐藏层在每一时刻得到更新，并在下一时刻作为“记忆”输入，隐

藏层包含神经网络是如何运算的关键信息。因此，本研究将使用 PCA 对隐藏层的活动状态进行分析，尝试分析网络活动过程中的主要特征因子。

### 3.2.2 模型训练和测试方法

使用人类被试数据划分出的训练数据对 FSNet 进行训练，采用 Adam 优化器和 MSELoss 均方损失函数对网络的参数进行优化，设定 loss 阈值为 0.034，达到阈值后停止模型的训练。为保证分析的模型不存在训练偏差，将该模型输出的反应时均值与分布偏度与研究 1 中 40 次训练的分布做 t 检验，结果显示并不存在显著差异，该模型能够代表普遍的训练结果。

采用人类被试数据划分出的测试数据集对 FSNet 进行测试，使用该测试集对模型进行 72 次输出，记录每一次输出中网络隐藏层的变化情况，将 72 对隐藏层变化序列拼接，得到后续分析需要的网络活动情况数据。

## 3.2 结果

### 3.2.1 PCA 结果

通过 python 程序语言的 sklearn 包(Pedregosa et al., 2011)中的 PCA 模块对拼接后的 RNN 网络的隐藏层向量进行分析，得到网络隐藏层的主要因子，因子解释率如图 8。为了确保因子能够更多的解释网络的活动特性同时保证因子的概括性，参考了 Yang 等人(2019)使用 ANN 在直觉决策任务上的筛选标准，最终选取 3 个因子。此时方差解释率超过 85%。

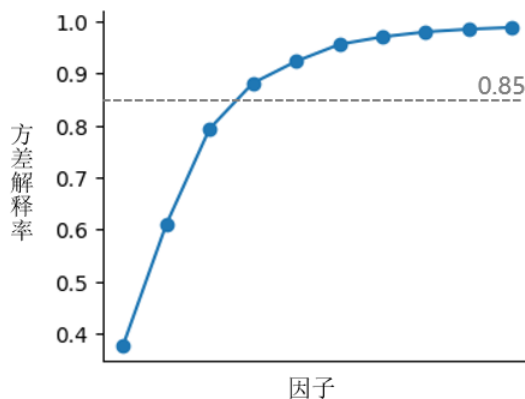


图 8 隐藏层活动的因子累计解释率

对原始的网络输出数据根据上述3个抽取的因子进行转换，得到这3个因子在时间维度上的变化曲线情况，见图9。每一条曲线代表一个试次中网络的活动情况。其中，红色曲线代表72次输出中“不同”刺激试次的网络变化情况，蓝色曲线代表72次输出中“相同”刺激试次的因子变化情况。绿色虚线代表做出相同判断的平均时间点，棕色虚线代表做出不同判断的平均时间点，黑色虚线分割不同的实验阶段，0-1为注视阶段，2-6为刺激呈现阶段，7-21为反应阶段。

曲线的y轴代表了因子的值，该值来自于神经网络隐藏层向量的运算结果，反映了隐藏层的活动情况。对于同一条试次曲线内，曲线数值随时间变化的趋势反映了模型的活动情况，高斜率的折线代表了网络在持续活动，对输入刺激和隐藏层数据进行运算。对于不同试次间的曲线，如果在某一时段曲线的分布较广，即试次间曲线值方差较大，则说明网络在这一时段对输入刺激较为敏感，反映了网络的活动水平较高。

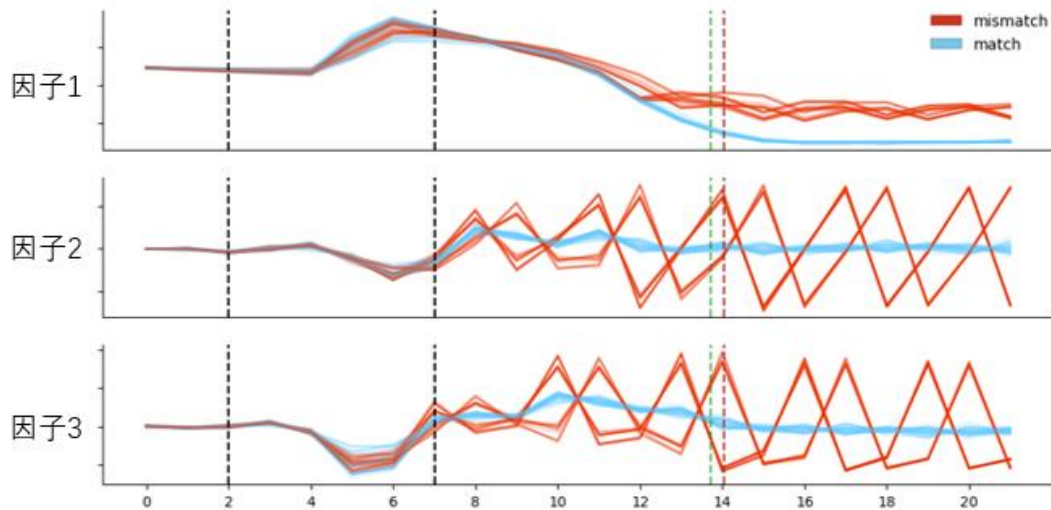


图9 各因子在时间维度上的变化情况

因子1的解释率为0.437，绘制成最上方的子图，因子时间变化曲线显示，从刺激呈现的时间点开始，相同和不同判断的试次均出现一个较大峰值，这一峰值在7时间步，刺激程序完成开始反应的时间点开始下降，对于“不同刺激”条件的试次，试次中的因子1的活动差异较大( $mean\_var = 0.0021$ )，对于“相同刺激”条件的试次，试次之间活动差异较小( $mean\_var = 0.0038$ )。这说明，“相同刺激”条件下，不同输入刺激与因子1的活动水平并没有显著关联性，而“不

同刺激”条件下，输入刺激会引起因子1的明显活动。“相同刺激”和“不同刺激”在因子1层面上表现出明显异质的活动。

因子2的解释率为0.158，绘制成中间子图，因子时间变化曲线显示，从刺激呈现的时间点开始，相同刺激和不同刺激试次均出现波动，相同试次的波动波幅较小( $mean\_var = 0.032$ )，而不同刺激的波动较大( $mean\_var = 2.2$ )且逐渐增大，在反应时间点附近达到峰值。因子2中“相同刺激”与“不同刺激”也表现出明显的活动异质性，

因子3的解释率为0.143，因子时间变化曲线显示，绘制成下方子图，因子3在刺激呈现阶段出现一个负峰值，且“相同刺激”的曲线间差异( $mean\_var = 0.22$ )比“不同刺激”的曲线间差异( $mean\_var = 0.066$ )较大，这可能说明相同刺激在因子3在刺激呈现阶段对输入刺激活动更敏感，活动水平更高。在反应阶段，因子3出现了类似因子1的活动情况，“不同刺激”条件的试次中活动差异较大( $mean\_var = 0.0021$ )，“相同刺激”条件试次之间活动差异较小，但因子1的波动起始时间要较为提前。

### 3.2.2 对因子的解释

根据对三个因子在时间维度上活动的描述，整体来看，FSNet处理“不同刺激”条件的试次与“相同刺激”条件的试次的过程中，活动水平和模式有着明显的差异，这种差异不仅表现在时间先后上，也表现在活动水平上，这可能说明FSNet对相同的刺激对和不同的刺激对进行处理时可能采用的是不同的算法。

从信息加工的角度来讲，RNN网络的隐藏层承担“记忆”的功能，如果随着时间输入的信息使隐藏层发生改变，则可以认为输入的新信息激发了网络的运算，改变了隐藏层存储的“记忆”。反之，如果随着时间隐藏层的值矩阵较为稳定，则可以认为外界输入的新信息没有被网络处理并改变记忆。因此，隐藏层矩阵的数据在某一时段趋于稳定，可能说明网络并没有在这一时间段上进行运算。网络隐藏层活动的平滑度，可以作为衡量网络对输入刺激以及已有信息的处理过程的活跃程度。

在因子1和因子3中，“相同刺激”的试次均在反应期出现了稳定的平滑段，在因子2中，也出现类似的平滑段，但其稳定性比因子1和因子3低，不过总体来看，要远高于“不同刺激”的试次。这表明这一时期网络可能活动水平较低，

而“不同刺激”的试次在三个因子中均有明显的波动，这说明网络隐藏层在持续参与运算并改变值，网络有较高的活动水平。根据这一解释，可以推测网络在反应期的活动特点：当输入相同刺激时，网络在反应期的活动水平较低，当输入不同刺激时，网络在反应期活动水平较高。

在因子3中的刺激呈现期中，出现了一个波，此时“相同”和“不同”两种条件下的因子3均有活动，但比起“不同”刺激，“相同”刺激的活动要更为强烈。这可能说明“相同”刺激的输入使得网络在刺激呈现期有着更多的活动。值得注意的是，相同刺激在刺激呈现阶段产生的活动波的结束时间步和不同刺激在反应期呈现的活动波开始时间步是相近的，这可能说明这两个活动是先后进行的。

在因子1的刺激呈现期中，“相同”和“不同”条件的刺激激发了网络相同的活动状况，因子1的值在刺激呈现后持续增长，并在呈现结束后迅速下跌，并且增长和下跌的过程都较为单调。对于RNN来说，持续的刺激呈现代表着将相同的刺激数据在一段时间步内持续输入到网络中参与运算，重复运算某个刺激带来的值矩阵变化很可能是单调的。因此，刺激呈现期因子1的活动很可能反映网络识别并存储输入刺激的过程。

### 3.2.2 快同效应的认知机制

根据对因子活动特点的描述和解释，我们能够结合现有的理论来证明本文的研究假设。

现有的理论认为，不论是异同判断的单过程模型还是双过程模型，都存在一个慢速比较器和一个同一性指示器来给出判断。争论的焦点在于这一过程是否是同时进行或者顺序进行的(Farell, 1985; 张之光, 禰宇明, 傅小兰, 2003)。

根据本研究对PCA分析结果的解释，FSNet在加工相同刺激和加工不同刺激时，有明显异质的加工过程，在相同刺激加工过程中，模型活动主要集中在刺激呈现阶段，而在不同刺激加工过程中，模型活动主要集中在刺激呈现后的反应阶段，而在刺激呈现阶段也有一定水平的活动。这两个加工过程与现有理论中的双过程模型的假设一致。这支持本研究的假设2。

FSNet在刺激呈现阶段，对于相同刺激和不同刺激均有加工，但相同刺激的加工比之不同刺激的活动水平较高，这可能说明不同刺激在刺激呈现之后有额外的加工过程。根据慢速比较器和同一性指示器的特性，相同刺激在呈现阶段的高

活动水平可能是由于编码和存储输入的刺激与同一性指示器的活动带来的。而不同刺激在刺激呈现之后仅进行编码和存储，而不进行额外的加工过程。

在反应期，FSNet 对于不同刺激有着更明显的加工，而对相同刺激的加工则几乎停止，并且对于不同刺激的加工过程是紧随着相同刺激在刺激呈现阶段的活动之后的。不同刺激在反应期的活动与双过程模型中假设的慢速比较器的活动规律是一致的。FSNet 对相同刺激处理和不同刺激处理这两个过程在时间上的顺序性，说明在异同判断任务中，可能基于刺激是相同或不同的而存在两个先后的加工过程，这支持了慢速比较器和同一性指示器是顺序进行加工的这一假说。

综上，根据相关理论和对模型活动的分析，可以对异同判断的过程做出解释，并以此解释快同效应的原理：

对于异同判断任务来说，加工很有可能是双过程的，并且其中的慢速比较器和同一性指示器是顺序加工的，对于输入的刺激，同一性指示器先进行处理，如果判断结果为相同，则直接输出结果而不继续后续加工，如果判断结果为不同，则使用慢速比较器后续进一步进行加工，最后输出判断结果。由于对相同刺激来说，只需要进行同一性指示器一个加工过程，而不同刺激不仅要进行同一性指示器的加工，还要调用后续的慢速比较器，需要更长的加工流程，进而导致了更长的反应时。因此，加工过程的不同导致了异同判断任务中快同效应的出现。

## 第4章 讨论

快同效应是异同判断这一任务中的稳定的结果模式,但其认知机制却仍然存在争议。本研究成功地训练出人工神经网络模型,能够复现出人类数据中的快同效应,并通过分析训练后的神经网络的特点,对异同判断任务中快同效应产生的假设进行了检验。在方法论上,本文使用人工神经网络这一新工具来研究快同效应,为今后的认知研究提供了借鉴。

本研究克服了人工神经网络决策时间依赖于硬件和软件运算速度的问题,使人工神经网络模型更“像人类”:传统的人工神经网络基于计算机科学,从本质上来说,比起基于神经生理学和神经解剖学的理论模型,生物学和神经生理学特性并不是 ANN 关心的重点(Yang & Wang, 2020)。因此人工神经网络不具有人类特有的反应特点,例如反应时。人工神经网络的决策时间基于硬件的运算速度和软件平台,这与人类真实的反应情况不符。有研究将神经网络模型纳入 DDM(Rafiei, Shekhar & Rahnev, 2023),使模型具有生成反应时的能力,但其产生反应时的过程来自于 DDM 而非神经网络,并且只适合于单一的视觉刺激任务。有研究者使用 RNN 来建模认知任务(Ehrliche et al., 2021),但未能使网络输出反应时。通过对人工神经网络模型进行改进,建立起与人类类似的反应时情况,能够使我们更加接近人类反应时的本质。本研究通过时间序列编码的方式,克服了神经网络无法输出反应时的弊端,使神经网络模型能够同时给出判断结论和反应时,这使得我们的模型更接近于真实人类被试的反应特点。

本研究通过 FSNet 验证了双过程理论,解决了长期的理论争议:双过程理论和单过程理论究竟哪个更符合与人脑真实的加工情况,目前的行为实验和脑电研究并没有给出明确的证据。一个原因是对于行为实验和脑电研究,无法直接深入认知过程本身,基于间接证据的解释为建立明确的加工过程理论带来了困难。

本研究采用人工神经网络直接对异同判断任务的加工过程本身进行分析,指出快同效应产生的原因在于对不同刺激对与相同刺激对的加工上的差异。这个结果与过去研究者使用行为实验与脑电实验得到的结果相符(魏景汉, 2010)。结合现有研究证据,本研究基于异同判断任务加工过程的直接证据,支持了异同判断任务的双过程理论,解决了长期的理论争议。



人工神经网络在本研究中的应用能为后续心理学研究提供多方面的启发,具有应用方面的潜力:

首先,人工神经网络能够建模复杂认知任务:得益于 ANN 的参数优化算法,ANN 能够自动学习认知任务的特点。本研究关注的异同判断任务的加工过程和反应特点会随着不同的任务条件发生改变(Farell, 1985; Walker & Cousineau, 2019),这说明异同判断任务是一个复杂的认知过程。快同效应一直是研究者建模异同判断任务中的一个难点(Harding & Cousineau, 2022)。ANN 能通过网络的复杂结构直接复现这一效应,展示了 ANN 在复杂任务建模上的优势。在此类复杂的认知任务上,ANN 有其独特优势。

除此之外,人工神经网络有可能取代人类被试作为实验对象:在目前的人工神经网络研究中,一类大规模参数的模型,如大语言模型(Large Language Model, LLM)在许多任务的成绩上比起传统的 ANN 模型取得了显著的进步,并且这类模型与人类有相似的行为和反应(Grossmann et al., 2023)。进而有研究者提出,LLM 有可能替代人类参与实验,以节省时间和成本,并可用于不适合人类参加的实验(Dillion et al., 2023)。有关 ANN 模型能否取代人类参与这一观点还存在许多争议,在本研究中,FSNet 模拟了参与异同判断任务的人类被试,对上述观点进行了初步尝试。为后续 ANN 在替代人类被试参与到心理学研究中的应用提供了理论支持。

神经网络工具在解决认知机制的问题上取得了一些成效,但作为新兴的工具,神经网络在应用层面还有诸多问题:

第一个问题是神经网络的可解释性。虽然分析 ANN 的结构比分析生物神经网络要更加容易,但对具有深层次结构的神经网络的分析则非常复杂。如果研究者希望模型能够处理更多不同情景下的任务,那研究者可能需要使用更复杂的网络结构和更深的参数层,这给理解网络的运行机制带来了困难(Yang & Wang, 2020)。由于 ANN 的训练基于优化算法来搜索参数值,对于复杂的神经网络来说,一个认知任务可能存在多个可行解,如何去分析这些不同的优化结果,如何解释这些结果与真实人类被试的心理机制之间的关系,仍旧需要更多的探索。

第二个问题是神经网络是否能够反映人类认知活动的结构。在一些基于神经网络的认知模型中,神经网络可能并不具备反映认知过程的能力。有研究者采用

CNN 神经网络和 DDM 决策模型共同构建混合模型来完成 MINST 手写数字数据集(LeCun et al., 1998)识别任务。该模型使用 CNN 来替代 DDM 模型中证据积累的过程,使模型具有生成反应时的能力(Rafiei et al., 2023)。但该模型中反应时生成过程本质上是基于 DDM 模型的, CNN 承担的主要功能则是对刺激进行模式识别,并输出置信度作为证据。模型的反应时输出主要基于 CNN 对刺激预测的概率,这一关联是由人工设定的模型结构决定的,而不是神经网络中内部生成的结构。这类模型能够为神经网络在传统的机器学习任务甚至心理学的认知任务上获得人类反应的特性,但其本质上是基于人工建构的理论模型,它能够反映潜在的人类认知活动中的心理结构仍旧是需要研究的问题。

还有研究者提出,神经网络并非真实学习了认知任务本身,而是由于其巨量参数而形成了对数据的过拟合。针对这一问题,有相关人工智能方面的工作表明,与我们的直觉不同,人工神经网络这样的大规模参数系统可能反而具有某种内在的数学特性,使其能够良好的泛化,这预示了神经网络或许能够自发生成某种特定的结构(Neyshabur et al., 2019)。还有研究者提出,人脑本身就是一个包含大量参数(例如突触连接,离子通道密度等)的适应性结构,神经网络的巨量参数特点反而适合于作为大脑的模型(Richards et al., 2019)。

## 第 5 章 总结与展望

对于在异同判断任务中发现的快同效应,研究者提出两种对立的理论来解释,这两种解释争论的焦点在于处理相同刺激对和不同刺激对的认知过程是否是同一个过程。研究者通过行为实验和脑电研究来研究这个问题,并基于人工建构的认知模型来解释异同判断任务和快同效应,但目前对于快同效应的产生原因还没有一个明确的解释。我们希望建立神经网络模型来研究异同判断任务的加工过程以及快同效应产生的机制,为异同判断的理论构建提供更多证据。

在本文的第一部分中,我们通过构建人工神经网络模型 FSNet,与人类被试的反应数据比对,发现 FSNet 在反应时分布,及快同效应的出现指标上均与人类被试的表现一致,证明本实验构建的模型有能力复现人类被试在异同判断任务上的反应特点。通过对实验数据的编解码和人工神经网络模型,本实验构建了与人类相似的知觉决策模型。

在本文的第二部分中,通过神经网络分析方法对 FSNet 模型的内部结构进行研究,分析模型中 RNN 网络活动在时间维度上的变化情况,从具体算法表征层面研究网络的信息处理机制,验证本文提出的两个假设: 1.对于异同判断任务来说,加工是双过程的,即对于相同刺激的加工和不同刺激的加工使用的认知过程是不同的。2.在异同判断中,慢速比较器和同一性指示器是顺序进行加工的。

综上,本文的研究结论如下:

(1)基于 RNN 神经网络,本研究建构了与人类被试有相同反应特点的计算机模型。该模型在反应时间、反应时间分布以及正确率等多个指标上展现出与人类被试相似的反应特征,并成功地模拟了显著的快同效应。

(2)基于对网络参数活动状态的分析,研究结果支持异同判断任务中双过程模型理论,并表明慢速比较器和同一性指示器可能是顺序加工的。快同效应的产生可能源于对相同刺激和不同刺激加工过程的不同。

本研究还有许多可以改进的部分,首先是任务类型上,本研究仅选择同时呈现刺激条件的异同判断任务作为建模对象,没有考虑刺激相继呈现条件下的异同判断任务。因此,现有的模型可能难以很好的推广到其他条件中去,在后续研究中,可以纳入多种任务条件下的数据,使模型具有更全面的行为能力。

然而, 本研究也存在一些局限性。首先, 本研究采用 PCA 分析对网络内部的动态活动情况进行分析, 而没有对网络静态的参数分布进行分析。机器学习领域有许多方法对网络的静态参数进行分析, 因此后续研究需要更多领域的研究人员合作, 对网络的具体结构进行深入研究。在 PCA 分析中, 由于无法比较因子的具体数值, 我们仅能评估不同条件下活动水平的相对变化。其次, 对于异同判断中的慢速比较器与同一性指示器是否存在功能重叠, 本研究无法提供确切的验证。未来的研究应采用更精细的参数分析技术以深入探究这一问题。最后, 受限于设备处理性能, 本研究仅选取单个代表性模型进行分析, 可能无法体现所有训练结果的总体情况。在后续的研究中, 可以采用更合适的算法来对多个模型的训练结果进行分析。

## 参考文献

- 常晶晶, 刘强, 邱江. (2012). 异同判断中“快同效应”的大脑机制:一项 ERP 研究. *心理科学*, 35(4), 829 – 835.
- 刘敬浩, 孙晓伟, 金杰. (2020). 基于主成分分析和循环神经网络的入侵检测模型. *中文信息学报*, 34(10), 105 – 112.
- 刘逸康, 胡传鹏.(2024).证据积累模型的行为与认知神经证据. *科学通报*, 69(8), 1068-1081.
- 魏景汉. (2010). 事件相关电位原理与技术. *生物技术通报*, 26(6), 250.
- 张之光, 禰宇明, 傅小兰. (2003). 不同 isi 下异同判断的同反应启动效应. *心理学报*, 35(3), 317-322.
- Abbott, L.F. (2008). Theoretical neuroscience rising. *Neuron*, 60, 489–495.
- Bamber, D. (1969). Reaction times and error rates for “same”-“different” judgments of multidimensional stimuli. *Perception & Psychophysics*, 6, 169-174.
- Barlow, H.B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication*, 1, 217–234.
- Chignell, M. H., & Krueger, L. E. (1984). Further evidence for priming in perceptual matching: Temporal, not spatial, separation enhances the fast-same effect. *Perception & Psychophysics*, 36(3), 257–265.
- de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, 8(85), 5351.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arxiv preprint arxiv:1810.04805*.
- Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can AI language models replace human participants?. *Trends in Cognitive Sciences*, 27(7), 597-600.
- Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating?. *Behavioral and brain sciences*, 11(3), 357-374.
- Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, 51(4), 380–417.
- Ehrlich, D. B., Stone, J. T., Brandfonbrener, D., Atanasov, A., & Murray, J. D. (2021). PsychRNN: An Accessible and Flexible Python Package for Training Recurrent Neural Network Models on Cognitive Tasks. *eNeuro*, 8(1).

- Erhan, D., Bengio, Y., Courville, A., & Vincent, P. (2009). Visualizing higher-layer features of a deep network. *University of Montreal*, 1341(3), 1.
- Fabian, P. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12, 2825.
- Farell, B. (1985). "Same"-"different" judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin*, 98(3), 419–456.  
<https://doi.org/10.1037/0033-2909.98.3.419>
- Fukushima, K., and Miyake, S. (1982). Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognit*, 15, 455–469.
- Grossmann, I., Feinberg, M., Parker, D. C., Christakis, N. A., Tetlock, P. E., & Cunningham, W. A. (2023). AI and the transformation of social science research. *Science*, 380(6650), 1108-1109.
- Goulet, M. A., & Cousineau, D. (2020). Sequential sampling models of same-different data and how they explain the fast-same effect. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 74(4), 284.
- Goulet M-A, Cousineau D. The fast–same effect of an exclusive-OR task. *Journal of Experimental Psychology: Human Perception and Performance*, 2020, 46(9): 991–1000.
- Harding, B., & Cousineau, D. (2022). Is the fast-same phenomenon that fast? An investigation of identity priming in the same-different task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(4), 520.
- Harding, B. (2018). A single process model of the same-different task (Doctoral dissertation, Université d'Ottawa/University of Ottawa).
- Harding, B., & Cousineau, D. (2022). Is the fast-same phenomenon that fast? An investigation of identity priming in the same-different task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(4), 520–546.  
<https://doi.org/10.1037/xlm0001076>
- Helmstaedter, M., Briggman, K.L., Turaga, S.C., Jain, V., Seung, H.S., and Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500, 168–174.
- Ila, A. B., & Polich, J. (1999). P300 and response time from a manual Stroop task. *Clinical neurophysiology*, 110(2), 367-373.

- Januszewski, M., Kornfeld, J., Li, P.H., Pope, A., Blakely, T., Lindsey, L., Maitin-Shepard, J., Tyka, M., Denk, W., and Jain, V. (2018). High-precision automated reconstruction of neurons with flood-filling networks. *Nat. Methods*, 15, 605–610.
- Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C.E., Kepecs, A., Mainen, Z.F., Qi, X.-L., Romo, R., Uchida, N., and Machens, C.K. (2016). Demixed principal component analysis of neural population data. *eLife*, 5, e10989.
- Kornblith, S., Norouzi, M., Lee, H., and Hinton, G. (2019). Similarity of Neural Network Representations Revisited. arXiv, 1905.00414 <https://arxiv.org/abs/1905.00414>.
- Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2, 249.
- Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.*, 25, 1097–1105.
- Krueger, L. E. (1978). A theory of perceptual matching. *Psychological Review*, 85(4), 278–304.
- Krueger, L. E. (1983). Probing Proctor's priming principle: The effect of simultaneous and sequential presentation on same–different judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(3), 511.
- Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science*, 197(4305), 792–795.
- LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10), 1995..
- Lindsey, J., Ocko, S. A., Ganguli, S., & Deny, S. (2019). A unified theory of early visual representations from retina to cortex through anatomically constrained deep CNNs. *arXiv preprint arXiv:1901.00945*.
- Neyshabur, B., Li, Z., Bhojanapalli, S., LeCun, Y. & Srebro, N. (2019). The role of over-parametrization in generalization of neural networks. in International Conference on Learning Representations (ICLR) 2019.
- Nickerson, R. S. (1965). Response Times for “Same”-“Different” Judgments. *Perceptual and Motor Skills*, 20(1), 15–18.

- Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, 503(7474), 78-84.
- Marr, D., & Vision, A. (1982). A computational investigation into the human representation and processing of visual information. WH San Francisco: Freeman and Company, San Francisco, 1, 1.
- Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of userdefined body parts with deep learning. *Nat. Neurosci*, 21, 1281–1289.
- O'Connell, R. G., & Kelly, S. P. (2021). Neurophysiology of human perceptual decision-making. *Annual Review of Neuroscience*, 44, 495-516.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. <https://pytorch.org>
- Posner, M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychological review*, 74(5), 392.
- Proctor, R. W. (1981). A unified theory for matching-task phenomena. *Psychological Review*, 88(4), 291.
- Proctor, R. W., & Rao, K. V. (1983). Evidence that the same-different disparity in letter matching is not attributable to response bias. *Perception & Psychophysics*, 34, 72-76.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2).
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, 20(4), 260–281.
- Reed, S. K. (2013). Psychological processes in pattern recognition. Academic Press.
- Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., ... & Kording, K. P. (2019). A deep learning framework for neuroscience. *Nature neuroscience*, 22(11), 1761-1770.
- Rigotti, M., Rubin, D. B. D., Wang, X. J., & Fusi, S. (2010). Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. *Frontiers in computational neuroscience*, 4, 24.



- Sussillo, D., & Barak, O. (2013). Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural computation*, 25(3), 626-649.
- Sussillo, D., Churchland, M. M., Kaufman, M. T., & Shenoy, K. V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. *Nature neuroscience*, 18(7), 1025-1033.
- Rafiei, F., Shekhar, M., & Rahnev, D. (2022). RTNet: A neural network that exhibits the signatures of human perceptual decision making. *bioRxiv*, 2022-08.
- Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Physica D: Nonlinear Phenomena*, 404, 132306.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Hassabis, D. (2017). Mastering the game of go without human knowledge. *nature*, 550(7676), 354-359.
- Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron*, 80(3), 791-806.
- Testolin, A., De Filippo De Grazia, M., & Zorzi, M. (2017). The Role of Architectural and Learning Constraints in Neural Network Models: A Case Study on Visual Space Coding. *Frontiers in Computational Neuroscience*, 11.
- Uchida, N., Kepecs, A., & Mainen, Z. F. (2006). Seeing at a glance, smelling in a whiff: Rapid forms of perceptual decision making. *Nature Reviews Neuroscience*, 7(6), 485–491.
- Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review*, 7(2), 208–256.
- Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory and Cognition*, 32(7), 1206–1220.
- Walker, J. A., & Cousineau, D. (2019). Into the Mind’s Eye: Exploring the Fast-Same Effect in the Same-Different Task. *The American Journal of Psychology*, 132(4), 421–437. <https://doi.org/10.5406/amerjpsyc.132.4.0421>
- Williams, A.H., Kim, T.H., Wang, F., Vyas, S., Ryu, S.I., Shenoy, K.V., Schnitzer, M., Kolda, T.G., and Ganguli, S. (2018). Unsupervised discovery of demixed,

- low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron*, 98, 1099–1115.e8.
- W. James. “Principles of Psychology,” Wiley, 1984.
- Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T., & Wang, X. J. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nature neuroscience*, 22(2), 297-306.
- Yang, G. R., & Wang, X.-J. (2020). Artificial Neural Networks for Neuroscientists: A Primer. *Neuron*, 107(6), 1048–1070.
- Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3), 356-365.
- Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the national academy of sciences*, 111(23), 8619-8624.
- Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. “Gradient-based learning applied to document recognition.” *Proceedings of the IEEE*, 86(11):2278-2324, November 1998.

## 致谢

在致谢这一部分，首先要感谢我的导师传鹏，这篇论文的选题是我之前不曾尝试但却十分感兴趣的，虽然遭遇了很多困难和失败，但最终能把它完成，多亏了传鹏的帮助和鼓励。在论文之外，传鹏的严谨的学术态度与对开放科学的热情也使我受益颇多。

也很感谢 Lab 的小伙伴们在学业上给予的帮助，现在觉得，当初选择加入我们的 Lab 真是一个非常正确的决定。还要感谢帮助过这篇论文肖剑同学，晚柯同学和元瑞同学，对于并不熟悉的神经网络部署调试，多亏了肖剑同学提供的支持；也很感谢提供了模型训练数据的元瑞同学；建模中途遇到的许多细节和概念问题，晚柯也帮忙一起理清了思路，十分感谢。

在完成论文期间，遇到许多生活和学习上的杂事，心院的刘老师和露姐总会全心帮忙协助解决，保证了我校园生活的精神状态，十分感动。

还要感谢校园里的小鸟，以及大晚上和白天蹲在池塘和河里的夜鹭，有空的时候我会带上相机和望远镜去校园的芳菲池和后山看看他们，让我疲劳的精神得到舒缓，十分感动。

感谢洛天依和星尘以及许多创作者给我带来的歌曲，一整天的写代码让我的精神有点迷糊，但是她们的歌声弥补了这一点。

感谢我的朋友们，浙江系，福建系和五湖四海的小伙伴们在这一年里的陪伴，支持和鼓励。谢谢你们在繁忙的考研毕业之间策划了一次难忘的旅行，十分感慨。

最后也是最重要的，感谢我的家人和我的爱人小邱。爸妈虽然平日里并不太多表达，但是每一次我能够取得小小进步和成功之后，都能从你们的言语中感受到你们的爱。

感谢我最爱的人小邱同学，在每一次我遇到挫折，不安，焦虑的时候，都会伸出手让我冷静下来，让我觉得有万分的勇气去面对所有的困难和挑战。从此不再是一个人的山遥路远，而是有一个人与我一同推开世界的门。

谢谢大家！

## 附录：预实验

### 1.研究目的

要能够学习到异同判断任务中的特殊结构，需要选择合适的模型结构，而由于训练 ANN 模型需要较高时间成本，在预实验中，将进行简略的尝试，对不同模型结构的指标进行比对，选择最优的模型结构。

### 2.模型结构简述

现有的使用 ANN 的建模思路主要有以下两种：1.主体基于 ANN，对出入输出进行特定编码的结构 2.使用 ANN 和其他模型结构结合，使 ANN 成为其他模型的某个特定工作环节。以下将分别介绍这两种思路

#### 2.1 基于 ANN 的模型

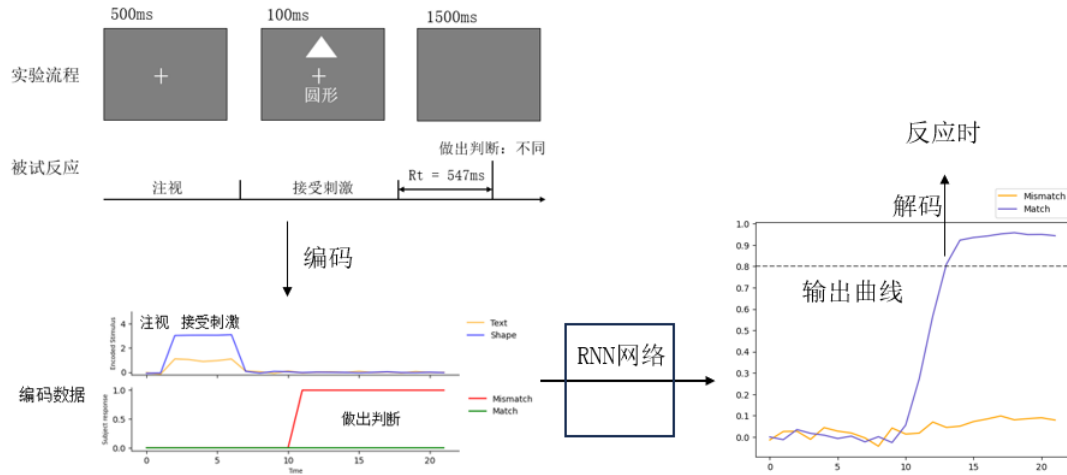
主体基于 ANN 的建模思路，将 ANN 模型作为模型主要的处理部分，但是对输入和输出按照某种规律做某种等价的变换，使其输出结果更具有清晰意义，这一种方式可以采用 CNN，RNN 或者全接连网络来实现。

现有的人工神经网络有许多常用的网络结构，例如卷积神经网络(CNN，适合处理空间，图像信息)，循环神经网络(RNN，适合处理序列信息)，生成对抗网络(GAN，适合进行内容的生成)等。由于本研究关注的反应时属性涉及时间信息的处理，因此本研究采用循环神经网络(Recurrent Neural Network，RNN)来作为模型的主体结构，RNN 是一类能够处理时序信息的神经网络。不同于 CNN，在认知科学的研究中，RNN 常被用于处理特定的认知任务。研究者通常会选定特定的认知任务来让人或动物进行执行，并用同样的任务来训练 RNN。因此，本研究基于 ANN 的模型将选用 RNN 作为模型基础

RNN 是一种对序列数据建模的神经网络，在每一时刻上，网络都会进行一次运算，结合存储的记忆信息对输入的信息进行处理，再将处理的结果作为新的记忆。一个标准的 RNN 网络的结构包括输入层、输出层和隐藏层(记忆)，隐藏层的信息在时间维度上不断继承更新，类似人脑中“记忆”的功能，隐藏层的输入不仅包括输入层的输出，还包括上一时刻隐藏层的输出，网络将隐藏层的输出

再接回隐藏层，形成一个循环。比较于传统的前馈网络，RNN 实际上是给网络加入了记忆单元，使其具有了处理时间序列信息的能力。

基于 RNN 的模型架构主要包括三个部分：1.输入编码部分 2.神经网络模型部分 3.输出解码部分，这三个部分的详细介绍见正文 2.2.2 模型架构。



附录图 1 RNN 网络模型的结构

## 2.2 ANN 和其他模型混合结构

使用 ANN 和其他模型结构结合，这种方法使用 ANN 充当其他模型的某个特定结构，来执行数据处理，例如基于 DDM 漂移扩散模型和卷积神经网络 (Convolutional Neural Networks, CNN) 神经网络的 RTNet 模型，CNN 取代 DDM 中原有的通过数学函数拟合的证据积累过程进行证据积累(Rafiei, Shekhar & Rahnev, 2023)。

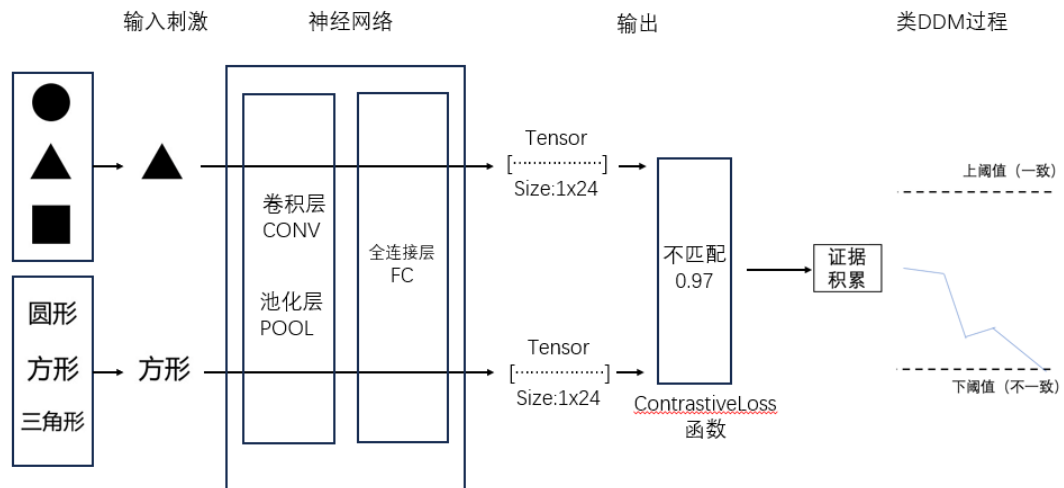
本研究采用证据积累模型(evidence accumulation model, EAM)与神经网络结合，这类模型认为人的决策遵循证据积累机制，外部刺激的呈现会促使认知系统中不断积累对于不同判断的证据，当某一判断证据的积累达到了阈值，决策者就会做出决策，证据积累的时间就是决策者的反应时。其中，最广泛使用的是漂移扩散模型(Drift Diffusion Model, DDM) (刘逸康, 胡传鹏, 2024)。

DDM 模型认为证据积累的过程主要由四个参数来表示: 漂移率(drift rate,  $v$ ), 代表接近某一阈值的平均速率, 表示每时间单位吸收的信息的相对数量; 偏向(bias)或者初始点(starting point)  $z$ , 代表反应前的初始偏向; 决策阈限(Threshold,  $a$ ), 反映的是两个决策阈值之间的距离; 以及非决策时间(non-decision time,  $t$ ), 表示

反应时的非决策部分。DDM 能够很好地解释反应时和正确率，并解释两者的关联，DDM 还可以很好显示出被试在决策过程中的非决策部分、反应偏向、反应标准和证据积累速度等信息。DDM 模型能够对反应时进行建模，因此，此结构将 DDM 模型的理论结构与各类神经网络进行结合，理论上能够使得 ANN 具有反应时的特点。

基于证据积累的孪生网络模型：

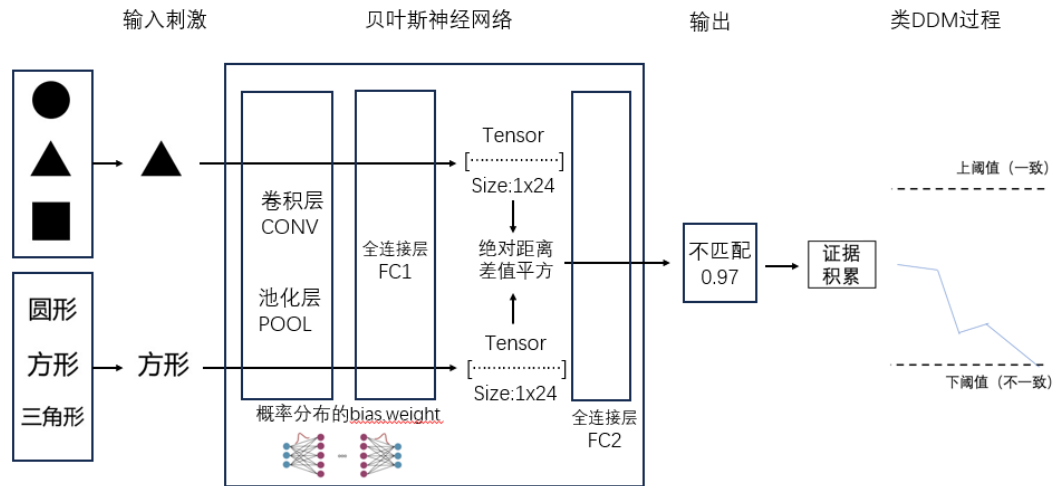
网络的结构如附录图 2，直接给神经网络呈现两个刺激，经由相同的神经网络结构处理，对网络输出的两个刺激编码后向量使用特殊的 Loss 损失函数进行处理，优化网络参数。网络根据输入的两个刺激，输出对于异同的判断的置信度，根据置信度结果进行证据积累，输出反应时。



附录图 2 孪生神经网络模型的结构

基于证据积累的贝叶斯 CNN 模型：

网络的结构如附录图 3，为了引入随机性，使用贝叶斯参数分布的神经网络的结构，使 CNN 网络具有输出随机性。网络将同时处理两个刺激，在内部对其异同进行比较，最后输出对刺激异同情况的判断以及置信度。



附录图 3 贝叶斯神经网络模型的结构

### 3.模型训练数据

实验数据来自研究组内研究者先前的研究数据(详见实验 1 方法部分), 为了加快训练的速度, 仅采取了其中 10%的数据进行训练, 其中 70%作为训练集, 30%作为测试集。采用 Python 程序语言和基于 Python 的 Pytorch 库来对模型进行建构和训练。

### 4 结果

基于证据积累的孪生网络模型直接对刺激进行学习, 采用孪生神经网络的结构进行异同判断, 建立的人工神经网络证据积累模型能够输出反应时, 但是由于神经网络的输出结果不具有随机性, 反应时的分布情况不太符合人类真实的情况。

基于证据积累的贝叶斯 CNN 模型得到了符合人类情况的反应时分布, 输出由于具有随机性而避免了极大极小值产生的反应时爆炸。但是网络的输出无法稳定复现快同效应的发生, 对“相同刺激”和“不同刺激”试次的差异在不同的模型训练结果中不存在显著方向性。

综上，基于证据积累的模型能够学习异同判断任务，并给出反应时，并且有与人类类似的正确率情况，但采用孪生网络以及 CNN 贝叶斯神经网络的模型都无法产生快同效应，相同与不同的刺激条件下反应时不存在差异。

主体基于 RNN 的模型，在使用少量数据训练后，损失函数值到达收敛，对输出结果曲线进行绘制，反应时分布与人类类似。对输出的反应时数据进行统计分析，存在明显的快同效应。

综上，主体基于 RNN 的模型能够最好的复现原始实验数据中产生的快同效应，后续实验中将以这种架构构建模型。