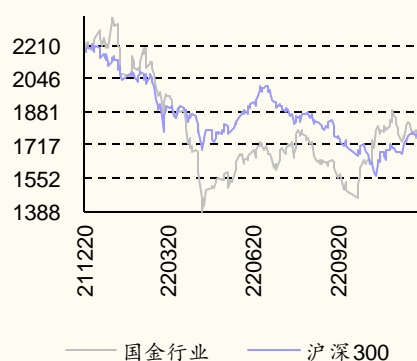


市场数据(人民币)

市场优化平均市盈率	18.90
国金计算机指数	1734
沪深300指数	3954
上证指数	3168
深证成指	11295
中小板综指	11959



相关报告

- 《高职院校教育新基建投资指南-教育信息化行业深度报告》，2022.12.14
- 《重磅车型放量，智能驾驶渗透率持续提升-智能汽车8月月报》，2022.9.13
- 《基本面拐点向上，龙头布局正当时-《2022-09-01行业中...》，2022.9.1
- 《畅享“新四化”，龙头布局时-计算机2022年中期策略》，2022.6.14
- 《金融科技板块21年报及22Q1梳理-《2022-05-03行...》，2022.5.4

王倩雯 分析师 SAC 执业编号: S1130522080001
wangqianwen@gjzq.com.cn

孟灿 分析师 SAC 执业编号: S1130522050001
mengcan@gjzq.com.cn

深度学习算法发展：从多样到统一

投资建议

- **行业策略：**深度学习的三要素包括算法、数据和算力，本文主要对算法的演进历程进行了回顾，认为深度学习底层算法被统一为 Transformer 之后发展放缓；而算法的行业落地应用、大数据的生成与处理、高算力芯片成为重点发展方向。
- **推荐组合：**建议关注受益于人工智能算法进步，并能成功实现商业化应用的海康威视、科大讯飞、中科创达、商汤科技等公司，以及关注可提供大算力 AI 芯片的海光信息等公司。

行业观点

- 神经网络的发展以 Relu 激活函数的提出为分水岭，可分为浅层神经网络和深度学习两个阶段。浅层神经网络阶段最重要的任务是解决梯度不稳定的问题，在这个问题未被妥善解决之前，神经网络受限于激活函数梯度过大或过小、以及神经元全连接对高算力的要求，因此应用性能不佳，而属于非神经网络的支持向量机 (SVM) 是当时解决人工智能模式识别的主流方法。
- 过去 10 年，深度学习经历了从多样化发展到融合统一的阶段。深度学习时代的开启依托于 2011 年 Relu 激活函数被提出、梯度消失问题被大幅缓解，此后深度学习算法和应用的发展均突飞猛进。最初卷积神经网络 (CNN) 通过对高层次特征的提取和压缩，擅长图像分类等任务；循环神经网络 (RNN) 通过对时序信息的提取，擅长文字、语音识别和理解等任务。2017 年 Transformer 的提出让深度学习进入了大模型时代、2020 年 Vision Transformer 的提出让深度学习进入了多模态时代。由于 Transformer 在大数据并行计算方面具备优势，且训练数据增长后对模型精度提升明显，自此各模态和各任务底层算法被统一为 Transformer 架构。
- 深度学习底层算法发展放缓，数据无监督学习、数据生成以及高算力芯片成为行业发展的重点方向。目前深度学习算法主要是基于 Transformer 骨干网络来进行分支网络的创新。如 OpenAI 在多模态主干网络 CLIP 的基础上引入扩散模型，即训练出能完成语义图像生成和编辑的 DALL·E2，引发 AIGC 浪潮；在 GPT-3 模型基础上引入了人类反馈强化学习方法 (RLHF)，训练出 InstructGPT 模型，并据此发布了对话机器人 ChatGPT，引起了互联网用户的注意。但随着 Transformer 基本完成底层算法统一之后，整个行业底层算法发展速度开始放缓，静待骨干网络的下一次突破。同时基于 Transformer 对大数据的需求，催生了无监督学习、高算力芯片的发展。

风险提示

- 海外基础软硬件使用受限；骨干网络创新放缓；应用落地不及预期

内容目录

1. 感知机：第一个神经网络.....	4
2. 多层感知机与 BP 算法——神经网络的再兴起.....	4
2.1 多层感知机解决了多元分类问题.....	4
2.2 BP 算法：神经网络训练的基本算法.....	5
3. 浅层神经网络在多应用领域各自演进.....	6
3.1 图像识别领域：“卷积”机制提取图像空间特征.....	6
3.2 自然语言处理领域：“循环”机制提取语言时序特征.....	6
3.3 支持向量机：深度学习出现之前的主流算法.....	7
4. 梯度消失问题的解决使神经网络向深层迈进.....	8
4.1 逐层无监督是深度学习的初次探索.....	8
4.2 ReLu 激活函数的提出开启了深度学习时代.....	8
5. Transformer 统一了各任务底层算法，开启多模态和大模型时代.....	9
6. 探索多模态主干网络.....	11
7. 探索 Transformer 时代的分支网络.....	12
7.1 AIGC 领域：多模态 Transformer+扩散模型.....	12
7.2 对话机器人 ChatGPT：语言 Transformer+强化学习模型.....	13
8. 小结及投资建议.....	13
9. 风险提示.....	14

图表目录

图表 1：感知机二元分类过程.....	4
图表 2：数据的线性可分性.....	4
图表 3：多层感知机通过引入隐藏层进行多元分类.....	5
图表 4：梯度消失使神经网络无法正常训练.....	5
图表 5：卷积神经网络典型结构.....	6
图表 6：卷积的实质是提取关键特征.....	6
图表 7：RNN 同时接受当前时刻输入与上一时刻输出的信息.....	7
图表 8：感知机与支持向量机目标函数不同.....	7
图表 9：在深度学习之前，支持向量机是解决人工智能任务的主流方法.....	7
图表 10：逐层无监督+BP 有监督解决梯度消失问题.....	8
图表 11：Sigmoid 和 tanh 函数存在梯度过大/过小问题.....	9
图表 12：残差学习进一步缓解了梯度消失问题.....	9
图表 13：Transformer 示意图.....	10
图表 14：GPT 系列模型参数呈指数级增长.....	10
图表 15：ViT 模型将图像当作文本进行处理.....	10
图表 16：每层 Swin Transformer 之间进行了类似 CNN 的 patch merging.....	11
图表 17：CLIP 多模态网络采用图像-文本联合训练.....	11

图表 18: 目前主流的多模态主干网络.....	12
图表 19: 扩散模型的前向扩散和反向生成过程.....	12
图表 20: DDPM 在 AIGC 领域的前沿成果.....	13
图表 21: 强化学习模型示意.....	13
图表 22: 神经网络各发展阶段经典模型梳理.....	14

深度学习的三要素包括算法、数据和算力，本文主要对算法的演进过程进行了回顾，认为过往神经网络的发展以 Relu 激活函数的提出为分水岭，分为浅层神经网络和深度学习两个阶段。

- 在浅层神经网络阶段，最重要的任务就是解决梯度不稳定的问题。在这个问题未被妥善解决之前，神经网络应用性能不佳，而属于非神经网络的支持向量机算法（SVM）是解决人工智能模式识别的主流方法。
- 2011 年 Relu 激活函数被提出、梯度消失问题被大幅缓解之后，神经网络进入深度学习时代，算法和应用的发展均突飞猛进。最初 CNN、RNN 等模型在不同的模态和任务中均各有擅长，2017 年 Transformer 的提出让深度学习进入了大模型时代、2020 年 Vision Transformer 的提出让深度学习进入了多模态时代，自此多模态和多任务底层算法被统一为 Transformer 架构。

目前深度学习算法主要是基于 Transformer 骨干网络进行分支网络的创新，如引入扩散模型、强化学习等方法。整个行业算法发展速度放缓，静待骨干网络的下一次突破。

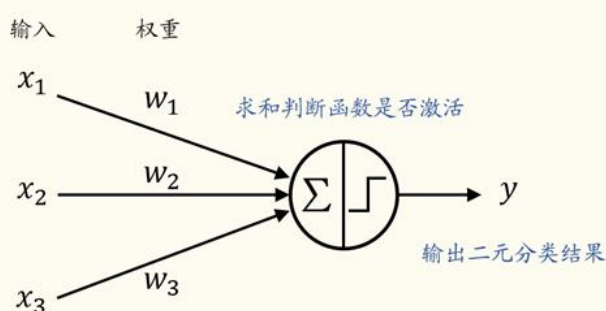
下文我们将对各发展阶段的经典模型进行回顾：

1. 感知机：第一个神经网络

感知机由 Rosenblatt 在 1958 年提出，是神经网络发展的起点。感知机是一个单层的神经网络模型，由输入值、权重、求和函数及激活函数三部分组成。整个模型对输入值进行有监督学习，权重部分是可训练参数；将对应输入与权重相乘求和，得到的求和值与常数比对，判断是否触发激活函数，最终依据输出的 0-1 信号判断图像类别。

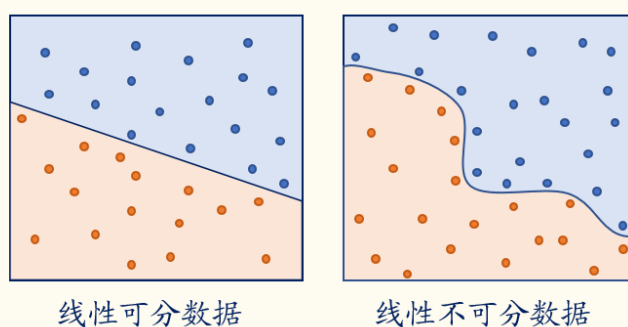
- 感知机提出了用神经网络模型解决人工智能任务。但作为神经网络模型的开山之作，还存在以下问题：
 - 1) 受阶跃激活函数限制，感知机只能输出 0 或 1，因此只能进行二元分类。
 - 2) 感知机只能处理线性可分数据，无法处理线性不可分的数据，而线性不可分数据是现实世界中的常态。该严重缺陷由 Minsky 于 1969 年提出，扼杀了人们对感知机的兴趣，也由此导致了神经网络领域研究的长期停滞。

图表 1：感知机二元分类过程



来源：AI 与强化学习公众号，国金证券研究所

图表 2：数据的线性可分性



来源：PaddleEdu，国金证券研究所

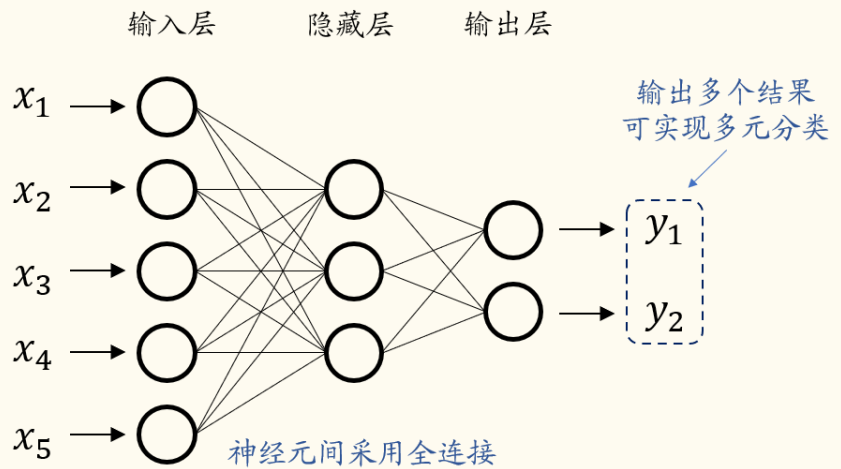
2. 多层感知机与 BP 算法——神经网络的再兴起

2.1 多层感知机解决了多元分类问题

20 世纪 80 年代，多层感知机（MLP）被提出。模型由输入层、输出层和至少一层的隐藏层构成，是一种全连接神经网络，即每一个神经元都会和上下两层所有的神经元相连接。各隐藏层中的神经元可接收相邻前序隐藏层中神经元的传递的信息，经过加工处理后将信息输出到后续隐藏层中的神经元。

由于隐藏层丰富了神经网络的结构，增强了神经网络的非线性表达能力，感知机的线性不可分问题得以解决，因而神经网络再次迎来兴起。

图表 3: 多层感知机通过引入隐藏层进行多元分类



来源: DeepAI, 国金证券研究所

■ 相较感知机，多层感知机主要进行了如下改进：

- 1) 解决了感知机的二元分类问题：引入隐藏层，并采用非线性激活函数 Sigmoid 代替阶跃函数，使得神经网络可以对非线性函数进行拟合。
- 2) 可进行多元分类任务：多层感知机拓宽了输出层宽度。

■ 多层感知机的发展受到算力限制。由于多层感知机是全连接神经网络，所需算力随着神经元的增加呈几何增长。而在算力相对匮乏的 20 世纪 80 年代，算力瓶颈阻碍了多层感知机的进一步发展。

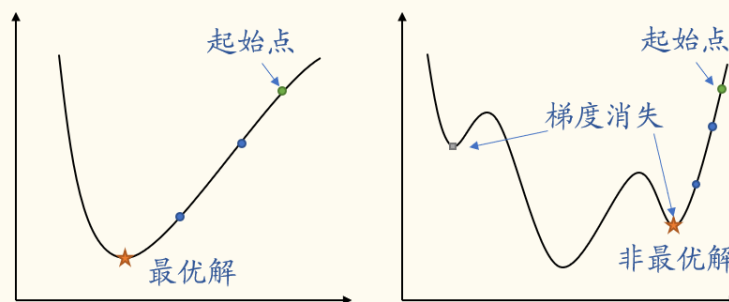
2.2 BP 算法：神经网络训练的基本算法

1986 年，Hinton 提出了一种适用于多层感知机训练的反向传播算法——BP 算法，至今仍是神经网络训练的主流算法。

- BP 算法的核心思想为：将输出值与标记值进行比较，误差反向由输出层向输入层传播，在这个过程中利用梯度下降算法对神经元的权重进行调整。
- BP 算法最大的问题在于梯度不稳定。由于当时 Sigmoid、Tanh 作为非线性激活函数应用广泛，而这两种激活函数都存在一定范围内梯度过大或过小的问题。神经网络停留在浅层时，连乘次数少、梯度较为稳定；而当神经网络向深层迈进，梯度的不稳定性加剧，使得深层神经网络无法正常训练。

因此，有效缓解梯度不稳定现象是神经网络向深层次迈进前必须解决的问题。

图表 4: 梯度消失使神经网络无法正常训练



来源: ScienceProg, 深度学习与图网络公众号, 国金证券研究所

3. 浅层神经网络在多应用领域各自演进

多层感知机的出现奠定了神经网络的基本结构，也使得神经网络的应用范围不再局限于图像识别，而是向自然语言处理、语音识别等其他领域拓展。由于各个领域的任务具有不同特点，神经网络产生了众多分支模型。这一阶段分支网络探索各自领域的任务特点，通过机制创新使神经网络获得对应的特征提取能力。

3.1 图像识别领域：“卷积”机制提取图像空间特征

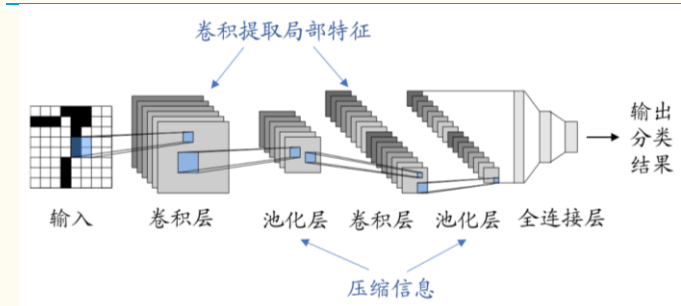
人类在进行图像识别时，能够从细小的特征推理得知事物的全貌，即“窥一斑而见全豹”。在多层感知机时代，由于二维图像被转化为一维向量后输入模型，因此丢失了图像的空间特征信息。为了使神经网络获得从特征到全局的图像识别能力，卷积神经网络应运而生。

1998年，LeNet 卷积神经网络（CNN）首次应用于图像分类。CNN 通过多个卷积层对特征进行提取和压缩，得到较为可靠的高层次特征，最终输出的特征可应用于图像分类等任务。

■ LeNet 网络由卷积层、池化层和全连接层构成：

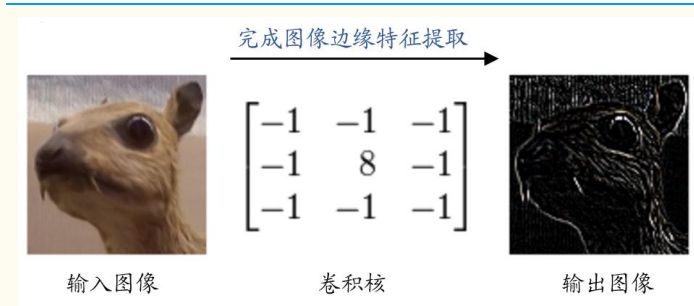
- 1) 卷积层：通过在输入图像上滑动卷积核，进行卷积操作提取关键特征。卷积核的尺寸比输入图像小得多，无需像多层感知机一样学习完整图片信息；
- 2) 池化层：对特征进行约减，从而提取主要特征，比如将卷积层的输出划分为若干矩形区域，取各区域的平均值或最大值，从而简化网络计算复杂度；
- 3) 全连接层：对提取到的特征进行汇总，将多维的特征映射为二维输出。

图表 5：卷积神经网络典型结构



来源：《Dive into Deep Learning》（Aston Zhang 等，2021），国金证券研究所

图表 6：卷积的实质是提取关键特征



来源：NVIDIA，国金证券研究所

■ 相较于多层感知机，卷积神经网络具备以下优点：

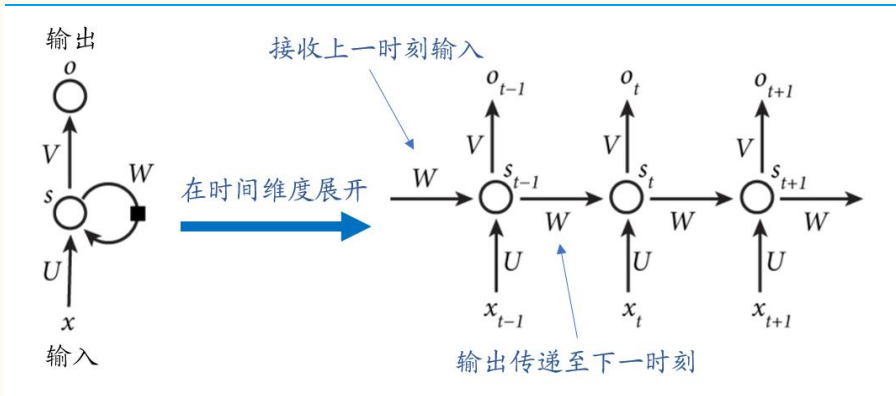
- 1) 计算量减少：神经元只与对应的部分局部连接；
- 2) 图像识别能力增强：利用卷积思想，使神经网络具备了局部特征提取能力；
- 3) 平移不变性：由于卷积核会在输入图像上滑动，所以无论被检测物体在图片哪个位置都能被检测到相同的特征。

3.2 自然语言处理领域：“循环”机制提取语言时序特征

人类在进行文字阅读、语音识别时，不仅会关注当前看到、听到的词句，还会联系上下文进行辅助理解。在多层感知机时代，所有的输入彼此独立，模型仅针对当前词句进行训练，而不关注前后信息，造成了时序信息的丢失。

为了使神经网络获得时序信息提取能力，1986年循环神经网络（RNN）被提出，将循环思想引入神经网络。在 RNN 中，每个神经元既接受当前时刻输入信息，也接受上一时刻神经元的输出信息，使神经网络具备了时序特征提取能力。

图表 7: RNN 同时接受当前时刻输入与上一时刻输出的信息



来源: Python 与机器学习公众号, 国金证券研究所

■ 相较多层感知机, 循环神经网络具备以下优点:

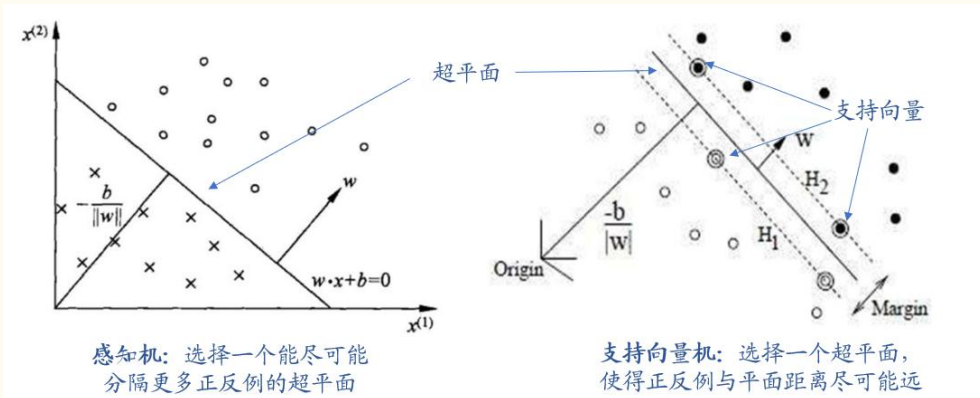
- 1) 可处理文本、语音等时序相关的信息;
- 2) 计算量大大降低: 在循环神经网络进行训练时, 参数 W 、 U 、 V 是共享的, 这减少了训练所需的计算量。

3.3 支持向量机: 深度学习出现之前的主流算法

支持向量机 (SVM) 是传统机器学习中最重要分类算法之一, 作为有监督的分类器, 其核心思想与感知机类似, 但对目标函数进行了改进:

- 感知机目标函数: 错误分类点到超平面的距离之和最小化。
- 支持向量机目标函数: 支持向量 (距离超平面最近的正反例) 到超平面的距离最大化。

图表 8: 感知机与支持向量机目标函数不同



来源: 人工智能技术公众号, AI 与强化学习公众号, 国金证券研究所

支持向量机的思想最早在 1963 年由 Vapnik 等人提出, 在 90 年代经过多次模型和算法改进, 能够解决感知机的线性不可分和多元分类等缺陷, 并可应用于人体动作识别、人脸识别、文本识别、生物信息学等领域。

图表 9: 在深度学习之前, 支持向量机是解决人工智能任务的主流方法

时间	理论提出者	理论内容
1963 年	Vapnik 等	在解决模式识别问题时提出支持向量方法
1971 年	Kimeldorf 等	构造基于支持向量构建核空间的方法, 可用于解决线性不可分问题
1987 年	Yang 等	将支持向量方法用于人体动作识别, 最高准确度达 96.05%
1995 年	Vapnik 等	提出“Chunking”的块算法, 正式命名为“支持向量机”, 并将其用于手写数字识别

1996年	Blanz等	将SVM用于物体识别
1996年	Schmidt等	将SVM用于说话人识别
1997年	Osuna等	提出分解算法，并将SVM用于人脸识别
1997年	Joachims	将SVM用于文本识别
1998年	Weston等	构造多值分类模型，可用SVM处理多元分类任务
1999年	Golub等	将SVM应用于基因表达微阵列数据分类

来源：《A tutorial on support vector machines for pattern recognition》(Burges等，1998)，《Molecular classification of cancer: class discovery and class prediction by gene expression monitoring》(Golub等，1999)，电子与信息学报公众号，声振之家公众号，国金证券研究所

由于支持向量机方法基本上不涉及概率测度及大数定律，具有更严格的理论和数学基础，因而拥有泛化能力强、收敛到全局最优、维数不敏感等优点。作为与神经网络并行的方法流派，在“深度学习”出现之前，支持向量机一度拥有更高的精度，是解决人工智能任务的主流方法。

支持向量机的最大瓶颈在于不适合进行多维度和大样本训练。其本质是借助二次规划获得最优解，当样本量过大时会极大增加机器内存和运算时间。

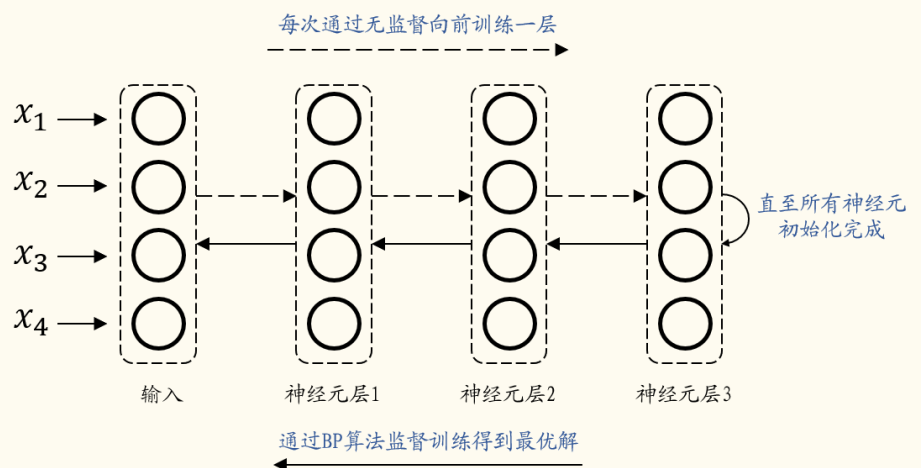
4. 梯度消失问题的解决使神经网络向深层迈进

4.1 逐层无监督是深度学习的初次探索

2006年，深度学习之父 Hinton 提出了一种梯度消失问题的解决方法，是深度学习的初次探索。

Hinton 的解决方法分为两步：1) 逐层无监督：即先通过无监督学习逐层初始各神经元的参数（而 BP 算法的初始参数为随机赋予），使各神经元的参数较为接近最优值，这一步极大地缓解了梯度消失问题；2) 通过 BP 算法进行有监督学习，得到神经网络的最优参数。Hinton 的论文发表之后，深度学习方法在学术界引起了巨大的反响，但由于此时模型的应用表现较为一般，因此深度学习的浪潮未向产业界蔓延。

图表 10: 逐层无监督+BP 有监督解决梯度消失问题



来源：《A Fast Learning Algorithm for Deep Belief Nets》(Hinton等，2006)，国金证券研究所

4.2 ReLu 激活函数的提出开启了深度学习时代

■ 2011年，ReLu 激活函数被提出，该激活函数的优点为：

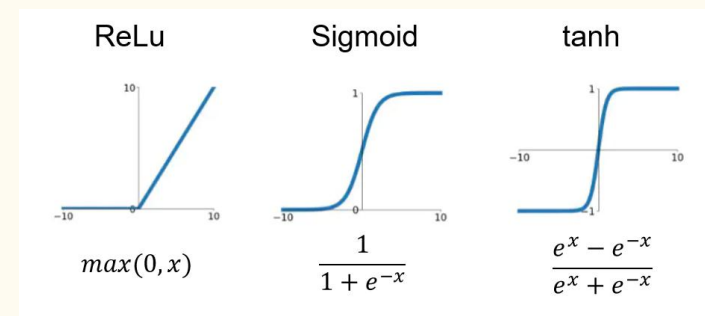
- 1) 具有更好的非线性拟合能力，缓解梯度消失问题；
- 2) 具有稀疏激活性，去掉了无关的噪声，能够更好地挖掘相关特征、且提升训练速度（导数为 0 或 1，非常简单）。

ReLu 激活函数被提出后，2011年微软首次将深度学习应用于语音识别，使得

识别词错误率稳定降低了 20~30%，在产业界引起轰动。

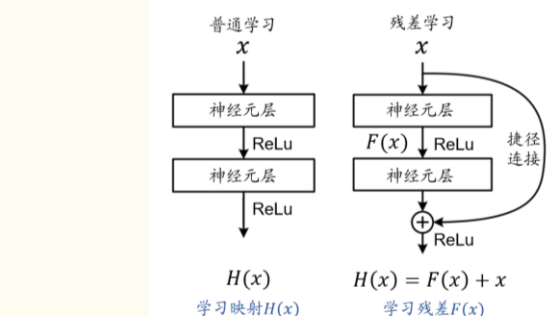
- 2012 年，Hinton 及其学生提出 AlexNet，自此奠定了深度学习的经典训练范式。AlexNet 采用了经典的 CNN 网络结构、使用 ReLu 激活函数、对输入值进行有监督学习（摒弃了此前的逐层无监督方法）、并采用 GPU 对训练进行加速。由于 AlexNet 将 ImageNet 数据集上图像分类的错误率由 26% 降至 15%，此后 5 年学术界均沿用 AlexNet 的范式进行深度学习训练。同时，由于错误率大幅降低，产业界开启了深度学习的应用。

图表 11: Sigmoid 和 tanh 函数存在梯度过大/过小问题



来源: PaddleEdu, 国金证券研究所

图表 12: 残差学习进一步缓解了梯度消失问题



来源: 《Deep Residual Learning for Image Recognition》(Kaiming He 等, 2016), 国金证券研究所

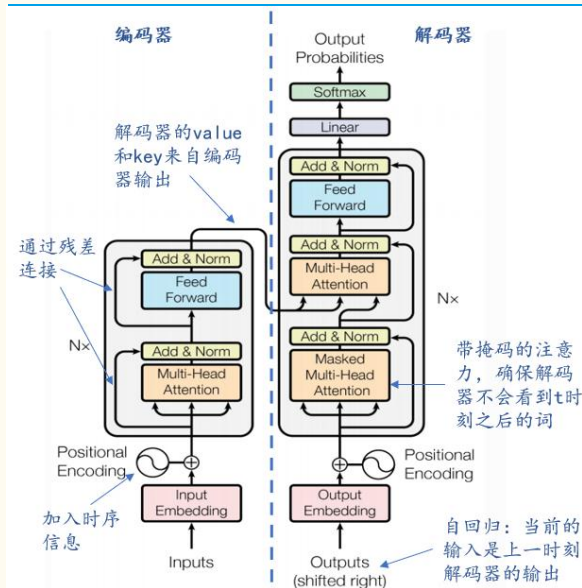
- 2016 年，何恺明等人提出 ResNet，使得神经网络可以对残差进行学习，更好地缓解了梯度消失问题，也让 ResNet 的神经网络深度高达 152 层。由于残差学习能很好地缓解梯度消失问题，在之后的 Transformer 架构中也得以沿用。

5. Transformer 统一了各任务底层算法，开启多模态和大模型时代

2017 年，Google 将注意力机制引入神经网络，提出了新一代深度学习底层算法——Transformer。由于 Transformer 在物体分类、语义理解等多项任务准确率超过 CNN、RNN 等传统算法，且能应用于 CV、NLP 等多个模态，因而 Transformer 的提出使得多任务、多模态的底层算法得到统一。

- 与 CNN、RNN 相比，3 种模型的算法复杂程度差异不大。Transformer 的主要优势在于：
 - 1) 可并行计算、减少训练时间、便于处理大数据：Transformer 用位置编码的方式代替了 RNN 的词句逐个输入、循环迭代，因此可以采用并行计算加速训练。
 - 2) 加强了神经网络对时序特征的提取：无论两个词相隔多远，都能捕捉到彼此之间的依赖关系。

图表 13: Transformer 示意图



来源: 《Attention Is All You Need》(Ashish Vaswani 等, 2017), 国金证券研究所

- 由于在处理大数据方面具备显著优势, Transformer 的出现开启了深度学习的“大模型”时代。如 OpenAI 发布的 GPT-3, 该模型在多个任务中都取得了第一的成绩, 而其在算法结构上与前代的 GPT、GPT-2 没有明显差别, 最大的区别是采用海量数据进行模型训练, 生动地诠释了“大就是好”的理念。

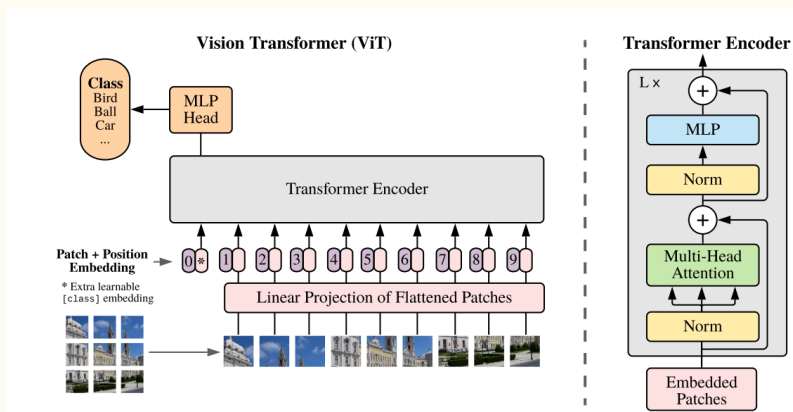
图表 14: GPT 系列模型参数呈指数级增长

模型名称	发布时间	参数量/亿	训练数据大小/GB	算力消耗/pfs-day
GPT	2018 年 6 月	1.17	5	0.96
GPT-2	2019 年 2 月	15	40	7.84
GPT-3	2020 年 5 月	1,750	45,000	3,640

来源: 《On the Comparability of Pre-trained Language Models》(Matthias 等, 2020), 国金证券研究所

- 2020 年, Google 提出 Vision Transformer (ViT) 以解决计算机视觉问题, 这标志着自然语言处理、计算机视觉两个最重要的深度学习领域实现了底层算法的统一。ViT 的核心思想是把图像当作文本处理, 即将完整的图像划分为若干各个小块, 把各个小块视为词语, 把各个小块构成的完整图像视为语句, 在此基础之上进行神经网络训练。

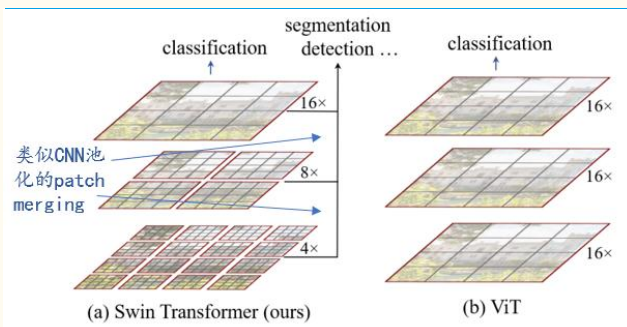
图表 15: ViT 模型将图像当作文本进行处理



来源: 《An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale》(Dosovitskiy 等, 2020), 国金证券研究所

- 相较 CNN, ViT 在计算机视觉领域任务处理的优点在于:
 - 1) 计算量降低: 注意力机制的引入使得神经网络可以将有限算力应用于关键信息, 降低了计算量、提升了神经网络的性能。
 - 2) 关注全局信息: ViT 不仅能对局部特征进行提取, 还关注各个小块之间的空间依赖关系, 能够依靠多个局部特征之间的空间关系辅助推理, 使得神经网络智能程度得到显著提升。
- 相较 CNN, ViT 在计算机视觉领域任务处理的缺点在于:
 - 1) 需要更大的训练样本: Transformer 相较 CNN 缺少一定的平移不变性和局部感知性, 因而需要更大的训练样本获得对局部特征的学习。在 ImageNet-21k、JFT-300M 等大数据集之中, ViT 较 ResNet 准确性更高。
- 2021 年 Swin Transformer 的提出很好地弥合了上述问题, 成为目前通用的骨干网络。与 ViT 始终进行 16×16 patch 的全局建模不同, 每层 Swin Transformer 模块之间进行了类似于 CNN 池化层的 Patch Merging 降采样操作, 用于缩小分辨率, 有效地降低了模型的计算复杂度。目前 Swin Transformer 在图像分类、语义分割、目标识别等领域均刷新了最优记录。

图表 16: 每层 Swin Transformer 之间进行了类似 CNN 的 patch merging



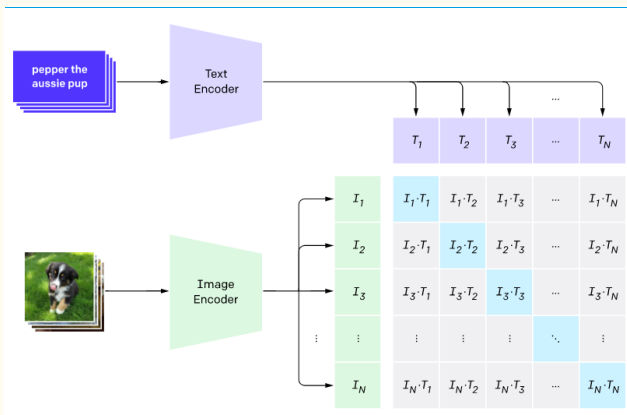
来源: 《Swin Transformer: Hierarchical Vision Transformer using Shifted Windows》(Ze Liu 等, 2021), 国金证券研究所

6. 探索多模态主干网络

ViT 开启多模态时代之后, 学术界开始尝试使用图像以及其对应的文本信息对神经网络进行训练, 以增强神经网络的理解能力。

- 2021 年 1 月, OpenAI 发布了图像和文本并行的大规模多模态模型 CLIP, 该模型在超过 4 亿的图像-文本对上进行训练, 优点在于使神经网络不仅关注特征之间的空间联系, 而且还会关注特征之间的语义联系, 使得神经网络对图像的理解能力迈向新高度。

图表 17: CLIP 多模态网络采用图像-文本联合训练



来源: OpenAI, 国金证券研究所

- 现阶段的多模态神经网络主要面向计算机视觉领域，还无法完成机器翻译、阅读理解等经典自然语言处理任务。其主要目的是通过引入文本信息帮助神经网络更好地完成传统视觉任务，并且使神经网络能够处理图文检索、语义图像生成、语义图像编辑等跨模态任务，拓宽深度学习能够面向的下游领域。

图表 18: 目前主流的多模态主干网络

公司	发布时间	模型名称	多模态任务
OpenAI	2021 年 1 月	CLIP	图文检索、地理定位、视频动作识别
微软	2021 年 8 月	BET-3	视觉推理、视觉问答、图像描述、图文检索

来源: OpenAI, 微软, 国金证券研究所

7. 探索 Transformer 时代的分支网络

Transformer 统一了多模态、多任务的骨干网络，这也使得其分支网络数量相对较少，往往是“Transformer+其他现有算法”的形式，学术界算法创新速度放缓。

7.1 AIGC 领域: 多模态 Transformer+扩散模型

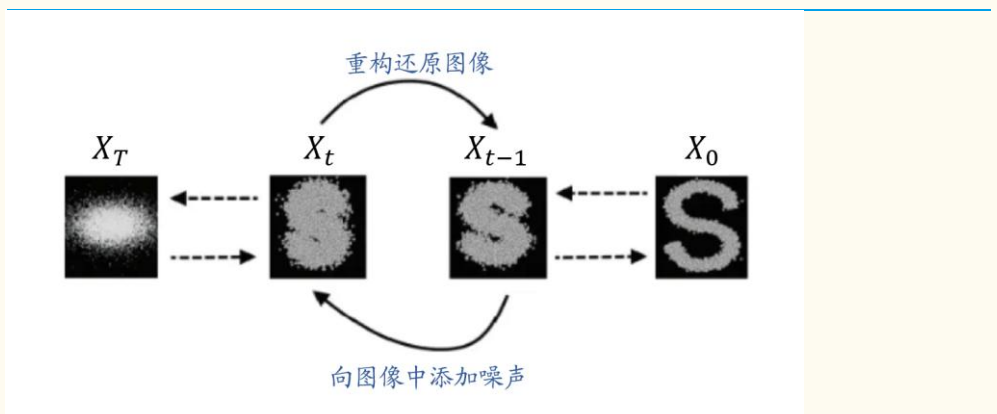
人工智能生成内容 (AIGC) 是指由人工智能进行创意创作，主要包括文学创作、代码生成、图像生成、视频生成等。其中文学创作和代码生成属于单一模态任务、难度较低，基于 Transformer 主干网络建立的大规模语言模型即可较好地适应这类任务。而图像生成、视频生成等任务难度较高，仅靠 Transformer 主干网络难以很好地完成任务，因而催生出一批由 Transformer 与其他现有算法结合而成的分支网络。

- 扩散模型是一种图像生成方法，其核心思想是污染与重构。扩散模型最早提出于 2015 年，2020 年 Ho 等人在原模型基础上进行了简化和建模方式改进，提出 DDPM 模型，目前 DDPM 是扩散模型应用的主流版本

扩散模型在训练时包括前向扩散和反向生成过程两个阶段。在前向扩散过程中，高斯噪音被逐步添加至图像中，直到图像成为完全随机噪声；在反向生成过程中，模型在每一个时间节点对噪声的概率分布进行预测，并尝试重构图像。

相较 GAN 而言，扩散模型训练更加稳定，且能生成更加多样的样本；缺点是训练和推理速度较慢。

图表 19: 扩散模型的前向扩散和反向生成过程



来源: TowardsData Science, 国金证券研究所

- 2022 年 4 月，OpenAI 基于多模态主干网络 CLIP、结合扩散模型方法，训练得到了新一代图像生成模型 DALL·E 2。该模型能够完成语义图像生成、语义图像编辑、图像风格迁移等多种任务，而用户无需任何计算机知识，仅需输入简短文字即可完成图像生成。

图表 20: DDPM 在 AIGC 领域的前沿成果

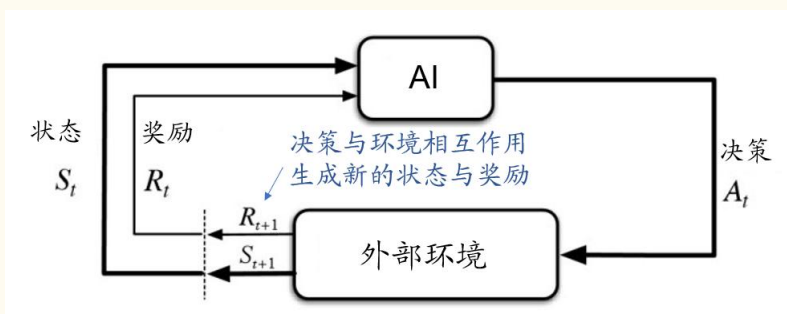
公司	发布时间	模型名称	多模态任务
OpenAI	2022 年 4 月	DALL-E 2	语义图像生成
Google	2022 年 8 月	Prompt to Prompt	语义图像编辑
Meta	2022 年 9 月	Make a video	语义视频生成/编辑、图像视频生成

来源: OpenAI, Google, Meta, 国金证券研究所

7.2 对话机器人 ChatGPT: 语言 Transformer+强化学习模型

强化学习最早出现于 20 世纪 60 年代, 核心思想是奖励机制, 使模型能够根据所处环境做出最优决策。Google 研发的著名围棋人工智能 AlphaGo Zero 便是基于强化学习方法训练得到的。

图表 21: 强化学习模型示意



来源: 清华大数据软件团队公众号, 国金证券研究所

2022 年 11 月 30 日, OpenAI 对外发布新一代对话机器人 ChatGPT, 这是一个基于 InstructGPT 模型, 能通过与人类互动进行强化学习的语言网络。相较于 GPT-3, InstructGPT 的参数量更小, 但引入了人类反馈强化学习 (RLHF) 的方法。

RLHF 是强化学习的一种, 其核心思想是将人类提供的范例作为学习目标, 当神经网络输出的内容接近范例时给予神经网络奖励。RLHF 的另一优点是能够帮助模型后续迭代升级。随着 ChatGPT 免费开放给公众使用, 广大用户形成的对话资料在经过数据处理后, 将成为最好的训练数据, 助力模型智能水平进一步提升。

8. 小结及投资建议

深度学习的三要素包括算法、数据和算力, 本文主要对算法的演进过程进行了回顾, 认为过往神经网络的发展以 Relu 激活函数的提出为分水岭, 分为浅层神经网络和深度学习两个阶段。

- 在浅层神经网络阶段, 最重要的任务就是解决梯度不稳定的问题。在这个问题未被妥善解决之前, 神经网络应用性能不佳, 而属于非神经网络的支持向量机算法 (SVM) 是解决人工智能模式识别的主流方法。
- 2011 年 Relu 激活函数被提出、梯度消失问题被大幅缓解之后, 神经网络进入深度学习时代, 算法和应用的发展均突飞猛进。最初 CNN、RNN 等模型在不同的模态和任务中均各有擅长, 2017 年 Transformer 的提出让深度学习进入了大模型时代、2020 年 Vision Transformer 的提出让深度学习进入了多模态时代, 自此多模态和多任务底层算法被统一为 Transformer 架构。

目前深度学习算法主要是基于 Transformer 骨干网络进行分支网络的创新, 如引入扩散模型、强化学习等方法催生出 AIGC、ChatGPT 等应用热点。但随着 Transformer 基本完成底层算法统一之后, 整个行业底层算法发展速度放缓, 静待骨干网络的下一次突破。同时基于 Transformer 对大数据的需求, 催生了无监督学习、高算力芯片的发展。

图表 22: 神经网络各发展阶段经典模型梳理

时代	模型	时间	核心思想	优点	缺点	
浅层神经网络时代	感知机	1958 年	· 通过激活函数进行二元分类	· 神经网络发展的起点	· 无法处理多维线性不可分的数据	
	多层感知机	1980s	· 引入隐藏层及非线性激活函数, 处理多元分类任务	· 可拟合非线性函数, 处理多元分类任务	· 对算力要求高; · 所有输入彼此独立, 缺失时间和空间信息	
	RNN	1986 年	· 每个神经元既接受当前时刻的输入信息, 也接受上一时刻神经元的输出信息	· 解决了时序信息的丢失问题; · 计算量大大降低	· 准确率待提升, 尤其是图像处理表现不佳	
	CNN	LeNet	1998 年	· 通过多个卷积层对特征进行提取和压缩, 得到高层次特征	· 解决了空间信息的丢失问题; · 计算量大大降低	· 准确率待提升, 尤其是 NLP 表现不佳
		-	2006 年	· 用逐层无监督初始化+BP 算法有监督学习的方法解决梯度消失的问题	· 缓解梯度消失问题	· 实际应用表现一般
深度学习时代	AlexNet	2012 年	· CNN+ReLU 激活函数+输入值有监督学习+GPU 训练	· 梯度消失问题大大缓解, 准确率更高 · 训练速度提升	· NLP 表现不佳	
	Transformer	-	2017 年	· 引入注意力机制, 对信息赋予不同的权重 · 加强了对时序特征的提取	· 开启了大模型时代, 可并行处理大数据, 训练时间减少; · 加强了对时序特征的提取	· 缺少平移不变性和局部感知性, 需要更大的训练样本 · 不能处理图像识别问题
		ViT	2020 年	· 分割图像当作文本处理	· 开启了多模态时代; · 计算量降低	· 需要更大的训练样本
		DDPM	2020 年	· 前向逐步添加高斯噪声, 在反向生成过程中对噪声的概率分布进行预测	· 模型稳定、数据生成更加多样	· 训练和推理速度慢
		Instruct GPT	2021 年	· 与人类互动进行强化学习	· 相较 GPT-3 参数量更小; · 训练数据更真实, 输出接近人类思维	· 准确率待提升, 目前仍未通过图灵测试

来源:《Dive into Deep Learning》(Aston Zhang 等, 2021),《A tutorial on support vector machines for pattern recognition》(Burges 等, 1998),《A Fast Learning Algorithm for Deep Belief Nets》(Hinton 等, 2006),《Deep Residual Learning for Image Recognition》(Kaiming He 等, 2016),《Attention Is All You Need》(Ashish Vaswani 等, 2017),《An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale》(Dosovitskiy 等, 2020),《Swin Transformer: Hierarchical Vision Transformer using Shifted Windows》(Ze Liu 等, 2021), DeepAI, PaddleEdu, 机器之心公众号, 新智元公众号, 深度人工智能公众号, AI 观察室公众号, AI 有道公众号, 易智瑞公众号, 腾讯云开发者公众号, 国金证券研究所

投资建议: 建议关注受益于人工智能算法进步, 并能成功实现商业化应用的海康威视、大华股份、科大讯飞、中科创达、商汤科技、四维图新、虹软科技等公司。以及关注可提供大算力 AI 芯片的海光信息、寒武纪等公司。

9. 风险提示

■ 海外基础软硬件使用受限

若因国际关系等原因, 高算力 GPU 等基础硬件或计算框架等基础软件使用受限, 可能会对国内人工智能算法应用产生影响。

■ 骨干网络创新放缓

目前 Transformer 成为深度学习骨干网络, 算法创新基本是基于 Transformer 做分支网络创新, 整体创新放缓。且 Transformer 本身作为骨干网络, 在处理部分任务时有一定局限性; 若骨干网络创新放缓, 可能部分任务解决进程会放缓。

■ 应用落地不及预期

若相关应用公司不能找到人工智能算法较好的商业应用落地场景, 或相关

场景客户没有较强的付费意愿，可能算法应用落地会不及预期。

公司投资评级的说明：

买入：预期未来 6-12 个月内上涨幅度在 15%以上；
增持：预期未来 6-12 个月内上涨幅度在 5%-15%；
中性：预期未来 6-12 个月内变动幅度在 -5%-5%；
减持：预期未来 6-12 个月内下跌幅度在 5%以上。

行业投资评级的说明：

买入：预期未来 3-6 个月内该行业上涨幅度超过大盘在 15%以上；
增持：预期未来 3-6 个月内该行业上涨幅度超过大盘在 5%-15%；
中性：预期未来 3-6 个月内该行业变动幅度相对大盘在 -5%-5%；
减持：预期未来 3-6 个月内该行业下跌幅度超过大盘在 5%以上。

特别声明:

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于C3级（含C3级）的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海

电话：021-60753903

传真：021-61038200

邮箱：researchsh@gjzq.com.cn

邮编：201204

地址：上海浦东新区芳甸路1088号

紫竹国际大厦7楼

北京

电话：010-66216979

传真：010-66216793

邮箱：researchbj@gjzq.com.cn

邮编：100053

地址：中国北京西城区长椿街3号4层

深圳

电话：0755-83831378

传真：0755-83830558

邮箱：researchsz@gjzq.com.cn

邮编：518000

地址：中国深圳市福田区中心四路1-1号

嘉里建设广场T3-2402