

汽车行业点评报告

AI 系列深度 05 期：模型如何学习表征并实现对齐？

增持（维持）

2026 年 07 月 29 日

证券分析师 黄细里

执业证书：S0600520010001

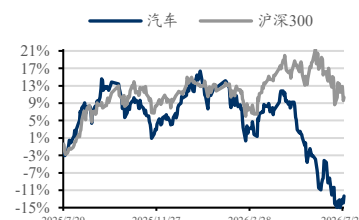
021-60199793

huangxl@dwzq.com.cn

投资要点

- **表征如何组织、如何学习、如何跨模态共享？** 嵌入、自监督学习和 CLIP 分别对应三层机制：嵌入是表征的工程形态，自监督是表征的规模化学习机制，CLIP 和多模态对齐则让图像与文字进入同一语义空间。本篇位于“表征是什么”与 CNN/ViT、Transformer 等具体架构之间，是理解后续技术路线的基础环节。
- **嵌入的本质，是把文本、图像、商品、用户等对象映射为数字向量，并使语义相近的对象在向量空间中距离更近。** 由此，搜索、推荐、以图搜图、向量数据库和 RAG 等应用，都可被理解为在高维向量空间中进行近邻检索。嵌入是抽象表征最可操作、最工程化的呈现方式。
- **自监督学习是通用表征规模化形成的关键机制。** 该方法不依赖人工逐条标注，而是利用数据自身构造训练信号，通过掩码预测、对比学习、下一个 token 预测等任务学习上下文、语义和结构关系。BERT、MAE、GPT 等路线任务形式不同，但均通过大规模自监督训练形成可迁移表征，这是预训练获得通用能力的重要基础。
- **CLIP 的意义在于把图像和文字对齐到同一表征空间。** 通过海量图文配对和对比学习，模型将正确配对的图文向量拉近、错误配对推远，使文字概念与对应图像在向量空间中自然接近。该机制成为图文检索、开放词表识别、文生图、视觉语言模型和多模态大模型的重要基础。
- **从表征体系看，本篇补齐三类关键问题：** 嵌入回答表征如何被工程化表达，自监督回答表征如何在无标注或弱标注数据中形成，CLIP 回答不同模态表征如何进入同一空间。
- **我们认为，数字 AI 底座模型发展路径已逐渐清晰、价值在于确定性，但其难以实现物理世界的建模闭环；我们认为物理 AI 将成为物理世界的 AI 解决方案，从而在 AI 的当前发展阶段成为增量更大的方向，智能驾驶作为最先跑通闭环的代表场景，我们认为应重点关注。**
- **风险提示：** 技术迭代不及预期的风险，大模型厂商 ARR 增速放缓的风险，云服务商资本开支不及预期的风险，智能驾驶及机器人商业化落地不及预期的风险。

行业走势



相关研究

《AI 系列深度 04 期：什么是表征？》

2026-07-29

《AI 系列深度 03 期：数字 AI 与物理 AI 的边界在哪里？》

2026-07-28

内容目录

1. 嵌入、自监督与 CLIP：表征落地的三类基础机制.....3

1.1. 三类机制分工：工程形态、学习机制与跨模态对齐.....3

2. 嵌入 Embedding：表征的工程化形态.....3

2.1. 嵌入的本质：将对象映射为向量.....3

2.2. 嵌入的落地：检索、向量数据库与 RAG.....3

3. 自监督学习：通用表征的规模化学习机制.....4

3.1. 自监督的必要性：标注成本高，未标注数据规模大.....4

3.2. 三类主流自监督任务.....4

3.3. 自监督是预训练的核心机制.....4

4. CLIP 与多模态对齐：让图文共享表征空间.....4

4.1. 跨模态对齐：图像与文字初始处于不同空间.....4

4.2. CLIP：基于图文配对的对比学习.....5

4.3. CLIP 成为多模态应用的重要基础.....5

5. 风险提示.....5

1. 嵌入、自监督与 CLIP：表征落地的三类基础机制

表征要进入工程系统，需要回答三个问题：1) 如何把对象表示为可计算结构；2) 如何在不依赖大规模人工标注的情况下学习表征；3) 如何让图像、文字等不同模态共享语义空间。嵌入、自监督学习和 CLIP 分别对应上述三类问题，是从“表征是什么”走向具体架构和多模态模型的基础环节。

1.1. 三类机制分工：工程形态、学习机制与跨模态对齐

嵌入、自监督学习和 CLIP 构成表征落地的递进链条。嵌入回答表征如何以向量形式组织和调用；自监督回答表征如何从大规模无标注数据中学习；CLIP 回答图像、文字等不同模态如何对齐到同一语义空间。三者共同将抽象表征转化为可训练、可检索、可跨模态调用的工程体系。

2. 嵌入 Embedding：表征的工程化形态

2.1. 嵌入的本质：将对象映射为向量

嵌入是将词、图像、商品、用户等对象映射为固定长度数字向量的过程。向量化之后，不同对象之间可以通过距离、方向和相似度进行计算，从而把语义关系转化为可操作的几何关系。对模型而言，嵌入是对象进入计算系统的基础接口，也是表征最常见的工程形态。

高质量嵌入的核心特征，是语义相近的对象在向量空间中距离更近。以文本为例，“猫”和“狗”通常比“猫”和“汽车”更接近；在推荐和检索场景中，用户兴趣、商品属性和文档主题也可以通过向量距离表达。由此，语义相似度判断被转化为向量空间中的近邻搜索问题。

嵌入空间还可以保留部分语义方向和结构关系。经典词向量研究发现，“国王-男人+女人 $\approx$ 王后”等向量运算能够反映性别、身份等语义方向，即“国王-男人”可近似于“王位”的向量表示，该向量表示加上“女人”则可近似表示为“王后”，说明嵌入并非简单编号，而是在连续空间中编码对象之间的关系。需要注意，这类线性关系并非对所有概念普遍成立，但揭示了嵌入作为语义几何结构的价值。

2.2. 嵌入的落地：检索、向量数据库与 RAG

嵌入是向量检索、推荐系统和语义搜索的底层接口。工程流程通常是先将文档、图像、商品或用户转化为向量并存入向量数据库，再将查询请求转化为向量，在库中寻找距离最近的候选项。语义搜索、以图搜图和推荐系统，本质上都是在向量空间中执行近邻检索。

RAG 同样建立在嵌入和向量检索之上。企业文档先被切分并嵌入向量库，用户提问时系统先检索相关片段，再将其放入提示词中交给大模型生成答案。该范式解决了大模型难以实时记忆最新或私有知识的问题，是当前企业级 AI 落地的重要路径之一。

3. 自监督学习：通用表征的规模化学习机制

3.1. 自监督的必要性：标注成本高，未标注数据规模大

传统监督学习依赖人工标注，但人工标注成本高、速度慢，难以覆盖互联网级图文数据。自监督学习的核心，是利用数据自身构造训练信号，使模型在无需人工逐条标注的情况下学习结构、语义和上下文关系。这一机制使海量无标注数据可被纳入预训练，是大模型形成通用能力的重要前提。

3.2. 三类主流自监督任务

第一类是掩码预测。模型遮盖输入的一部分，再根据上下文恢复被遮盖内容；文本侧代表为 BERT，图像侧代表为 MAE。为了完成恢复任务，模型需要学习上下文依赖、局部结构和整体语义，由此形成可迁移表征。

第二类是对比学习。模型将同一对象的不同增强视为正样本，将其他对象视为负样本，通过拉近正样本、推远负样本来学习表征。SimCLR 是典型代表，LeCun 的 JEPA 也可归入在表征空间中进行预测和比较的路线。对比学习的关键在于不直接依赖类别标签，而是通过样本间关系构造学习信号。

第三类是下一个 token 预测。模型根据已有上下文不断预测下一个 token，训练答案来自文本序列自身，无需额外人工标注。GPT 系列正是通过该任务在海量文本上学习语言表征，并进一步形成生成和推理能力。

3.3. 自监督是预训练的核心机制

自监督学习构成预训练的内核：先在海量数据上通过自监督任务学习通用表征，再在少量高质量数据上进行微调、指令对齐或后训练。嵌入是表征的工程化结果，自监督则是形成该结果的训练机制。今天大模型的通用能力，很大程度来自自监督预训练所获得的可迁移表征。

4. CLIP 与多模态对齐：让图文共享表征空间

4.1. 跨模态对齐：图像与文字初始处于不同空间

图像表征通常由视觉网络提取，文字表征通常由语言网络提取，二者初始处于不同向量空间。模型并不会天然知道一张猫的照片与“猫”这一文字概念对应。要实现文生图、

以文搜图和视觉语言模型，首先需要把图像与文字对齐到可比较的共享语义空间。

4.2. CLIP: 基于图文配对的对比学习

CLIP 由 OpenAI 在 2021 年提出，核心思路是利用互联网中的大规模图文配对数据进行对比学习。训练过程中，模型将正确配对的图像和文本向量拉近，将不匹配的图文向量推远；训练完成后，图像和文字被映射到同一向量空间，对应概念在空间中更接近。

CLIP 的重要能力之一是零样本识别。对齐完成后，模型可将候选类别名称转化为文字向量，再与输入图像向量比较相似度，从而判断图像所属类别，而不必为每个新类别重新训练分类器。这种以文字作为开放标签的能力，是多模态模型泛化的重要来源。

4.3. CLIP 成为多模态应用的重要基础

CLIP 式图文对齐成为文生图、图文检索、视觉语言模型和多模态大模型的重要底座。Stable Diffusion 等文生图模型利用文字向量引导图像生成；VLM 则通过图文对齐获得看图、读字和语义推理的基础能力。能否把多模态信息映射到统一表征空间，直接影响后续多模态模型的泛化能力和应用边界。

5. 风险提示

技术迭代不及预期的风险：数字 AI、物理 AI 技术进展低于预期，任务或应用场景拓宽速度低于预期。

大模型厂商 ARR 增速放缓的风险：若客户增长、续费率或客单价不及预期，模型厂商 ARR 增速可能放缓。

云服务商资本开支不及预期的风险：若自由现金流承压，或 AI 算力需求及投资回报低于预期，云服务商可能下调资本开支。

智能驾驶及机器人商业化落地不及预期的风险：若产品可靠性、成本下降或用户需求低于预期，相关产品商业化进程可能放缓。

免责声明

东吴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。

本研究报告仅供东吴证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

在法律许可的情况下，东吴证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

市场有风险，投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

本报告的版权归本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的，应当注明出处为东吴证券研究所，并注明本报告发布人和发布日期，提示使用本报告的风险，且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数，新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的），北交所基准指数为北证 50 指数），具体如下：

公司投资评级：

买入：预期未来 6 个月个股涨跌幅相对基准在 15%以上；

增持：预期未来 6 个月个股涨跌幅相对基准介于 5%与 15%之间；

中性：预期未来 6 个月个股涨跌幅相对基准介于-5%与 5%之间；

减持：预期未来 6 个月个股涨跌幅相对基准介于-15%与-5%之间；

卖出：预期未来 6 个月个股涨跌幅相对基准在-15%以下。

行业投资评级：

增持：预期未来 6 个月内，行业指数相对强于基准 5%以上；

中性：预期未来 6 个月内，行业指数相对基准-5%与 5%；

减持：预期未来 6 个月内，行业指数相对弱于基准 5%以上。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况，如具体投资目的、财务状况以及特定需求等，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

东吴证券研究所
苏州工业园区星阳街 5 号

邮政编码：215021

传真：（0512）62938527

公司网址：<http://www.dwzq.com.cn>