

**研 究 生 课 程 论 文**

****

|  |  |
| --- | --- |
| **题目：** | **数据驱动的角色动作生成技术** |
| **学号：** | **22251246** |
| **姓名：** | **马云涛** |
| **专业：** | **电子信息** |
| **所在学院：** | **软件学院** |
| **课程名称：** | **计算机动画与交互技术** |

**一、动作生成技术研究现状及进展**

使用图形学方法对角色动作生成问题进行的研究大致分为两个方向：基于物理的角色控制与基于数据的角色控制。前者主要是通过模拟一个真实的物理环境，设置人在运动过程中的约束，比如关节之间的旋转范围，人运动的平衡让“人”在这个环境里去尝试实现一些任务，在这个过程中人会摔倒，或者实现了目标设置奖惩函数，通过强化学习的方法，最终让人学会去完成任务。其中最有名的工作就是发表在SIGGRAPH 2018上的DeepMimic。

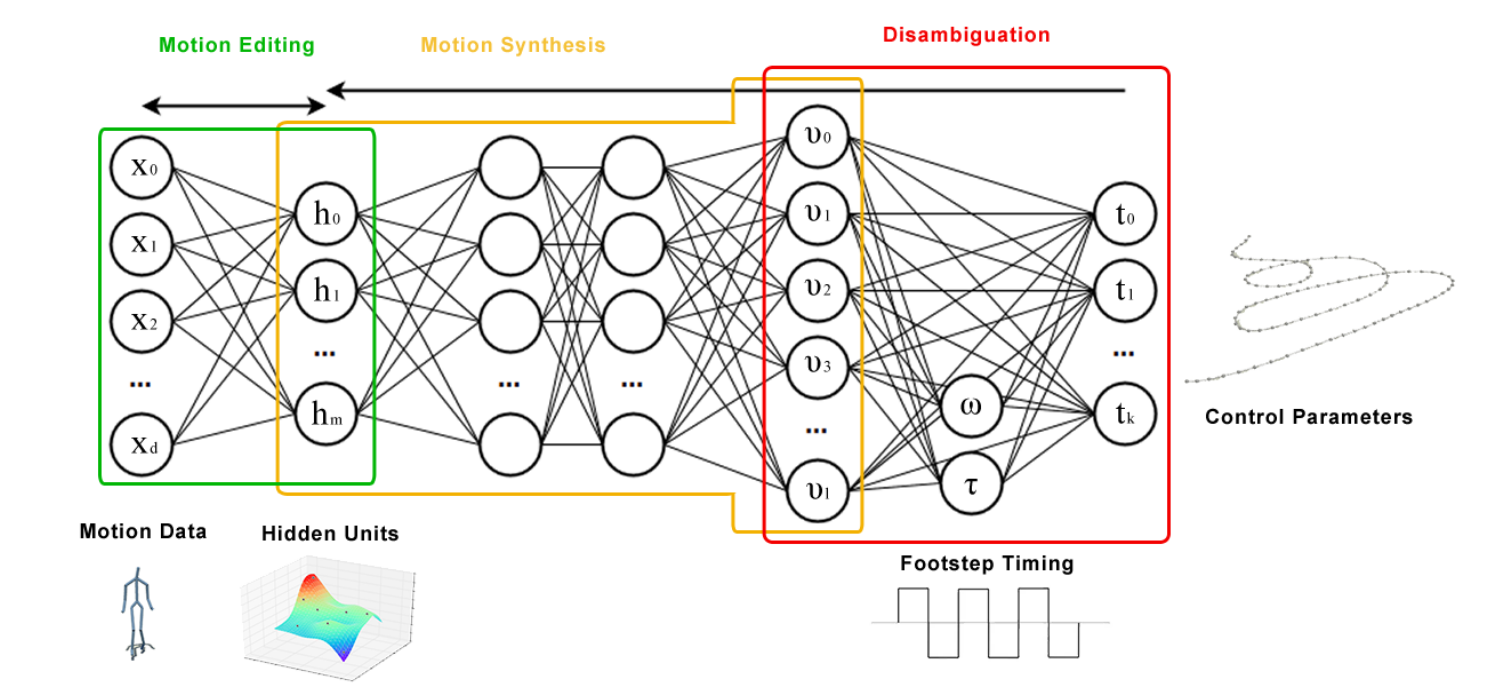
但是用强化学习来解决角色动画生成的问题不是一件容易的事情，其中的问题主要体现在当前强化学习训练较为困难，难以处理动作控制这类复杂任务。相比较来说，另一类基于数据的角色控制就更加直观。为了生成动作首先对动作进行学习，一般分为三步：通过动作捕捉的方法，获取大量人运动的数据；用神经网络学习动作的表达；根据用户指定的路线或者方向，生成动作。本文主要介绍基于数据驱动的角色动作生成方法。

**二、SIGGRAPH 2016《A Deep Learning Framework for Character Motion Synthesis and Editing》**



本篇文章提出了一种可以使用高阶参数来指导动画的生成与编辑的深度学习框架，解决了之前的工作中存在需要对动作数据进行人工标注并且运行时鲁棒性不好的问题。

作者使用了他们在SIGGRAPH Asia 2015中的前置工作，其提出了一种卷积自动编码器将人体动作降维到低维的隐空间上，为了不让解码器出来的动作出现数据损失的情况，隐空间的维数并不低，但相比较原空间来说，在隐空间形成的流形更适合用来生成动作。本文利用了上述自动编码器的解码部分。也就是说，所有生成的动作都是对应隐空间上的一个结果，再由解码器解码得到最终的结果，这就要求隐空间要足够通用。文中作者使用了一个大规模数据集对其进行预训练得到。



本文提出的框架分为两个部分，动作合成部分和动作编辑部分。动作合成部分，最直观的想法，输入就是控制信号，输出就是动作，但这样会由于收集到的动作数据有歧义而出现问题，例如在同一条轨迹下，可以对应有两个完全不同频率的运动，一个快一个慢，那么在训练的时候，同样的输入，就会对应不同的输出，从而网络学到的结果就是它们的平均，也就会出现太空步的效果。

因此作者在在动作合成部分前添加了用于消除歧义的网络，这部分通过识别数据中人的脚离地的距离来判断当前人的动作状态，然后计算出人运动的频率和时长。这个消除歧义的网络可以理解为通过增加维度来使任务变得更容易学习。

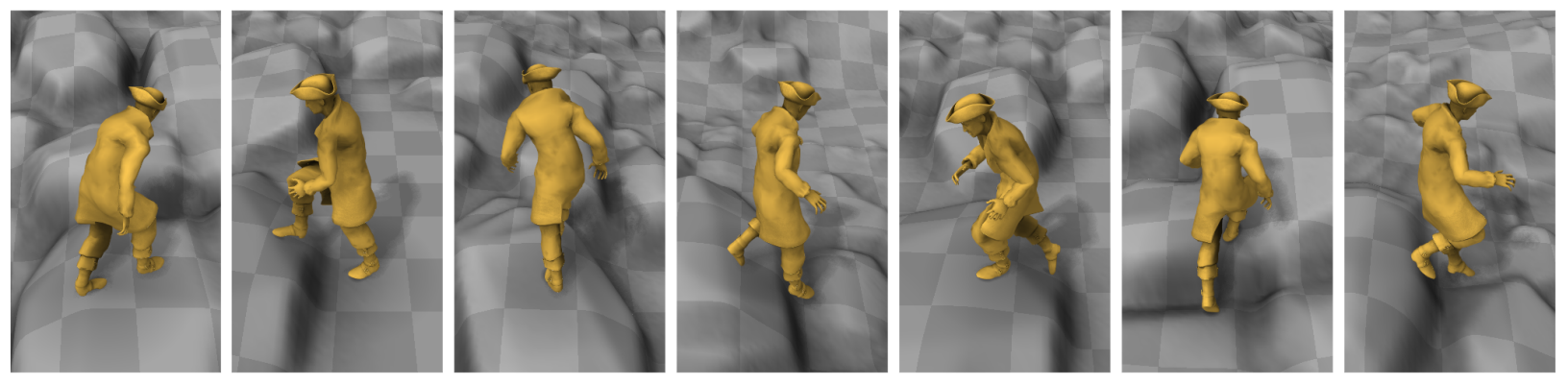
动作编辑部分的作用主要存在于生成的东西往往不是完美的控制，必须要通过进一步的编辑才能得到完美的结果。

直接在结果上编辑，并不是一件容易的事情，那样的话即使生成的结果很好，经过编辑可能又会出现不自然的结果，因此，本文将把该问题放到隐空间上来解决。隐空间保证了生成的动作是自然的，将编辑看作三维空间中的目标优化问题，将该问题转化为隐空间上的优化问题，就能保证出来的结果依然保持自然的效果，不会出现人工编辑带有的缺陷。

并且该部分可以用来生成风格动作，只要在优化项中加入动作风格的约束即可得到较好的结果。

本文并没有使用端到端的框架，而是将大问题拆分为几个子问题，然后分别进行解决，最后将问题的解决方案进行融合，得到了一个具有较好效果的深度学习框架。

**二、SIGGRAPH 2017《Phase-Functioned Neural Networks for Character Control》**



本文主要解决前文中提到的歧义问题，实现了角色动画的实时生成控制，通过当前状态以及用户指定的控制参数求取下一刻的角色姿态。

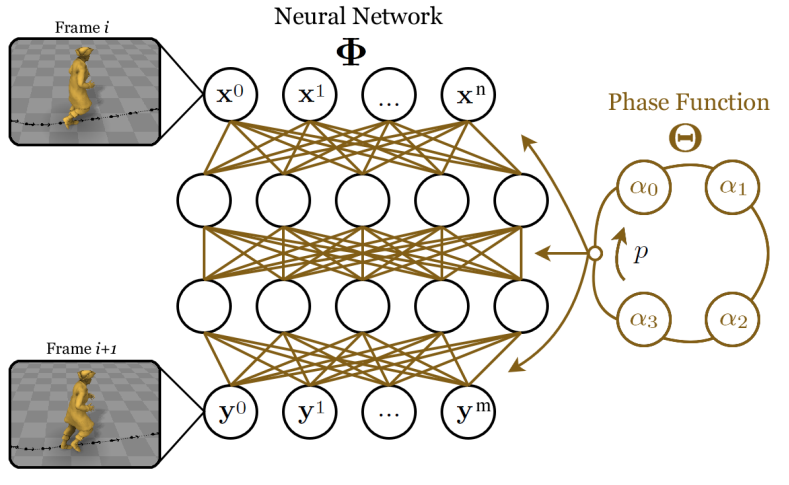
为了消除动作歧义，前文中作者增加了一个模块专门提取输入高阶动画的频率和时长信息。但是，这样的做法只适合离线的生成，并不适合实时的控制。在实时的控制下，这些参数是随时可变的。为了解决这个问题，一个思路是给路径添加约束，另一个思路则是从运动本身下手——之前求频率和时长就是基于人的运动有相位性的假设，若已知当前动作在相位的哪个状态，求下一个时间点的状态就不再困难，同时，歧义的问题也同样得到了解决，因为周期性增加了动作所处状态的约束信息。

本文的流程一共分为三个部分：预处理、训练与测试。

预处理部分中运动相位的信息在动作捕捉数据集中是没有的，需要标记。脚与地面的接触很好标记，但是这个经常出错，所以作者又人工检查了一遍。相位是个0到2π的值，当右脚触地，相位值为0，当左脚触地，相位值为π，右脚又触地，相位值为2π。站立的状态是没有周期的，所以人为定义了一个很短的时间作为周期。

作者同时通过人为标记动作的类别降低歧义，也方便用户设计想要的运动。

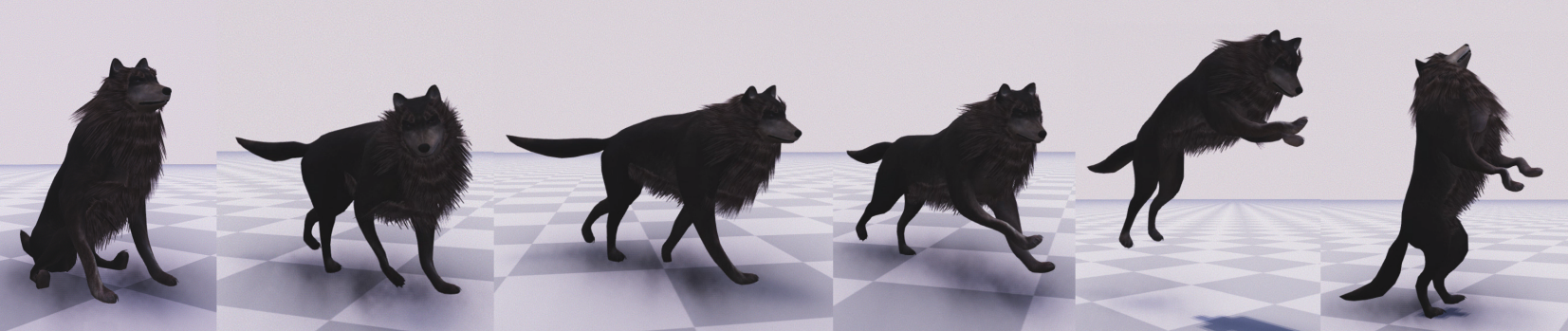
轨迹与地形拟合控制人的运动是通过给定人应该遵循的运动轨迹，数据中的轨迹通过投影得到，地形拟合则是由于在动作捕捉数据时，为了能获得爬山这种动作的数据，有录具有地面高度变化的数据，但是由于地面高度并没有在捕捉数据中，所以只能重新反求。地形拟合为了能训练对地形更加鲁棒的模型，需要对数据中的地形做增强。



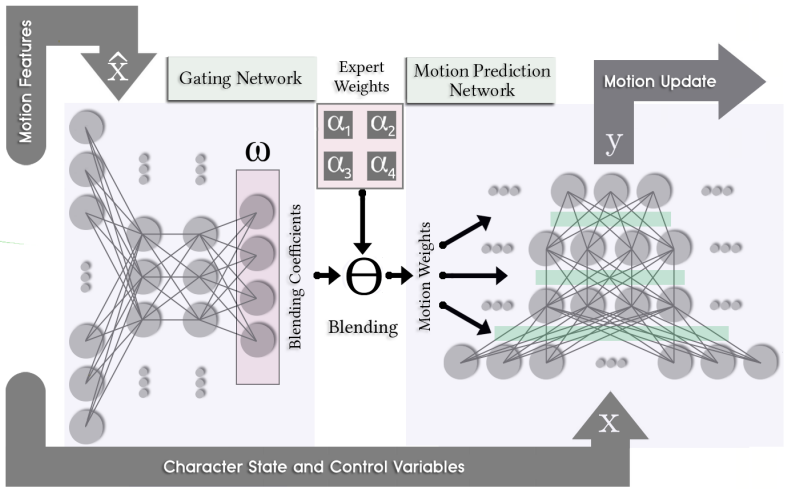
在训练与测试部分中，网络结构分为两部分，一部分是预测的网络，用于运动估计，另一部分则是相位网络，输出当前的相位，相位会进一步控制预测网络的权重。这里就有一个问题了，前面说到相位是[0-2π]的函数，如果按照0到2π循环，那么输出和预测网络做混合得到的就不是连续的变化。因此，对相位函数要采取一个可循环的输出，文章用的是cubic Catmull-Rom spline函数。

相比上一篇工作，PFNN不管是框架、效果还是实验都有了重大提升，而且真正有了可以实用的影子。相比之下，求phase的长时间问题也不是那么大。

**三、SIGGRAPH 2018《Mode-Adaptive Neural Networks for Quadruped Motion Control》**



本文将重点放在了四足动物的角色动画生成上。这部分工作比人的运动又难了一些，主要原因有两个：四足动物的数据难采集，四个脚运动肯定比两个脚要复杂。



PFNN用相位解决了人运动的歧义性问题，对于四足动物的动画相位情况比两个脚就要复杂度多了。本文架构如图所示，将之前人为定义的周期函数改为全连接网络，将网络的输出经过softmax变成一个和为一的向量，再用这个向量控制预测网络的权值。这种方法称为Mixture of Experts。

虽然本文引入MoE虽然只是一个小的改动，但是这是一种更加直观的动作身体问题的方法，一切都在往完全的端到端模型靠拢。

**四、SIGGRAPH Asia 2019《Neural State Machine for Character-Scene Interactions》**



环境交互一直是数据驱动方法难以处理的问题，因为环境的情况千变万化，要生成足量的数据分析环境是一件非常困难的事情。

在这篇文章中，作者将重点集中在了生成角色在有限个交互情况下的动作，如上图中的第一个弯腰过墙，第二个坐在桌子上，第三个搬东西，第四个坐在椅子上，第五个开门。

在本文中，在通过动作捕捉得到的人的行为之外，作者通过IK来反向计算动作来获取环境数据。以椅子为例，作者从ShapeNet中选取了大量各种各样的椅子，然后计算人与椅子交互时接触的关节点（手、臀部），再通过IK的方法重新计算动作的信息，这样，一方面不仅不需要采集环境的信息，还可以通过这种方式获得多种数据，这是文章的重要贡献之一。

本文中提出的模型依然是MoE的模型，但是为了和环境交互，增加了几个新的关键点：首先，输入中出了演员的姿态、轨迹，还包含了角色与环境相对的位置信息。输出中也增加了目标视角下的演员坐标、接触点以及相位的更新信息。作者在本文中重新启用了相位的信息，但与PFNN不同在于这次相位信息不是通过网络学习而是显性定义的，并且作用在输入而不是网络的权重上。

**五、总结**

动作匹配是一种强大且通用的动画系统实现方案。这种基于机器学习的方法让数据库可以扩展到巨大的数据集上。通过这种技术，可以避免使用状态机，不再需要大量的动画控制器，对人力成本是一种极大的节约。在提高生产力的同时让动作设计师能够有更多的精力释放创造力，创造出能够对不同情况做出真实而独特反应的角色。

近几年的顶会SIGGRAPH 都发表了动作生成方向的高质量论文，进一步解决了其在应用落地上的短板。这个领域取得了很多优秀的成果，但仍然处于起步阶段。真实性，实用性，普适性，行为理解，环境交互，留给后来人研究的问题还有许多许多。

**六、参考文献**

[1] Holden D, Saito J, Komura T. A deep learning framework for character motion synthesis and editing[J]. ACM Transactions on Graphics (TOG), 2016, 35(4): 1-11.

[2] Holden D, Komura T, Saito J. Phase-functioned neural networks for character control[J]. ACM Transactions on Graphics (TOG), 2017, 36(4): 1-13.

[3] Zhang He,Starke Sebastian,Komura Taku, Saito, Jun. Mode-Adaptive Neural Network for Quadruped Motion Control. SIGGRAPH 2018.

[4] Starke S, Zhang H, Komura T, et al. Neural state machine for character-scene interactions[J]. ACM Transactions on Graphics (TOG), 2019, 38(6): 1-14.

[5] Sebastian Starke, Yiwei Zhao, Taku Komura and Kazi Zaman. Local Motion Phases for Learning Multi-Contact Character Movements, ACM Transactions on Graphics 39(4) (Proceedings of SIGGRAPH 2020)