讲座 2：机器翻译和多语言大模型

机器翻译的起源：Warren Weaver, 1949：
分析自然语言——研究目标，发掘共同特性——特别关注

机器翻译的发展：

- **发展历程简介**
  - 基于规则的机器翻译
  - 基于实例的机器翻译
  - 统计机器翻译
  - 神经网络机器翻译
  - 大语言模型和多语言大模型

1. 基于规则的机器翻译：
   由语言学方面的专家进行规则的制定

2. 基于实例的机器翻译
   利用类比思想，查找接近的翻译实例，并进行逐词替换以完成翻译

3. 统计机器翻译
   从双语平行语料中自动进行翻译规则的学习和应用

- **可以一定程度上从数据中自动挖掘翻译知识**
- **流程相对复杂，其中各个部分都不断被改进和优化**
- **翻译性能遇到瓶颈，难以大幅度提升**

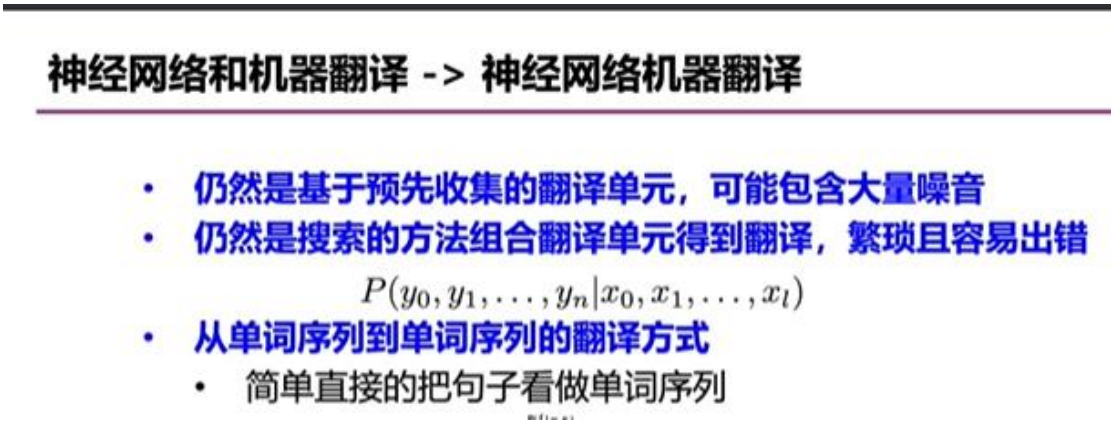
3. 神经网络机器翻译

神经网络语言模型（NLPM）：使用神经网络建模 n-gram 概率

- **类似分类，学习（x, y）的映射关系**
  - 假设映射关系是语言内通用的
    - 描述语言的是映射?
  - 相近的词可能得到相近的计算结果
    - 极大缓解了数据稀疏问题，不会因为数据出现少导致零概率
- **参数用于建模映射，即描述概率计算过程**
  - 参数规模 O(|V|*d)
    - 降低了参数规模

*** 将one-hot映射为低维的参数可以作为词的表示**

神经网络和机器翻译：符号表示到连续表示的变革降低了数据稀疏性
超越单个模型的建模方式：多个模型共同建模
尝试直接生成目标短语



神经网络机器翻译：

- **从单词序列到单词序列的翻译方式**
  - 简单直接的把句子看做单词序列
  - **不需要建模规则的组合关系**
  - 例如：习近平主持召开中央全面深化改革领导小组会议
- **机器翻译能力随着机器计算能力的迅速发展而增长**
  - 神经网络的引入从统计稀疏性和建模两个方面提升了机器翻译系统
  - 神经网络机器翻译是一种能够更加充分发挥机器长处的自动翻译方法
- **进一步的提升?**

4. 大语言模型和机器翻译：更大规模的数据+更大规模的模型

No Language Left Behind (NLLB):

- **单语数据+双语数据挖掘**
- **148含英语语言对-761m句对**
- **1465非英语语言对-302m句对**

- **1.3B dense/54.5B MoE**
- **Flores-200**

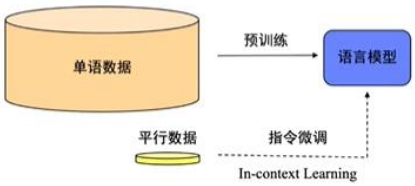
大模型的翻译能力表现：

- **通用模型 v.s. 执行特定任务（翻译）**
  - Instruction Following 通过指令指定模型行为
  - In-context Learning (ICL) 从上下文中学习

学习范式正在转变：

- **神经机器翻译: 主要从平行数据中学习翻译知识**

- **大语言模型: 主要从单语数据中学习通用知识（包含翻译知识）**



新的学习范式带来新的研究问题

- **当前状况评估**
  - 大语言模型翻译表现的好不好?
  - 不同的表现方式有什么区别?
- **探索模型潜力**
  - 大语言模型翻译潜力如何?
  - 如何能够激发模型的翻译能力?
- **模型能力的可能应用**

当前状况评估 → 探索模型潜力 → 模型能力应用

评估结果-v.s. 有监督模型

- **大语言模型仍然落后于强大有监督基线模型:**
  - GPT-4 (40.91%) v.s. NLLB (59.19%)
  - 尤其在低资源语言上远远落后于商用翻译系统 Google Translate

额外观察-能力激活问题

- **通过in-context exemplars激活翻译能力并不稳定**
  - 德英翻译使用其他语言激活效果不好
  - 中英翻译使用其他语言激活效果更好

研究思路

- **ICL可能无法有效激活模型的能力**
- **如何能够更准确地评估模型的翻译能力?**

- **显式要求模型完成翻译指令**
  - 将l_s语言的句子x翻译为l_t语言y

Translation: [l_s]: x [l_t]: y

主要结果

- **指令学习之后，模型确实展现了更强的翻译能力/潜力**

研究动机：大语言模型中包含大量的长尾信息；大语言模型参数规模巨大，解码开销高
传统知识迁移方法：

传统方案的问题：无外推能力

- **完全依赖语言现象在单语数据中的显式出现**
  - 无法外推到新的上下文
  - 无法外推到其他的词

MT-Patcher：基于LLM的高效知识迁移框架

- **识别小模型中的翻译错误，并进行修正（Feedback/Post Editing）**
  - 可以针对性地选择错误的知识
- **根据现有错误词对进行联想（Word Analogy）**
  - 可以外推到更多相关的知识
- **生成包含目标词对的平行数据（Parallel Data Synthesis）**
  - 可以外推到更多上下文

- **充分发挥大语言模型的能力，充分利用原有模型的现有能力！**

部分结果：中文成语翻译

- **MT-Patcher 显著提升了对于新的上下文/词的翻译准确率**
  - 中文成语翻译上也能观察到类似的提升。

小结：

- **长尾知识翻译仍然是机器翻译中一个尚待解决的问题**
  - 平行数据中覆盖较困难
  - 小模型中也可能不能完全覆盖
- **高效地从大模型向小模型中迁移长尾知识**
  - 利用大型语言模型的通用语言能力
  - 利用小模型的现有能力

多语言大模型：
能力来源？运作机制？可靠性问题？

偶然的双语信号Incidental Bilingualism

- **PaLM的数据中出现了涉及至少44种语言的双语信号**
  - 包括翻译、参考、字符混合等情况

自发形成的对应

- **当语言之间存在隐式的对应关系时，可以成为学习信号的来源**
  - Unsupervised Neural Machine Translation

模型的多语言能力如何工作?

- **可能是借助英文发挥作用（通过观察中间层状态/输出）**
- **可能是借助英文发挥作用（通过激活神经元情况进行推测）**
- **存在语言无关和语言相关的神经元**

模型中存在跟语言相关的特定区域

- **语言核心区：仅占1%左右参数，但是对输出语言有重要影响**
  - 这些区域的参数改变会影响模型在一种或多种语言上的能力

多语言的不平衡：

大语言模型的多语言能力——从知识来看

- **多语言大模型的跨语言能力不平衡**
  - 用英文和非英文回答相同问题，后者准确率更低 (SeaEval，Wang et al., 2023)
  - 英语知识编辑对其他语言上影响较小(Qi et al. 2023)
- **基础模型、微调策略的影响（Ye et al., 2023）**
  - 跨语言预训练的语言覆盖率较为重要
  - 指令微调时用目标语言比用英语略好

- **不同语言中学到的是相同的知识吗?**
  - 性能一致是否代表能力一致?
  - 预训练微调是否可以提升一致性?

系统性评估多语言知识一致性

- **准确率（Performance, PF）：**
  - 模型在不同语言上回答知识相关问题的准确率是否接近?
- **一致性（Consistency, CT）：**
  - 模型用不同语言回答相同的知识相关问题时，是否给出相同的答案?
- **传导性（Conductivity, CD）：**
  - 模型在一种语言上学习到的知识，能否用另一种语言检索出来?

- **已有文献关注 PF 和 CT，很少有更本质的对 CD 的研究**
- ***考虑到跨语言对齐的效果受到语言基础能力的影响，因此应额外评估基础能力**

LLM的跨语言知识对齐程度

- **目前看来LLM的知识迁移能力不高**
  - 用中文回答英语中学习到的知识

LLM的“新”知识推理

- **目前看来利用预训练知识进行推理的能力不足（Out-of-Context Reasoning）**
  - 可能跟the reversal curse (Berglund et al., 2023)有关

LLM的推理能力迁移

- **纯推理（knowledge-free reasoning）比知识推理容易跨语言迁移**
  - 具有更高的隐藏表示一致性和激活神经元重合比例

增强语言模型的特点语言能力：

- **模型的能力由数据决定**
  - 数据覆盖较少的语言能力较弱
- **增强指定语言的能力**
  - 增加该语言数据
  - 利用语言间的对应关系

LLM的语言能力不平衡

- **更强能力往往仅在英文上表现，在中低资源语言上能力明显不足**
  - 指令执行，多轮指令执行
  - 安全性

以泰文提问时，ChatGPT误解了要求，给出了检

对话大模型的跨语言能力迁移框架TransLLM

提升解决特定任务的能力：Q-Align

- 让LLM像理解英语一样去理解其他语言的问题，然后利用其英语能力对问题进行求解

提升解决特定任务的能力：MAPO

- 利用优势语言推理过程对其他语言的推理过程进行偏好优化
- Multilingual-Alignment as Preference Optimization(MAPO)

仍然有进一步提升的空间

- 各语言通过与英文对齐取得了显著提升
- 但仍没能完全达到英文的水平

小结：

- 大模型具有令人惊奇的多语言能力
  - 其原理和工作机制正在被逐步揭示
- 语言间不平衡影响可靠性和公平性
  - 减小语言间差异、提升特定语言性能尤为重要
  - 翻译发挥着重要的桥梁作用