



CCAI 2017  
中国人工智能大会

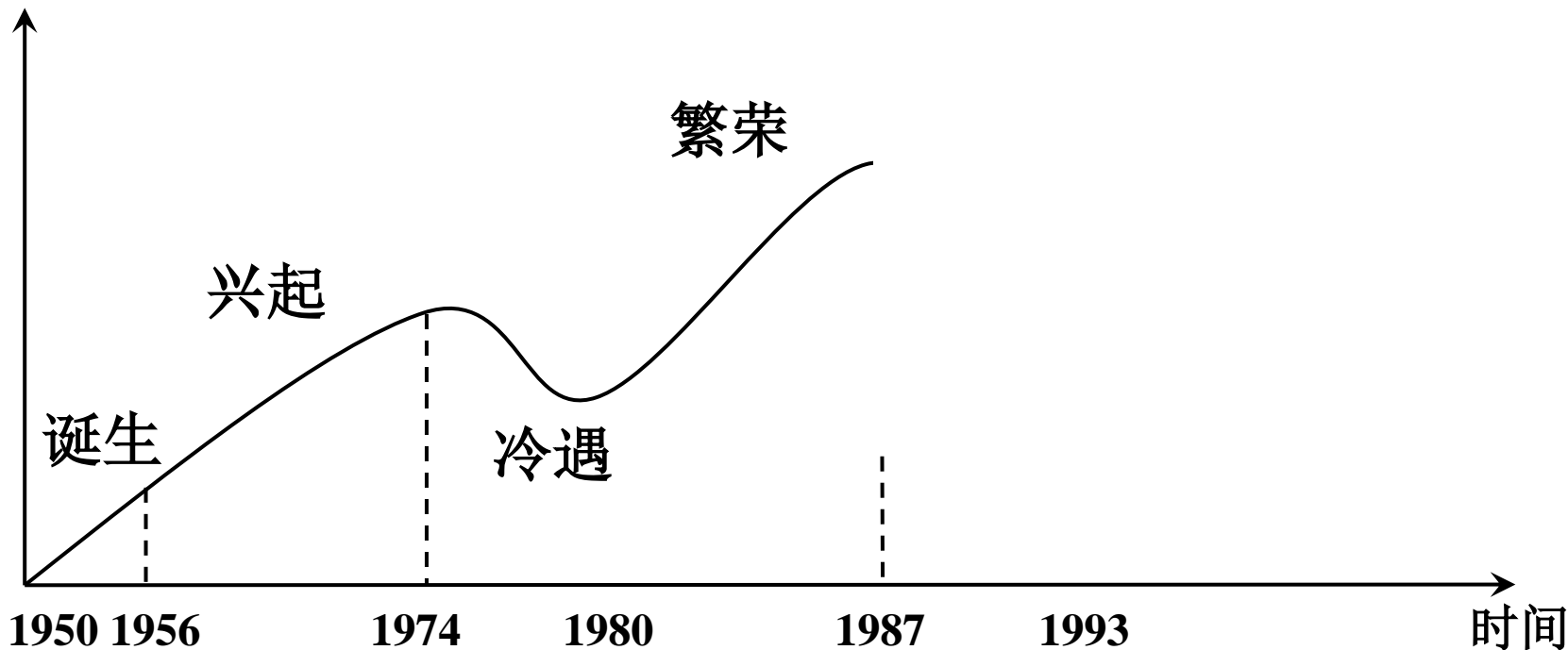


# 语言智能与应用

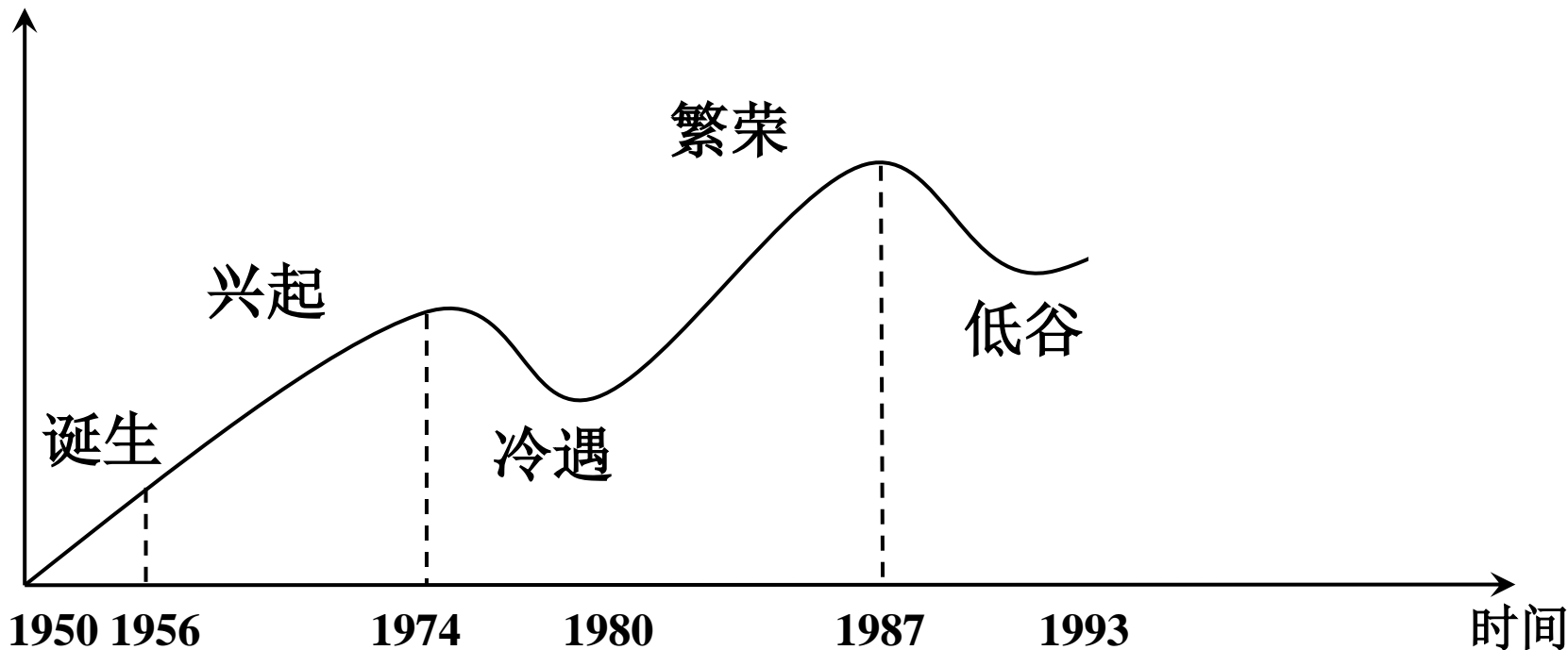
宗成庆

中国科学院自动化研究所

## ● 1987：笑傲江湖



## ● 1993：欲哭无泪



# 关于AI

● 2017

你有，我有，  
全都有啊！



感知、认知、  
控制、行为

## 1. Deep Learning 在语音识别、图像处理等中取得了很大的成功



6-11 July 2015, Lille, France

# 关于语言智能的讨论



Neil Lawrence said ...

“NLP is kind of like a rabbit in the headlights of the deep learning machine, waiting to be flattened.”



A Professor of Machine Learning  
at the University of Sheffield



但实际上并没有出现 Neil 所说的情况。您认为，主要原因是什么？

2. Deep learning (DL) 需要大规模训练样本，而且测试样本与训练样本比较一致时，才会获得较好的结果，可是对于儿童学习和使用语言来说，不需要太多的例句，因此小样本学习成为学术界关注的问题，但尚未建立有效的模型。企业界在建立自己的应用系统时，都采用了DL方法吗？是如何处理训练数据缺乏这一问题的（例如，面对一个新的领域、新的任务）？



3. 曾几何时，在NLP领域理性主义方法（specifically, rule-based approaches）与经验主义方法（Statistics-based approaches）针锋相对，各不相让，而近10余年来，理性主义者的呼声似乎销声匿迹了，难道说经验主义方法彻底击败了理性主义方法吗？换句话说，统计方法可以完全替代规则方法吗？企业在建立自己的系统时是否还使用规则方法？还需要学界进一步研究的规则方法问题吗？

4. 学术界在从事研究时往往“一窝蜂”，如近年来的DL方法和知识图谱、问答系统等。而往往这些问题都是由企业界引领的（不管是问题和方法的提出，还是在数据和平台的领先优势上）。企业界的专家能够为学界提出一些建议或希望？（无论是研究的问题、所采用的方法，还是研究方式）

5. 句法分析是NLP中一个非常困难的问题，性能一直很难提高，而且中文 Parser 的性能始终比英文的 Parser 低约5个百分点。在实际系统中，Parser 有什么具体的应用吗？如果继续研究这项技术，应该如何进行？（实际上，汉语分词也有类似的问题）

6. 产学研成功合作是学界和业界共同关注的问题，但在实际执行过程中，往往不是那么容易。您认为成功合作的模式应该是什么样的？这种模式是可以复制的吗？是否可以分享一下成功的经验或失败的教训？

7. 近年来，在网络信息处理，包括NLP领域，学界与业界从事的研究内容界限越来越不明显，而业界的平台和待遇优势却远远超过学界，因此包括美国在内，越来越多的学界大老走向企业界，大学和研究所培养的一流人才相当大的比例流向了各大企业。您如何看待这一问题？学界的出路在哪里？

8. 一群鸡和兔子，放在同一个笼子里，上面有35个头，下面有94只脚，问有多少只鸡、多少只兔？



常识与专业知识的是如何区分的？常识是如何表示、学习和更新的？



如何判断哪些是异常用词或有背于常理的语言表达？判断的标准是什么？

谢谢  
Thanks!