

AiCon

全球人工智能与机器学习技术大会

通过语音和语言技术打造AI教育平台

秦龙

先声教育/联合创始人&CTO



极客时间

重拾极客精神·提升技术认知

下载极客时间App

获取有声IT新闻、技术产品专栏，每日更新



扫一扫下载极客时间App

人工智能基础课

“通俗易懂的人工智能入门课”

王天一
博士
副教授



扫一扫，免费试读

AI技术内参

你的360度人工智能信息助理

洪亮劫
Etsy
数据科学主管



扫一扫，免费试读



关注落地技术，探寻AI应用场景

- 14万AI领域垂直用户
- 8000+社群技术交流人员，不乏行业内顶级技术专家
- 每周一节干货技术分享课
- AI一线领军人物的访谈
- AI大会的专家干货演讲整理
- 《AI前线》月刊
- AI技能图谱
- 线下沙龙



扫码关注带你涨姿势

QCon

全球软件开发大会

成为软件技术专家的 必经之路

[北京站] 2018

会议：2018年4月20-22日 / 培训：2018年4月18-19日

北京·国际会议中心

8折 购票中, 每张立减1360元

团购享受更多优惠



识别二维码了解更多

ArchSummit

全球架构师峰会

2018 · 深圳站

从2012年开始算起，InfoQ已经举办了9场ArchSummit全球架构师峰会，有来自Microsoft、Google、Facebook、Twitter、LinkedIn、阿里巴巴、腾讯、百度等技术专家分享过他们的实践经验，至今累计已经为中国技术人奉上了近千场精彩演讲。

限时**7折**报名中，名额有限，速速报名吧！

● 2012.08.10-12 深圳站



● 2018.07.06-09 深圳站

会议：07.06-07.07

培训：07.08-07.09



TABLE OF CONTENTES

简介

K12的作业问题

智能语音测评

智能写作批改

AI+教育的未来



先声教育

教育领域的人工智能技术服务商



TABLE OF CONTENTES

简介

K12的作业问题

智能语音测评

智能写作批改

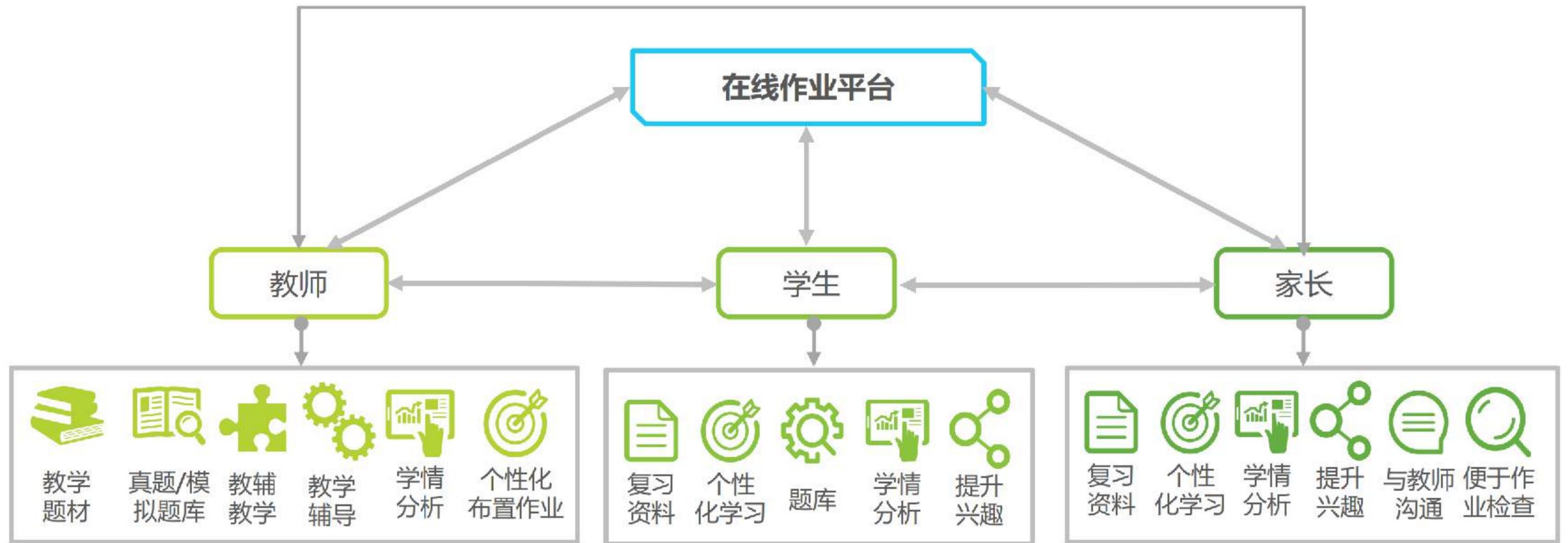
AI+教育的未来

K12教育

- 小学、初中、高中
- 28万所学校
- 1.6亿学生
- 家庭教育支出： >30% 家庭支出
- 国家教育支出： >4% GDP

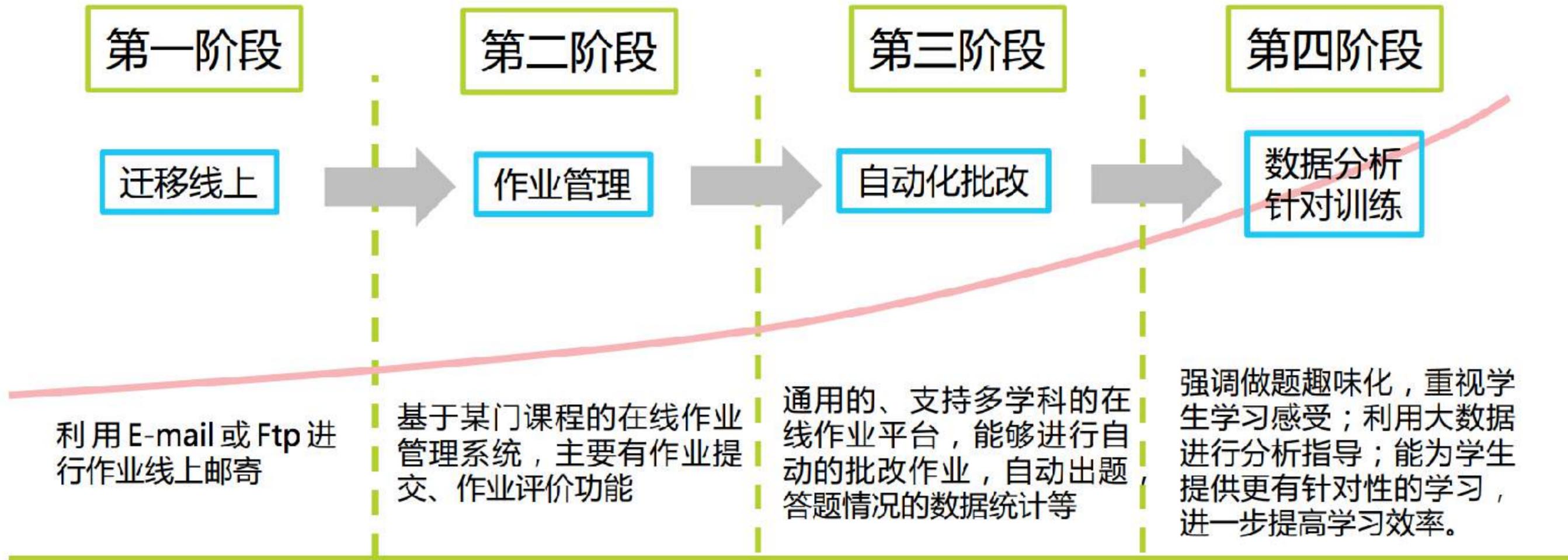
作业的问题

- 老师
 - 组织作业，批改作业，大量重复性劳动
- 学生
 - 作业多，无反馈，缺乏有针对性的指导
- 家长
 - 陪伴孩子做作业，平均每月19天



2017年中国在线作业平台的概念及作用

来源：艾瑞咨询研究院自主研究及绘制



2017年中国在线作业平台的发展历程

来源：艾瑞咨询研究院自主研究及绘制

自动批改

- 以英语作业为例
 - 客观题：单选、完形填空、阅读理解等
 - 主观题：口语、翻译、作文

TABLE OF CONTENTES

简介

K12的作业问题

智能语音测评

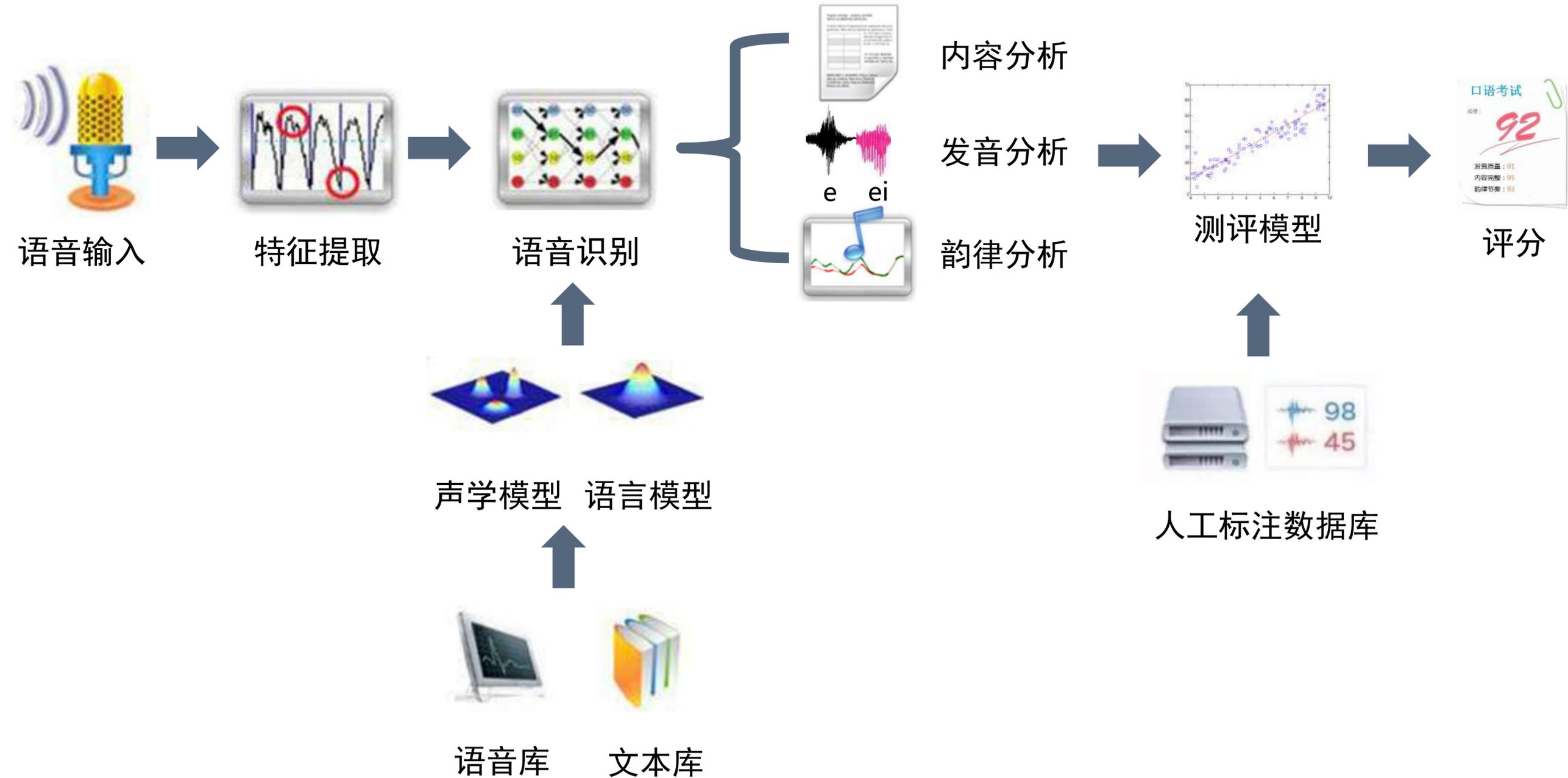
智能写作批改

AI+教育的未来

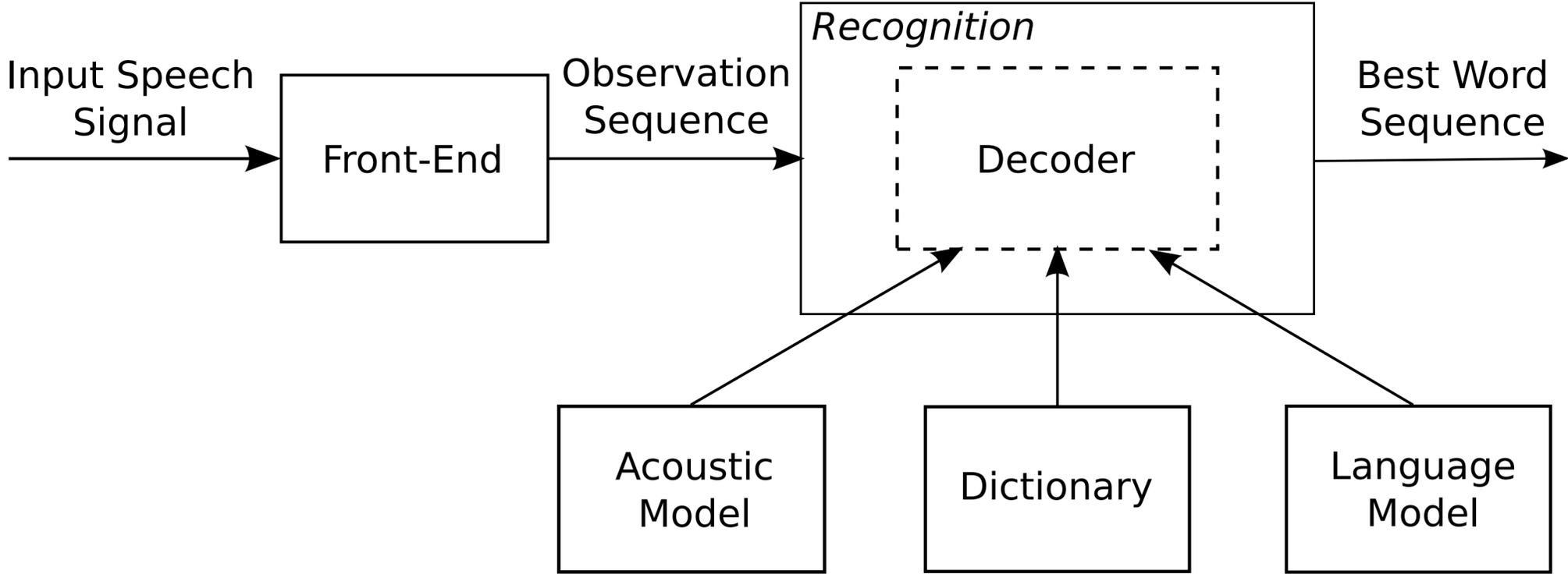
智能口语测评

- 自动打分，精准反馈
- 语音识别 (LSTM) 提取特征
- 测评模型 (LR、SVM、DNN) 估计打分

口语测评原理



语音识别任务



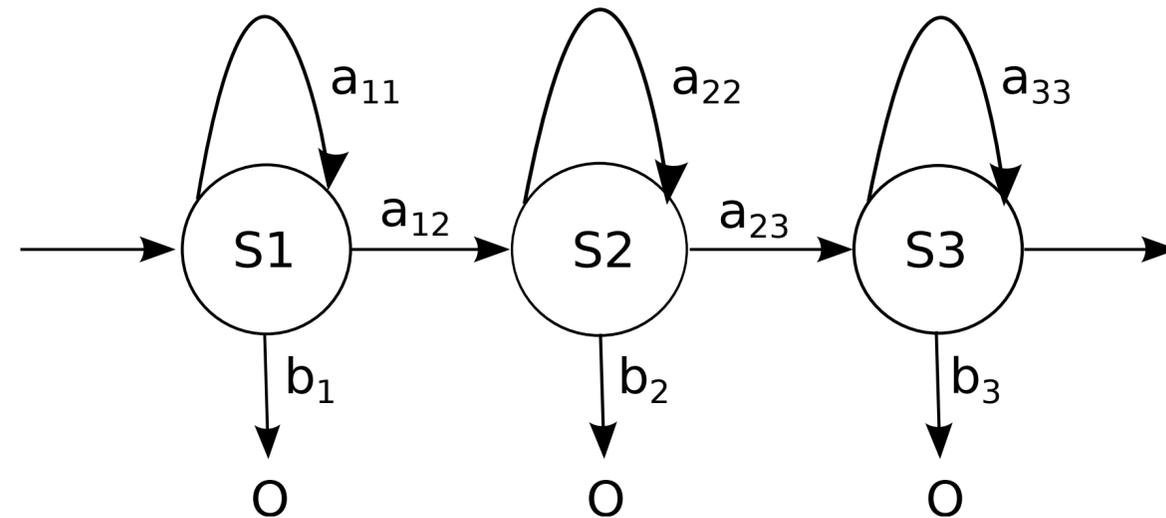
$$\hat{W} = \arg \max_W P(W | O) = \arg \max_W \frac{P(O | W) * P(W)}{P(O)} = \arg \max_W P(O | W) * P(W)$$

声学模型
Likelihood

语言模型
Prior

声学模型

- 声学模型用来计算语音信号和给定的发音之间的概率分布
- 根据语音信号时序性，我们使用HMM (hidden Markov model)对语音信号建模



- 常见的HMM观测模型有GMM (Gaussian mixture model)和深度神经网络 (LSTM or CNN)

语言模型

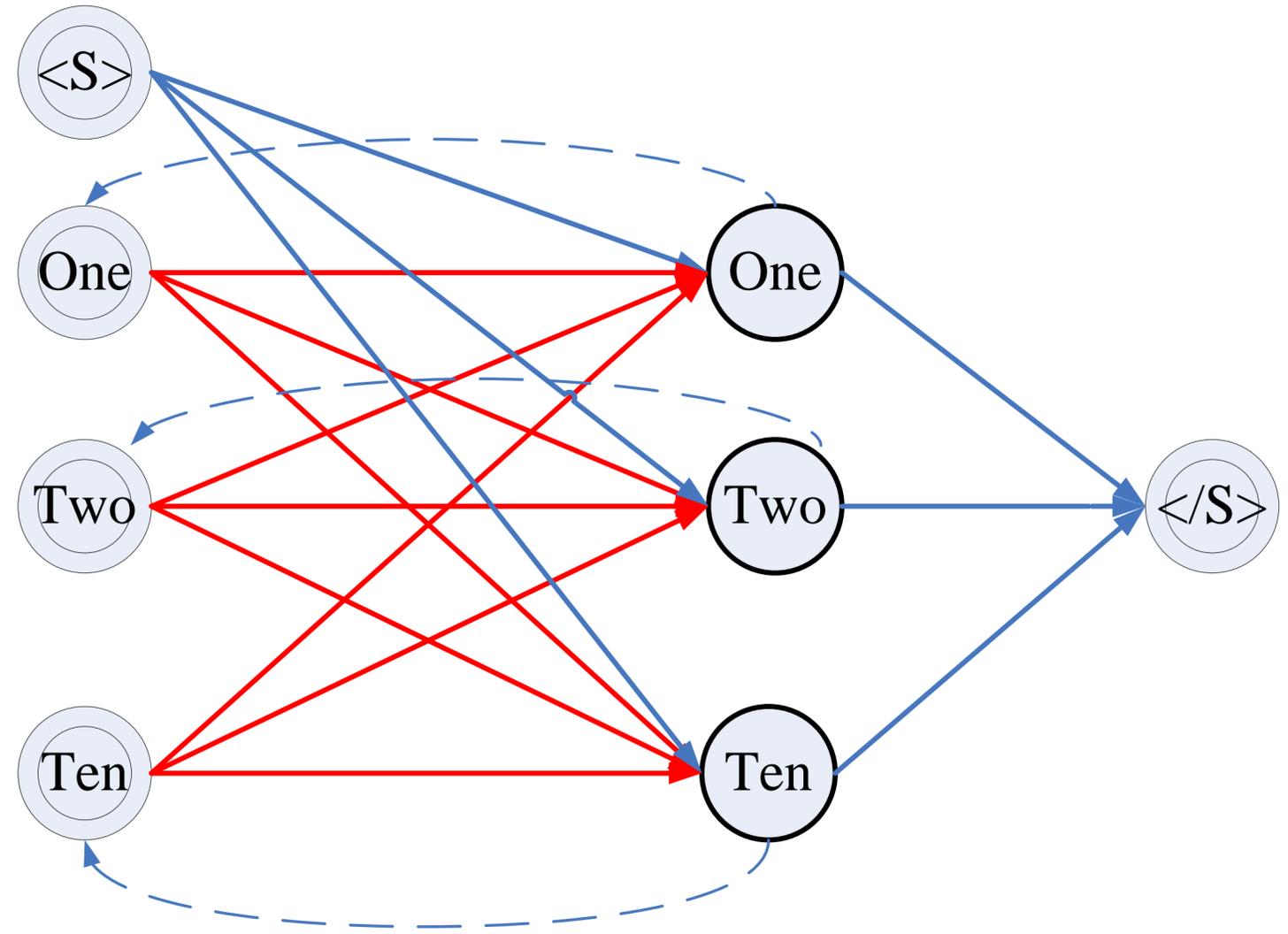
- 语言模型用来计算一句话 $W(w_1, w_2, w_3 \dots)$ 出现的概率
- 最基本的语言模型为n-gram

$$P(W) = P(w_1) * P(w_2 | w_1) * P(w_3 | w_1, w_2) * \dots * P(w_N | w_1, w_2, \dots, w_{N-1})$$

- RNN (recurrent neural network) 模型也开始广泛使用

解码过程

- 寻找搜索空间内的最佳路径
- 采用动态规划方法 (Viterbi search)
- 为了提高效率，采用 Beam Search 裁剪得分低的分支



测评特征

- 多种发音、流利度、韵律、后验概率特征用于测评模型的输入
- 发音特征：声学模型似然度得分、置信度得分等
- 流利度：语速、停顿等
- 韵律：基频的均值和方差、能量分布、语调等
- 后验概率：深度神经网络输出的后验概率

测评模型

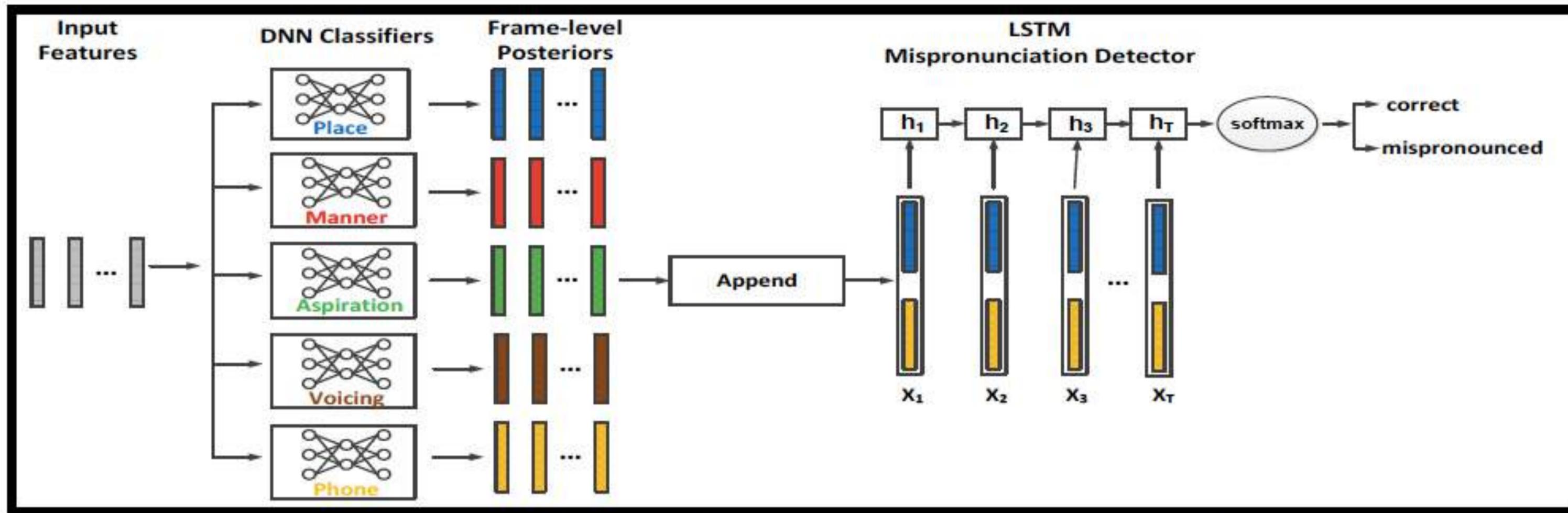


Figure 2: Overview of mispronunciation detection framework

口语测评题型

朗读题

- 音标
- 单词
- 句子
- 篇章
- 选择题

半开放题

- 翻译
- 短问答
- 复述作文
- 要点作文

开放题

- 话题作文

口语测评准确性

题型	相关度
朗读题	0.82
短问答	0.91
复述题	0.85
总分	0.93

口语测评展示

- 评分准确度达到93%
- 提供15个维度教学场景的反馈
- 覆盖中高考全部近30个题型
- 支持2个平台和5个客户端
- 匹配多种网络环境，自动切换离线/在线使用
- 高可靠性，每秒可处理10000并发评测请求
- 日均测评1000万次，处理语音数据1万小时



TABLE OF CONTENTES

简介

K12的作业问题

智能语音测评

智能写作批改

AI+教育的未来

智能写作批改

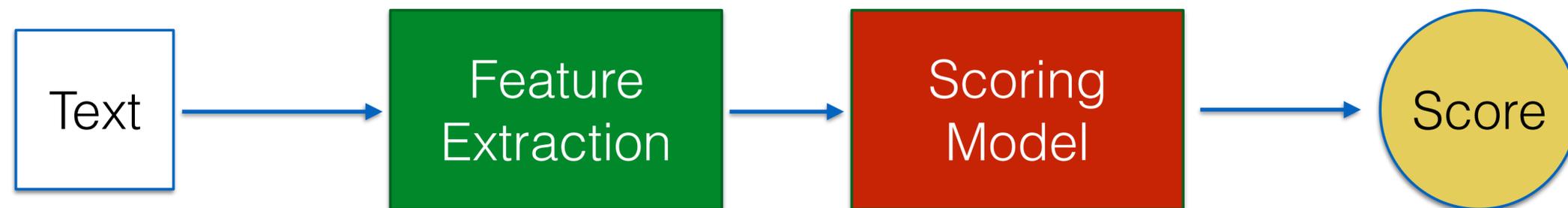
- Essay Grading: 对作文进行自动打分
- Error Correction: 纠正拼写、语法、用词错误
- 与人工批改相比: 高效、及时、客观

自动批改方法

- Regression Model: 基于语言学特征, 利用机器学习技术构建评分的回归/分类模型
- Neural Network: 通过深度神经网络 (LSTM) 构建从文字到语意到评分的映射模型

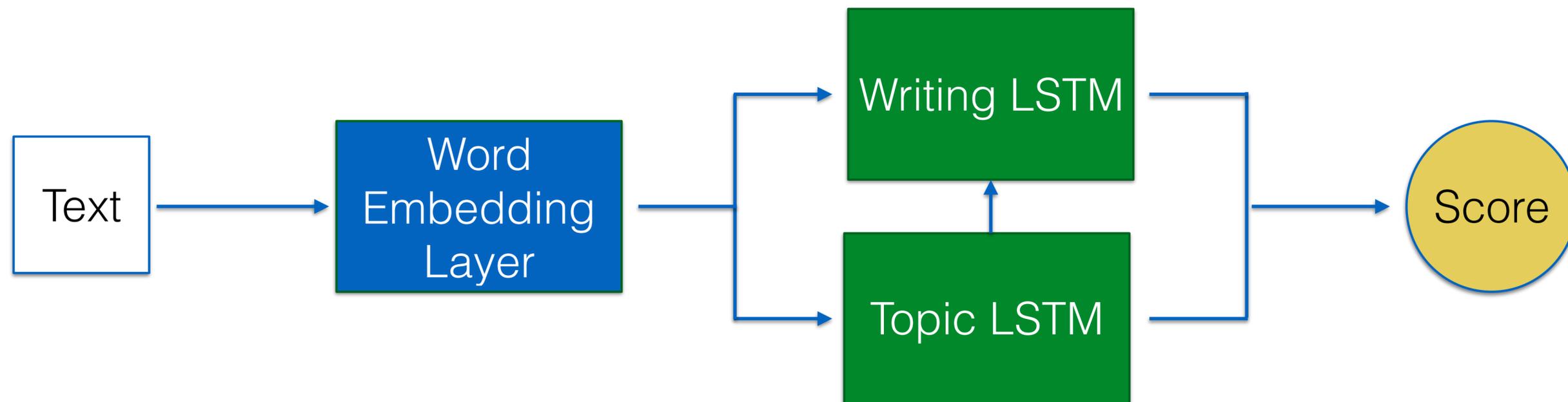
Regression Model

- 特征：词汇、词性 (part of speech)、文章长度、ngram分布、语法特征、主题等
- 模型：Linear regression、SVM、Boosting等



Neural Network

- 通过神经网络将词向量投射到高维空间，然后通过卷积/回归神经网络提取深度语音信息
- 融合写作网络和主题网络



自动批改评分准确性

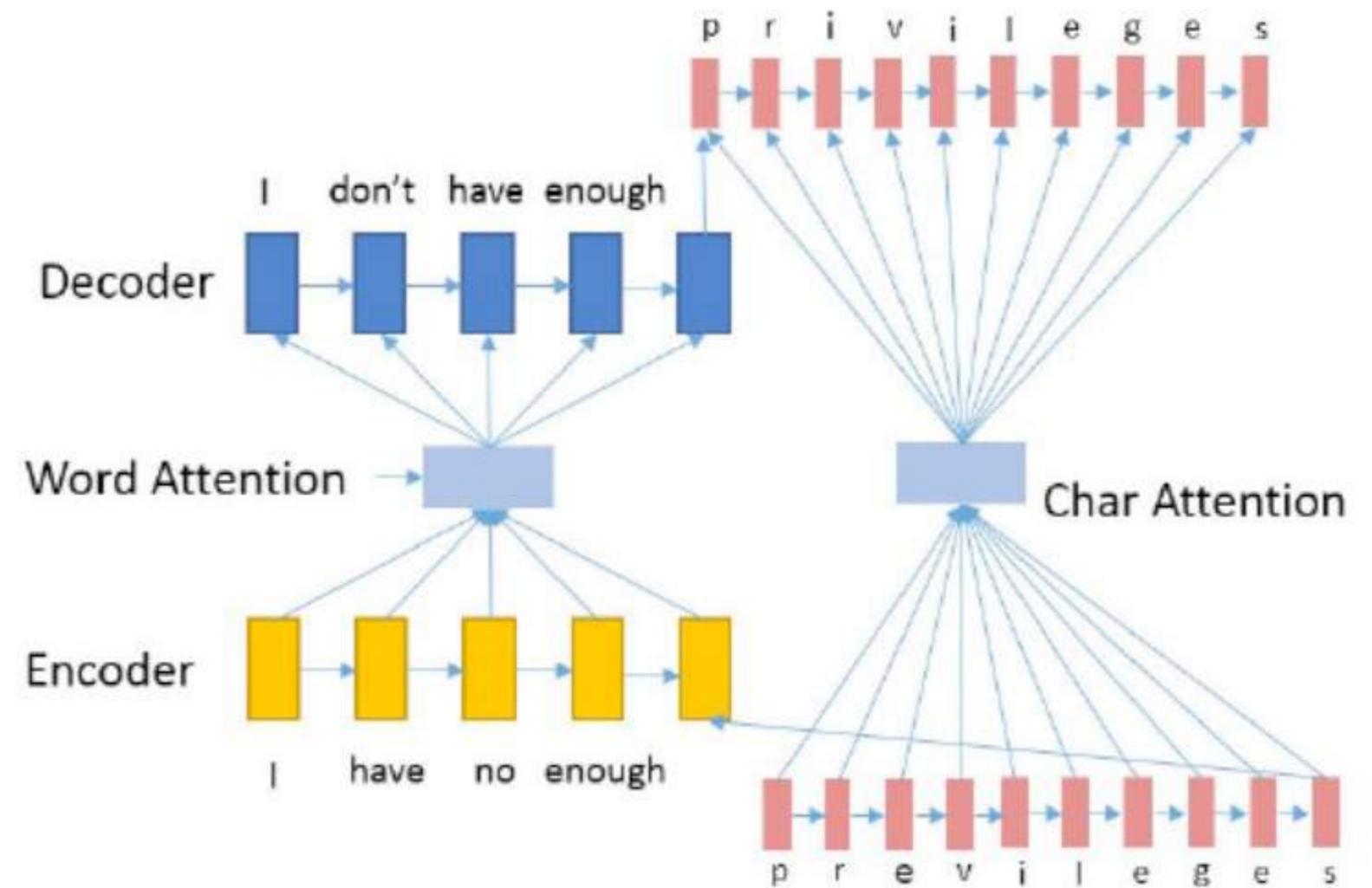
System	QWK
Baseline Regression Model	0.679
Baseline Neural Network	0.716
SingSound Regression Model	0.739
SingSound Neural Network	0.754

错误纠正方法

- 基于人工制定的规则的方法
- 基于不同错误类型的分类的方法
- 基于机器翻译的方法

基于机器翻译的错误纠正方法

- Encoder-Decoder神经网络模型
- 把有错误的句子作为源序列, 把正确的句子作为目标序列
- 引入character level的对齐来处理typo



错误纠正准确性

System	F _{0.5} Score
Baseline SMT Model	51.5
SingSound SMT Model	52.3
SingSound NMT Model	53.4

写作批改展示

- 无监督批改，无需大量范文
- 提供词汇、语法、内容、可读性单项评分
- 反馈拼写、用词、语法错误

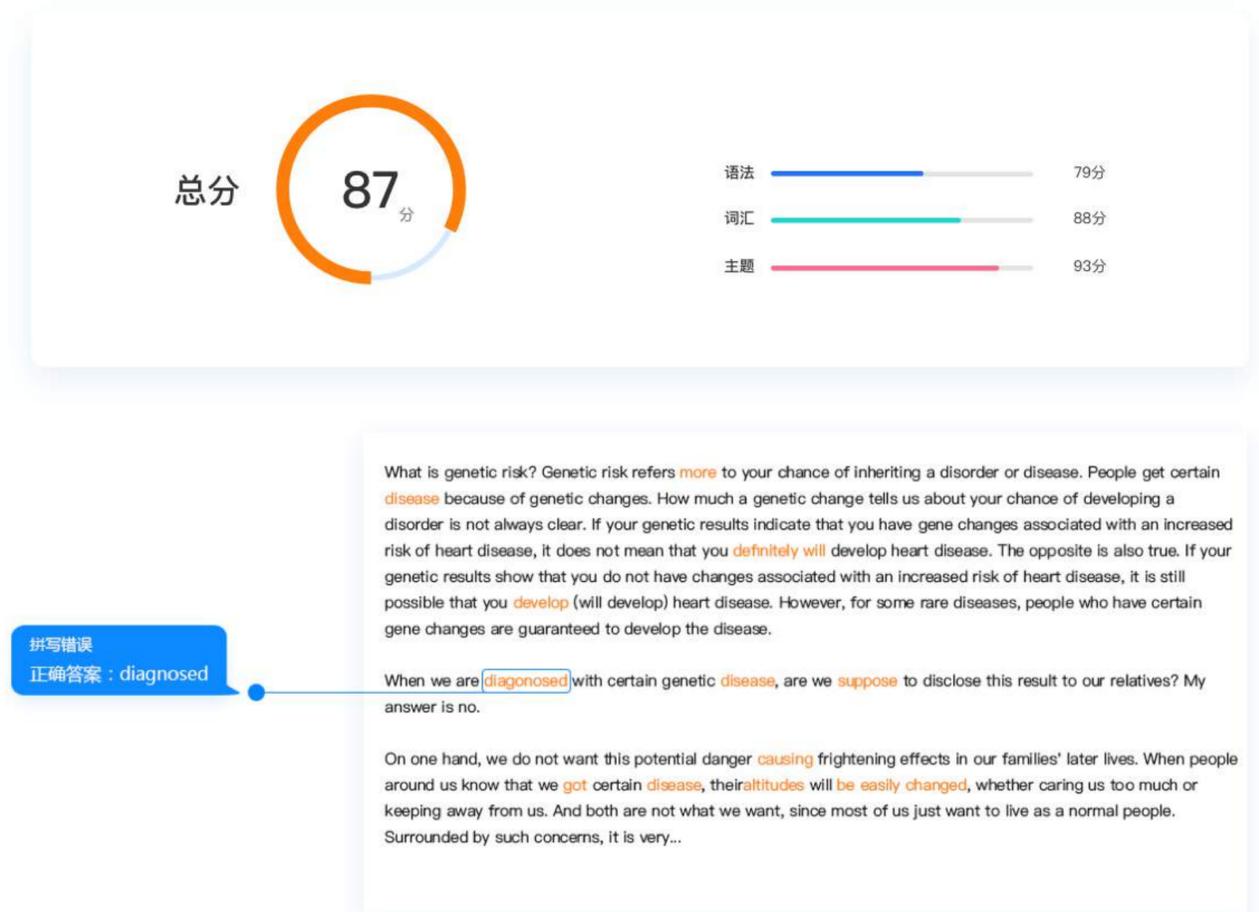


TABLE OF CONTENTES

简介

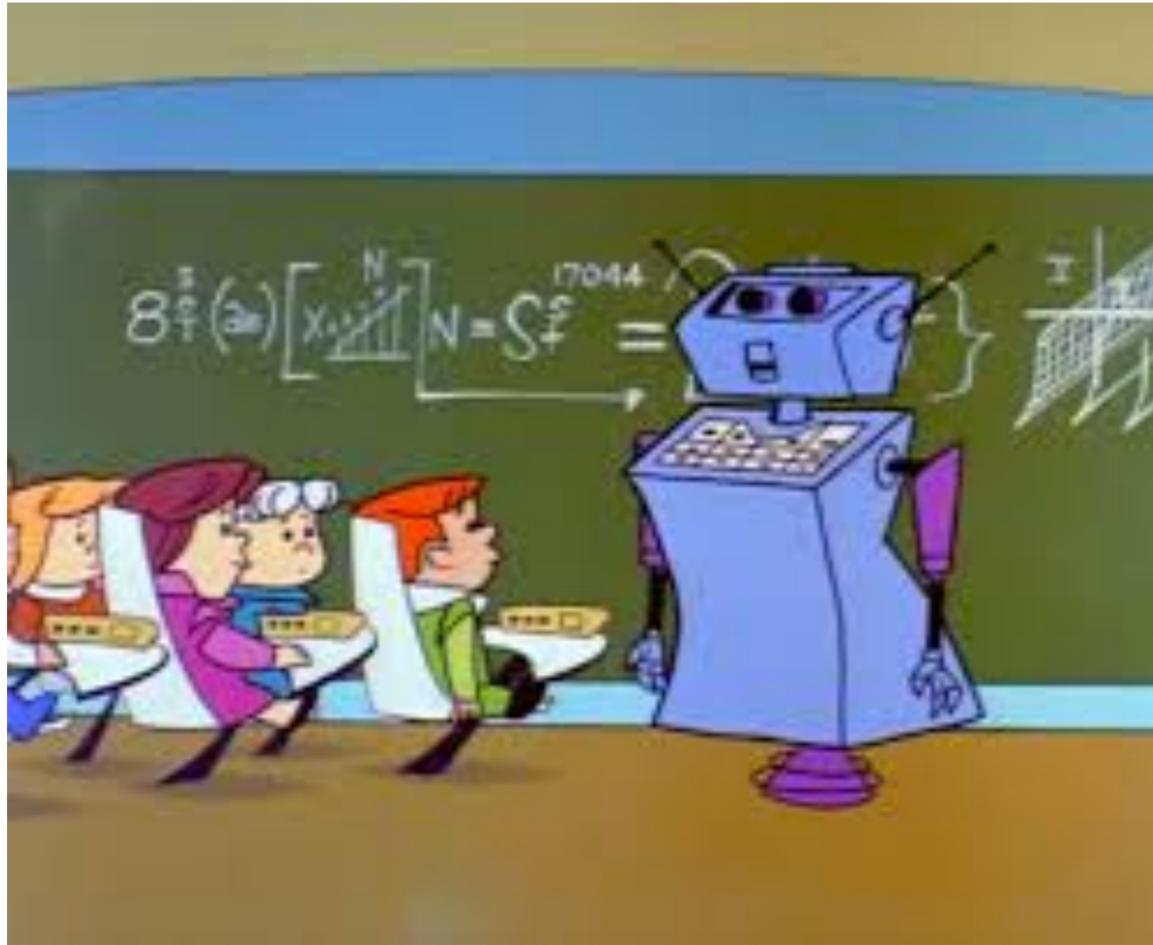
K12的作业问题

智能语音测评

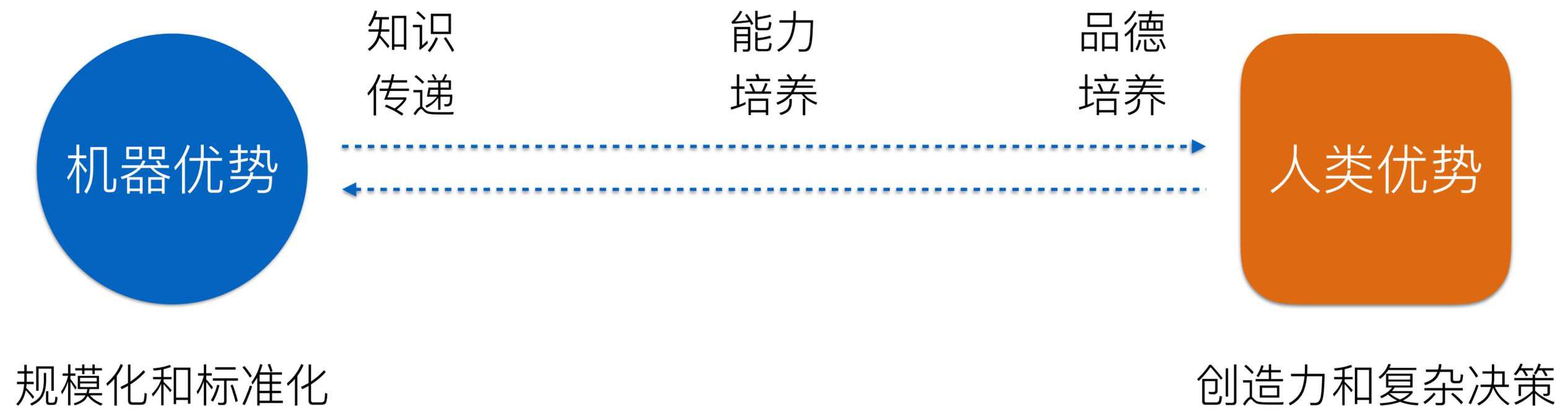
智能写作批改

AI+教育的未来

教育的未来=AI老师?



教育的未来=AI+老师



Thanks!