

自适应学习 机器学习在开心词场中的应用

王新义@HJ



日程

- 关于教育及互联网教育
- 机器学习在沪江的应用
- 机器学习在开心词场中应用
 - 自适应词汇量测试
 - 记忆模型
- DKT

关于教育及互联网教育

- 教育是最传统、最复杂、涉及面广的社会活动
- 教育痛点：公平、效率、痛苦
- 互联网教育：低频、高交互



真正的互联网教育是具备大规模、复杂交互行为的普惠性知识学习和教育。

——阿诺

AI背景下互联网教育

- 愿景：帮助教育进化
 - 因材施教的千年愿景
 - 智能化，让不可能成为可能
 - 应试？还是要真正帮助人类成长
- 使命：使用机器学习(AI)技术改造和促进人类自身学习(提高学习效率和学习效果)



注:古希腊：Socrates & Plato & Aristotle & Alexander

古先秦：孔子一生，据史书记载，教导了三千多学生。教导3000学生如何做到因材施教？？

学习，成为更好的自己

机器学习在沪江的应用

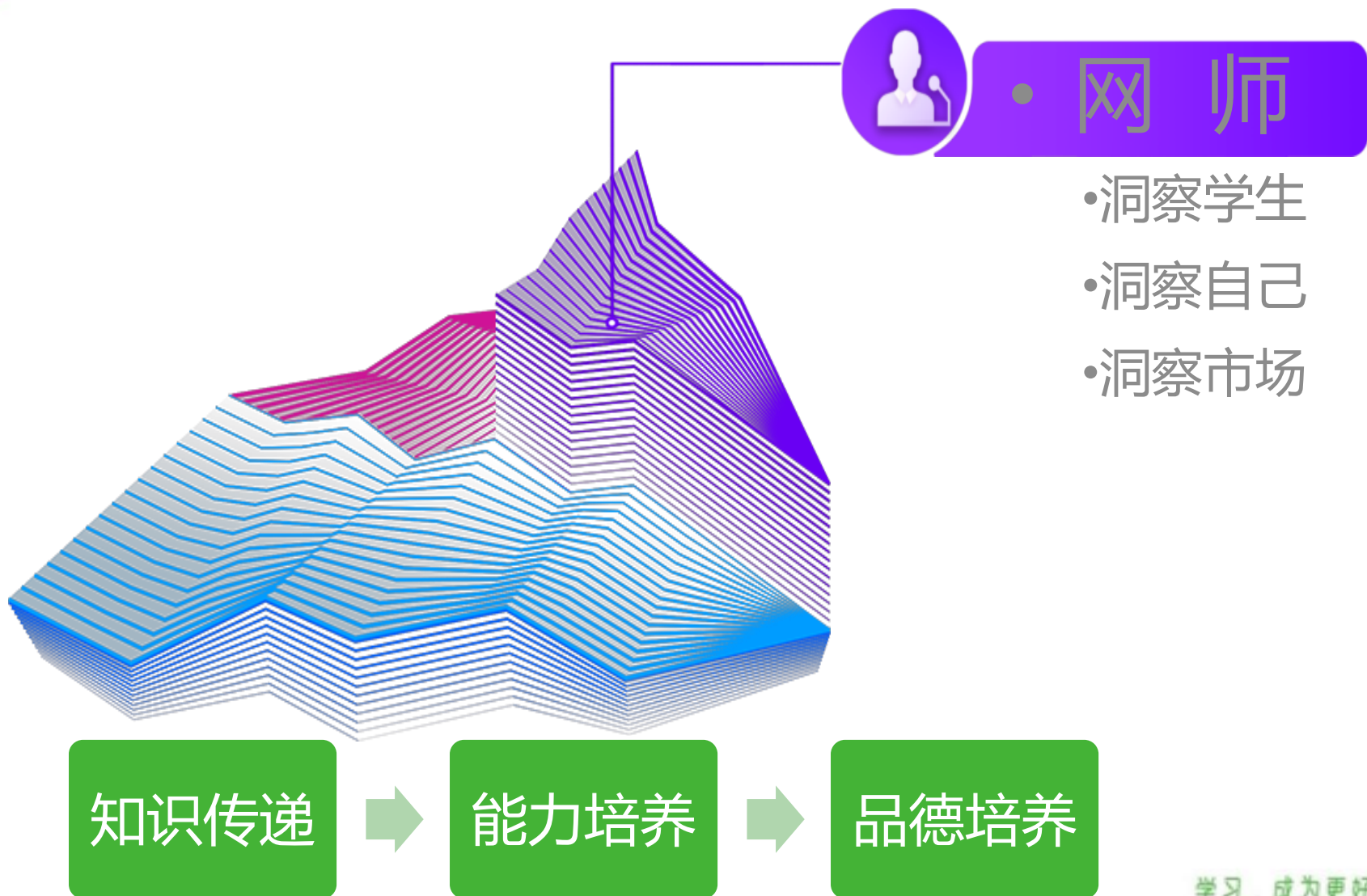
AI在HJ的四个应用场景：

- 自适应学习
- 人机交互
- 教学过程监控
- 内容加工

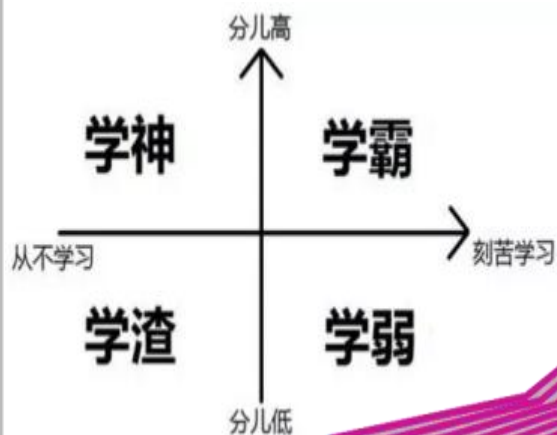


学习，成为更好的自己

机器学习在沪江的应用

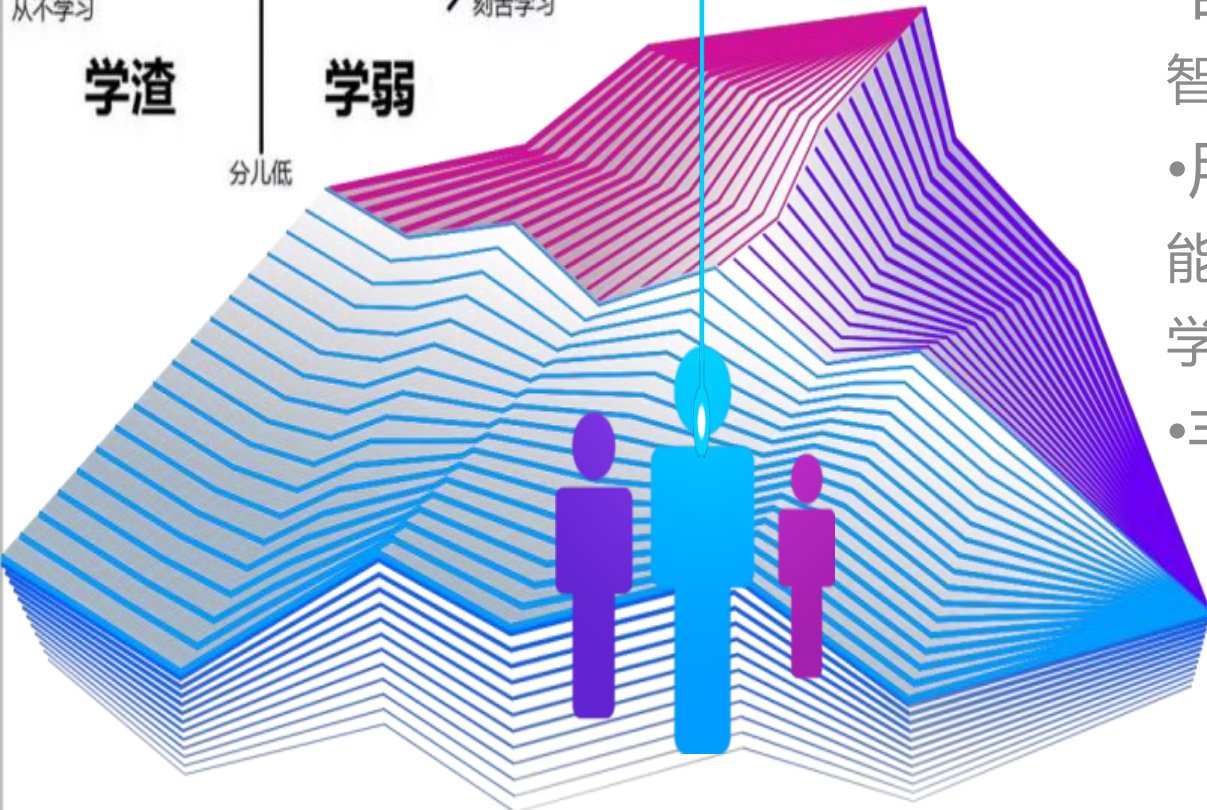


机器学习在沪江的应用



• 学习者

- 自适应学习：智能导学服务、智能学习助手
- 用户模型：学习需求、学习能力、学习动力、学习态度、学习信念等
- 丰富多样的学习体验

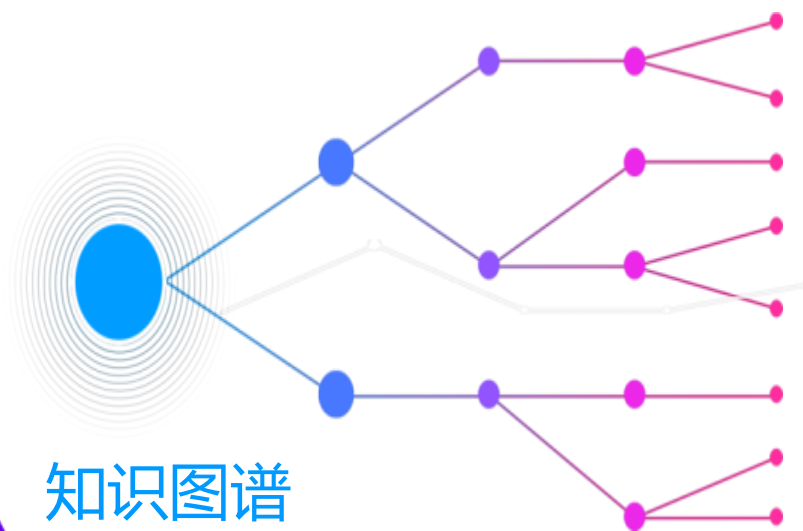
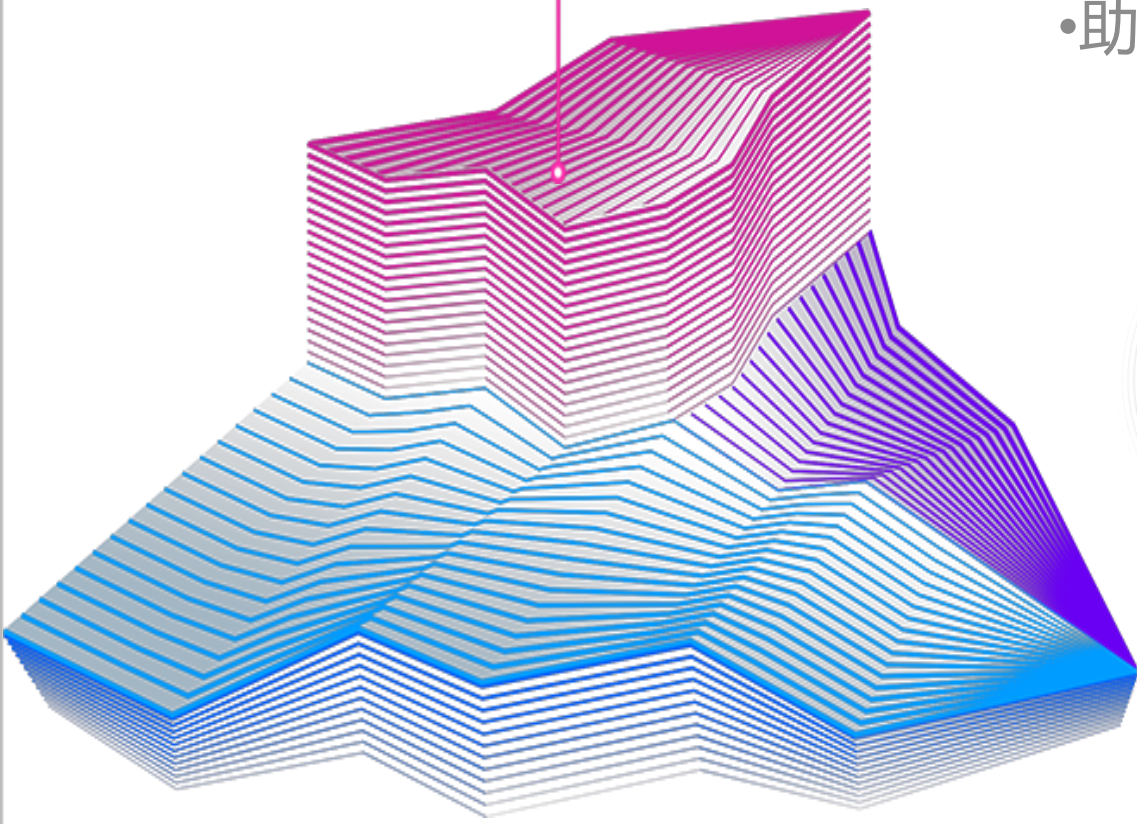


机器学习在沪江的应用



• 内 容

- 加速内容产品化
- 助推内容商品化



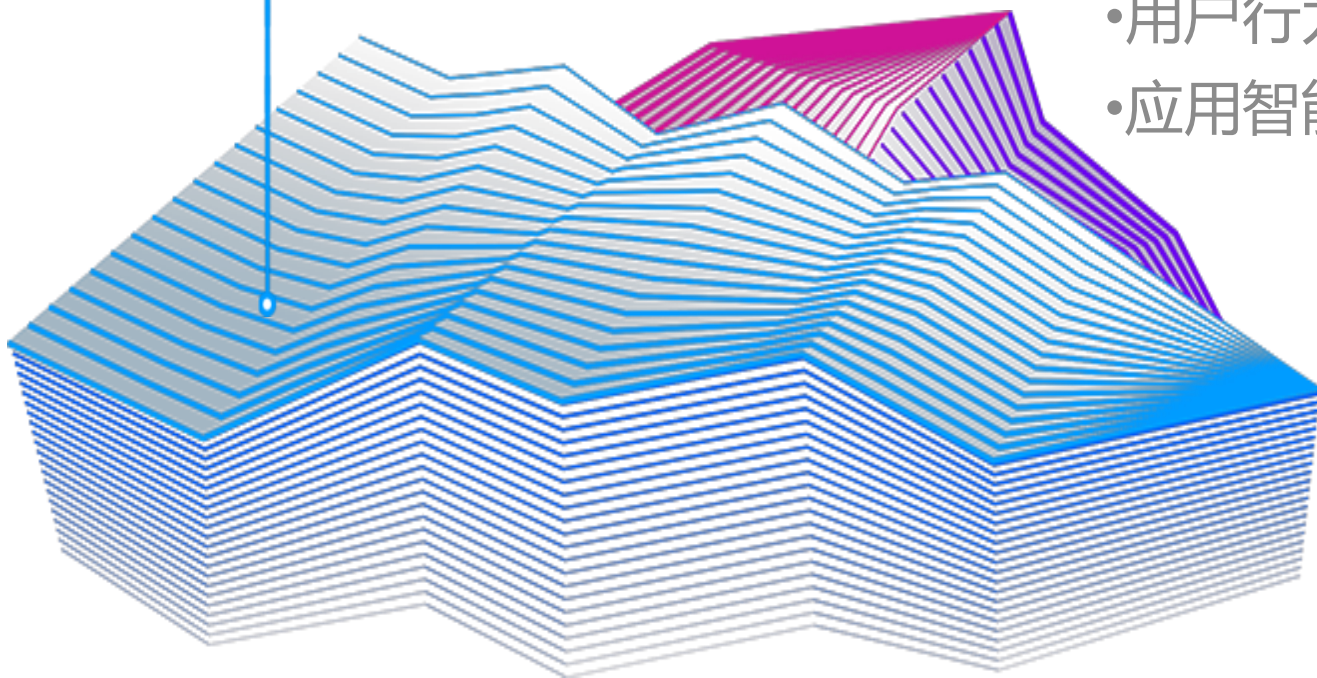
学习，成为更好的自己

机器学习在沪江的应用



• 工具

- 通过开放接口与教学场景深度融合
- 用户行为感知和数据分析
- 应用智能云服务



机器学习在沪江的应用

售前

- 推荐系统
- 例子系统

售中

- 金融风控

学中

- 推荐（自适应学习）
- 工具线产品的算法支持

学后

- 教师及课程质量评价
- 学生的测评及评价

工具线

- 教研：打标系统
- bot：教学助理
- NLP应用

基础平台层

自适应学习系统

推荐系统

NLP服务系统

机器学习平台

特征工程

行为及特征模型

学习者模型

质量评价

转化/时间序列预测模型

知识图谱

模型算法、框架

机器学习框架

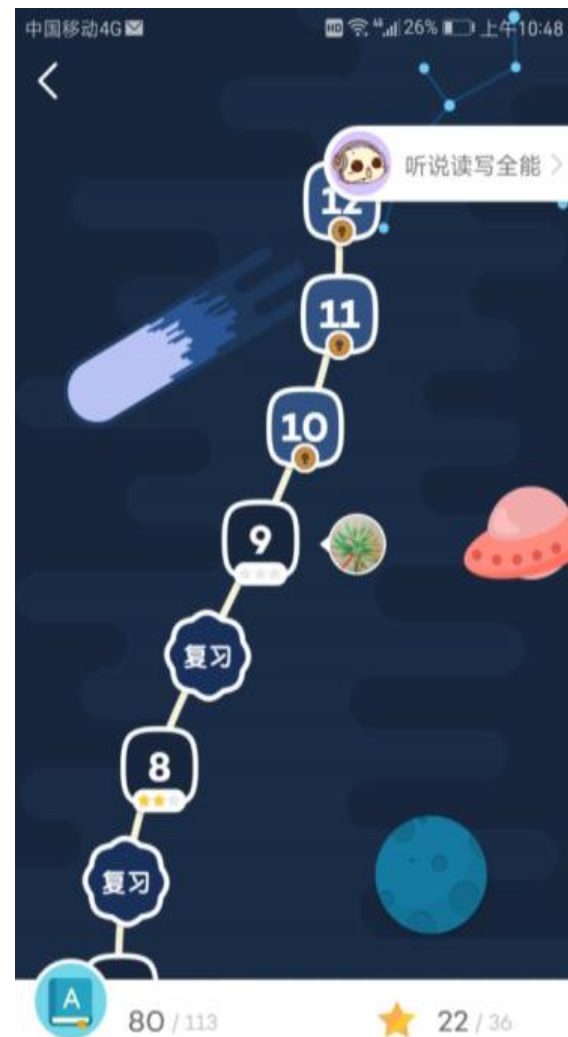
深度学习框架

NLP/NLU模型算法

图片、语音识别

DKT/BKT、IRT、RM理论

机器学习在开心词场中应用



学习，成为更好的自己



机器学习在开心词场中应用

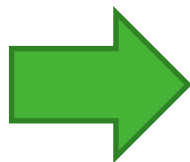
- 自适应词汇量测试
- 记忆模型
- 词性标注



自适应词汇量测试

自适应词汇量测试

- | | | | |
|-------------------------------------|------------------------------------|-------------------------------------|------------------------------------|
| <input type="checkbox"/> skirt | <input type="checkbox"/> sword | <input type="checkbox"/> blink | <input type="checkbox"/> natter |
| <input type="checkbox"/> chop | <input type="checkbox"/> bounce | <input type="checkbox"/> bait | <input type="checkbox"/> deflect |
| <input type="checkbox"/> stall | <input type="checkbox"/> stance | <input type="checkbox"/> juggle | <input type="checkbox"/> flurry |
| <input type="checkbox"/> hunt | <input type="checkbox"/> bat | <input type="checkbox"/> tinker | <input type="checkbox"/> whim |
| <input type="checkbox"/> storm | <input type="checkbox"/> mercy | <input type="checkbox"/> drought | <input type="checkbox"/> shaggy |
| <input type="checkbox"/> rope | <input type="checkbox"/> cellar | <input type="checkbox"/> dangle | <input type="checkbox"/> reproach |
| <input type="checkbox"/> stake | <input type="checkbox"/> bald | <input type="checkbox"/> outright | <input type="checkbox"/> trundle |
| <input type="checkbox"/> constraint | <input type="checkbox"/> portray | <input type="checkbox"/> beware | <input type="checkbox"/> rampant |
| <input type="checkbox"/> hint | <input type="checkbox"/> grim | <input type="checkbox"/> wriggle | <input type="checkbox"/> judicious |
| <input type="checkbox"/> lawn | <input type="checkbox"/> tan | <input type="checkbox"/> glide | <input type="checkbox"/> whiff |
| <input type="checkbox"/> hedge | <input type="checkbox"/> refurbish | <input type="checkbox"/> ledge | <input type="checkbox"/> tandem |
| <input type="checkbox"/> nest | <input type="checkbox"/> saddle | <input type="checkbox"/> loot | <input type="checkbox"/> scathing |
| <input type="checkbox"/> truck | <input type="checkbox"/> scent | <input type="checkbox"/> feeble | <input type="checkbox"/> rind |
| <input type="checkbox"/> drift | <input type="checkbox"/> wallet | <input type="checkbox"/> seep | <input type="checkbox"/> strife |
| <input type="checkbox"/> harvest | <input type="checkbox"/> meadow | <input type="checkbox"/> greed | <input type="checkbox"/> grouse |
| <input type="checkbox"/> lick | <input type="checkbox"/> clap | <input type="checkbox"/> shrine | <input type="checkbox"/> bawl |
| <input type="checkbox"/> knit | <input type="checkbox"/> tickle | <input type="checkbox"/> stoop | <input type="checkbox"/> spool |
| <input type="checkbox"/> beam | <input type="checkbox"/> hollow | <input type="checkbox"/> numb | <input type="checkbox"/> pittance |
| <input type="checkbox"/> dull | <input type="checkbox"/> forbid | <input type="checkbox"/> bustle | <input type="checkbox"/> shrivel |
| <input type="checkbox"/> withdraw | <input type="checkbox"/> drip | <input type="checkbox"/> burp | <input type="checkbox"/> easel |
| <input type="checkbox"/> shave | <input type="checkbox"/> warranty | <input type="checkbox"/> swivel | <input type="checkbox"/> mayhem |
| <input type="checkbox"/> crush | <input type="checkbox"/> strive | <input type="checkbox"/> drab | <input type="checkbox"/> rascal |
| <input type="checkbox"/> summit | <input type="checkbox"/> cute | <input type="checkbox"/> throttle | <input type="checkbox"/> fledgling |
| <input type="checkbox"/> carve | <input type="checkbox"/> alley | <input type="checkbox"/> streamline | <input type="checkbox"/> signet |
| <input type="checkbox"/> lap | <input type="checkbox"/> junk | <input type="checkbox"/> napkin | <input type="checkbox"/> dowdy |
| <input type="checkbox"/> harsh | <input type="checkbox"/> hamper | <input type="checkbox"/> resilient | <input type="checkbox"/> lore |
| <input type="checkbox"/> clutch | <input type="checkbox"/> gleam | <input type="checkbox"/> topple | |
| <input type="checkbox"/> scrub | <input type="checkbox"/> plank | <input type="checkbox"/> blunder | |
| <input type="checkbox"/> toss | <input type="checkbox"/> grumble | <input type="checkbox"/> stead | |



- 减少测试时间
- 提升用户体验

帮助学生推荐合适的

- 词书
- 句子
- 文章

<http://testyourvocab.com/>

静态勾选
需要做很多题目

个性化，交互式
~40 道题目

学习，成为更好的自己

基本原理

静态考卷

- 每个学生所做的题目相同
- 在以下题目上浪费较多时间，影响用户体验
 - 肯定会做的容易题
 - 肯定不会做的难题

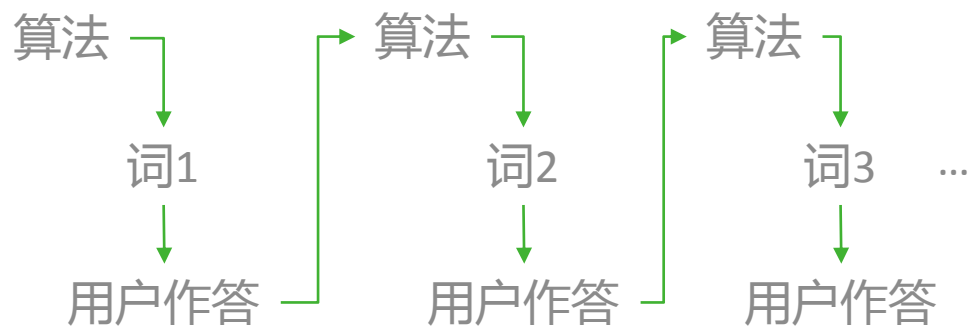
动态交互式测试

- 每个学生所做的题目不相同
- 下一道题目根据历史做题的反馈动态改变
- 算法可以聚焦于算法不确定的题目请学生回答，而避免在肯定会做和肯定不会做的题目浪费太多的时间

基本假设：

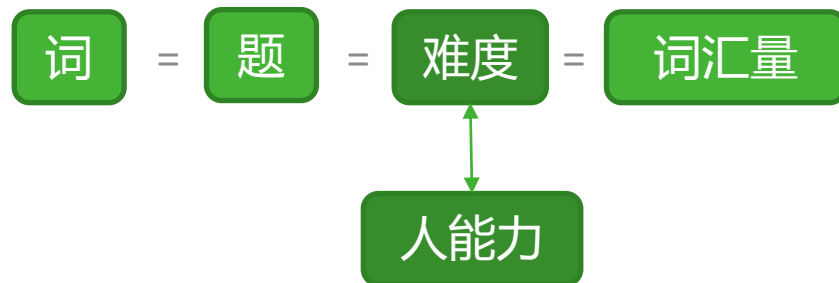
- 单词和单词之间有相对顺序
- 难度接近的单词，用户的作答也接近

流程



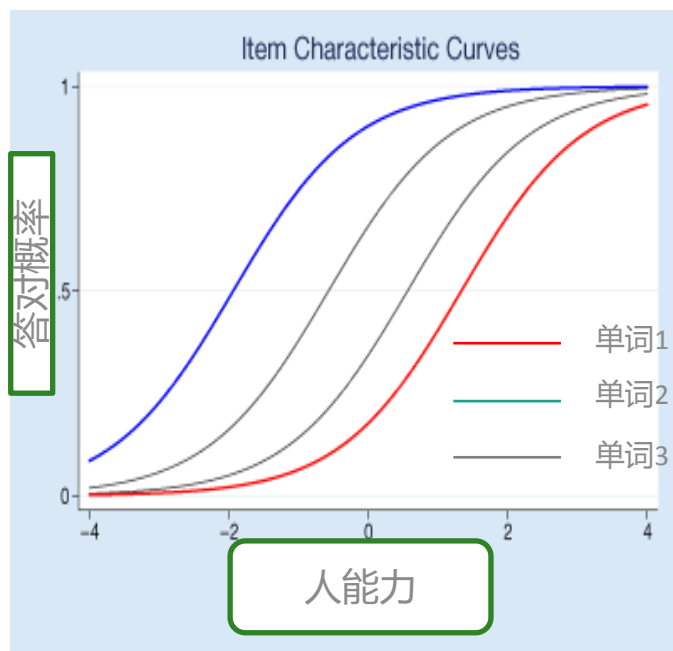
算法:

1. 根据做题记录，估计人能力
2. 出下一道题



- 词，题，难度，词汇量是对应的
- 难度已知:
 - 冷启动的时候，由词频保序变换而来
 - 有了用户作答数据，可以通过计算而得到
- 人能力和词难度可比。有了人能力，通过对应的难度，可以算得词汇量

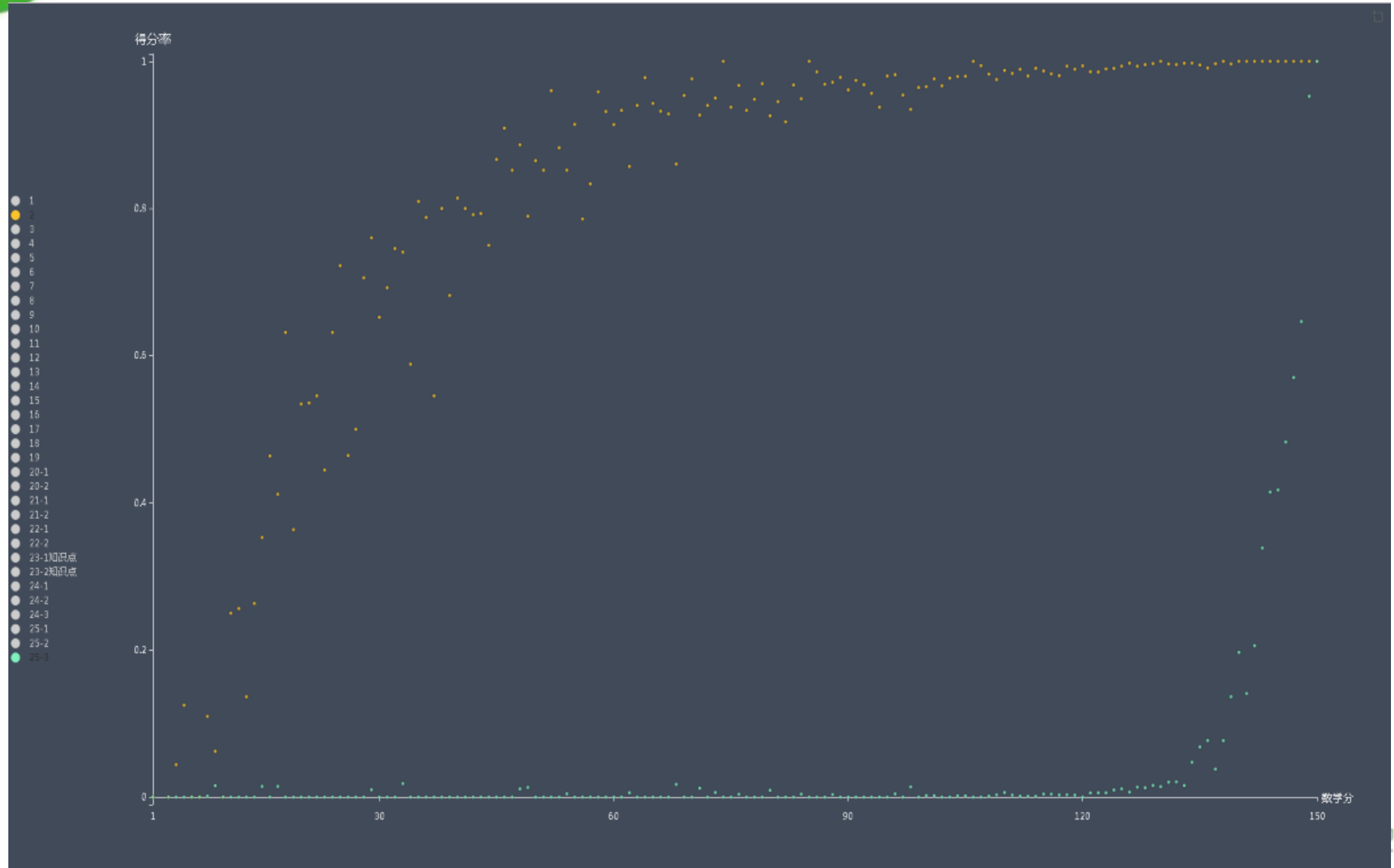
IRT (Item Response Theory)



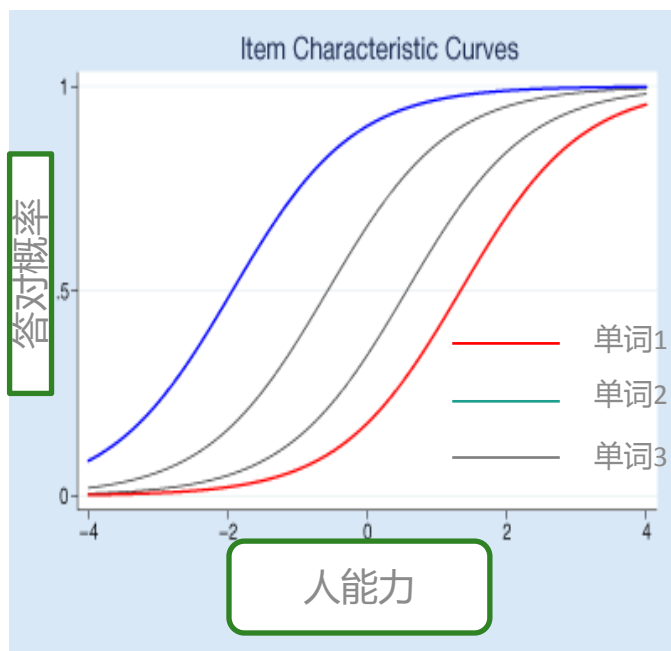
$$P(\text{答对} | \text{词难度}, \text{人能力}) = \frac{1}{1 + e^{-(\text{人能力} - \text{词难度})}}$$

- 人能力越高，答对概率越大
- 词难度越高，答对概率越小，曲线越靠右
- 人能力=词难度, 则答对概率0.5

IRT (Item Response Theory)



如何估计人能力？



word	correct	词难度
depress	1	0.9
take	0	0.1
delight	1	0.3
explain	1	0.4

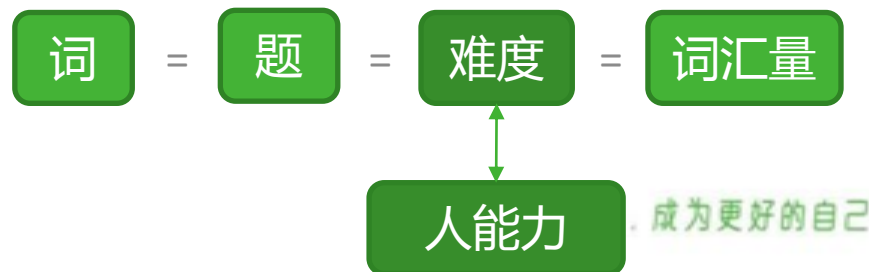
极大似然估计

max
人能力

$$\underbrace{\left(\frac{1}{1+e^{-(\text{人能力}-0.9)}}\right)}_{\text{depress}} \underbrace{\left(1-\frac{1}{1+e^{-(\text{人能力}-0.1)}}\right)}_{\text{take}} \underbrace{\left(\frac{1}{1+e^{-(\text{人能力}-0.3)}}\right)}_{\text{delight}} \underbrace{\left(\frac{1}{1+e^{-(\text{人能力}-0.4)}}\right)}_{\text{explain}}$$

$$P(\text{答对}|\text{词难度, 人能力}) = \frac{1}{1+e^{-(\text{人能力}-\text{词难度})}}$$

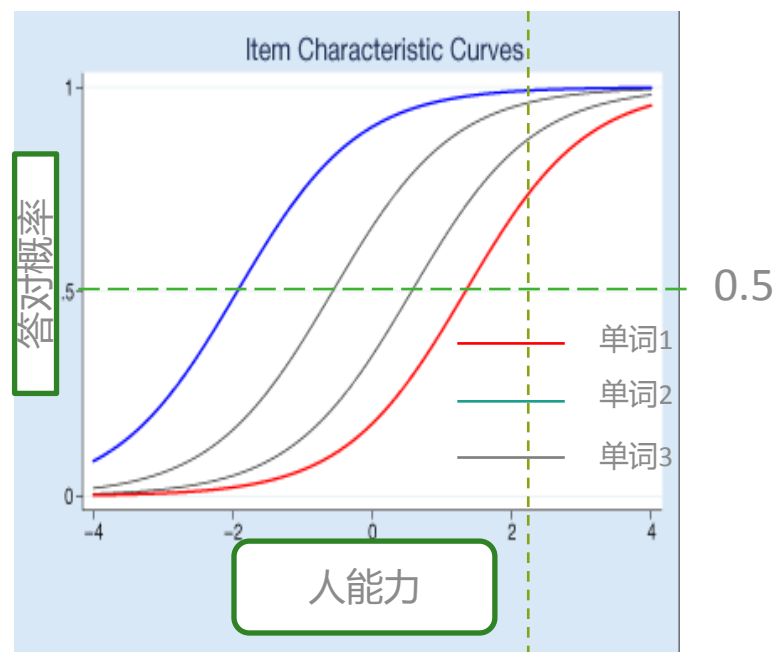
$$P(\text{答错}|\text{词难度, 人能力}) = 1 - \frac{1}{1+e^{-(\text{人能力}-\text{词难度})}}$$



如何选下一个词？

- 选难度和当前能力估计最接近的，且用户没有回答过的词
 - 对于该词， $P(\text{答对}|\text{词难度}, \text{人能力}) \sim 0.5$, 即最不确定
 - 从而避免两种可能浪费时间的词(肯定会做，肯定不做)

$$P(\text{答对}|\text{词难度}, \text{人能力}) = \frac{1}{1 + e^{-(\text{人能力} - \text{词难度})}}$$



选红色的单词1

使用IRT根据用户数据进行难度校准

y	take	explai n	huge	read	talk	term
用户1	1	0	1			
用户2	0	0		1		
用户3	1		1		1	
用户4		1		1	1	1
用户5	0	1	1			1
用户6			1	0	0	1

$$\max_{u,d} \prod_{(i,m) \in S} \left(\frac{1}{1 + \exp(-(u_i - d_m))} \right)^{y_{im}} \left(1 - \frac{1}{1 + \exp(-(u_i - d_m))} \right)^{1-y_{im}}$$

- u : 用户难度
- d : 单词难度
- 此问题具有全局最优解
- 在数据量很大的时候, 可以采用随机梯度下降的方法优化该代价函数
- 计算出来的单词难度和冷启动时设置的大体趋势一样但是有区别
- 细节区别往往由于词和题目之间的存在gap
 - 简单的词出的选项混淆太强, 导致学生选错
 - 难的词由于例句原因, 可以猜出答案

为什么不对每个词的正确率求平均得到难度值?

- 因为在词汇量测试里, 每个用户所做的词都不一样
- 总共1000个词, 每个用户只做了少于40道题目



可以改进的方向

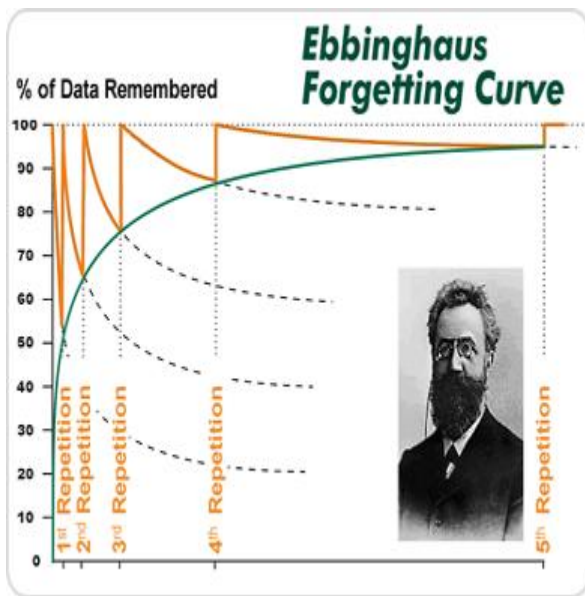
- IRT模型的改进，如何更好的建模用户答题(用户-词)矩阵
 - 混合IRT (如果存在多个学习路径)
 - 深度神经网络
 - 可以很好的解释现有的数据
- 推题策略的改进：
 - MDP?



记忆模型

记忆模型

艾宾浩斯记忆曲线



- 个性化复习策略
- 非个性化模型

Supermemo 2004



anki



机器学习模型

- Act-R
- IRT
- MCM
- Duolingo
- ...

- 个性化复习策略
- 个性化的模型

c.f.

- Predicting and Improving Memory Retention: Psychological Theory Matters in the Big Data Era
- [ACL16] A Trainable Spaced Repetition Model for Language Learning

学习, 成为更好的自己

记忆模型

艾宾浩斯记忆曲线 → 间隔效应 → Act-R → MCM

$$P_r(\text{recall}) = m(1 + ht)^{-f}$$

概率随着时间指数衰减:

其中 m, h, f 是常数, 分别解释为初始学习的程度 ($0 < m < 1$), 时间的缩放因子 ($h > 0$), 以及记忆的衰减指数 ($f > 0$)

艾宾浩斯记忆曲线: $y = 1 - 0.56x^{0.06}$

$$\text{Optimal ISI} = 0.097RI^{0.812}$$

多次学习对记忆的影响: 间

隔效应 (Spacing effect)

两次学习的间隔记作 ISI

(intersession interval), 第

二次学习和最后的测验的时

间记作 RI (retention interval)

Act-R :

$$m_n = \ln \left(\sum_{k=1}^n b_k t_k^{-d_k} \right) + \beta$$

ACT-R 假设每次学习会有不同的记忆概率轨迹, 而且记忆概率随着时间的增长成幂函数衰减: t_k, d_k 指的是第 k 条轨迹的记忆时间和衰减指数, β 是和学生或者记忆事物有关的影响记忆强度的参数。 b_k 指的是每条记忆轨迹的显著性, 这个数越大表示一次学习的效果越好。

$$d_k(m_{k-1}) = ce^{m_{k-1}} + \alpha$$

轨迹的衰减和学习发生的时间点有关: 这里 c 和 α 是常数, 如果第 k 次学习和前一次的间隔比较短, 会导致当前的一条衰减的较快。

$$P_r(\text{recall}) = 1 / (1 + e^{\frac{\tau - m}{s}})$$

回忆的概率和记忆活性 m 单调相关: 其中 τ 和 s 是相应的参数。整个模型有 6 个自由的参数。

记忆模型

MCM提出了一个假设，每次新的学习学到的东西是分别存储在不同的轨迹中，而且会按照不同的速率衰减。虽然每条迹会指数衰减，这些轨迹的和随着时间的衰减是一个幂函数，举例来说，第*i*条轨迹， x_i 的衰减如下面公式所示：

$$x_i(t + \Delta t) = x_i(t) \exp(-\Delta t / \tau_i)$$

其中是衰减时间常数，而且后续的轨迹具有比较小的衰减时间常数，轨迹1-k使用了一个加权平均，最后合成了一个总的轨迹强度。

$$s_k = \frac{1}{\Gamma_k} \sum_{i=1}^k \gamma_i x_i$$

其中 $\Gamma_k = \sum_{i=1}^k \gamma_i$ 。 γ_i 是一个权重因子，代表了第*i*条轨迹的贡献，在总共*k*条轨迹中，记忆的概率是其中的最小值：

$$P_r(\text{recall}) = \min(1, s_k)$$

间隔效应发生的主要原因是轨迹的更新规则（Staddon et al., 2002）。一条轨迹只有在其它轨迹无法保持对材料的记忆的时候才会更新。这个规则影响了信息在不同发生频率和不同环境下的记忆效果。当一个材料被学习的时候，第*i*条轨迹贡献的上升和前面轨迹的总强度负相关：

$$\Delta x_i = \epsilon(1 - s_i)$$

其中是 ϵ 步长。

个性化策略 vs 个性化模型

学生历史记录

time	word	correct
1/1/2012 13:00	depress	1
1/1/2012 13:10	depress	0
1/1/2012 13:20	depress	1
1/2/2012 13:00	depress	1
1/2/2012 13:40	depress	0
1/3/2012 13:50	depress	0
1/3/2012 14:00	depress	1

- 每个学生的历史记录都不一样

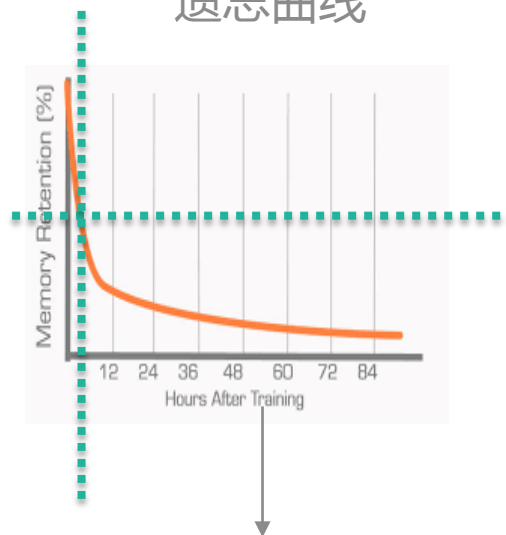
+

记忆模型

输出: 遗忘曲线

=

遗忘曲线



下次复习时间

time	word	correct
1/7/2012 13:00	depress	1

- 非个性化记忆模型
 - 艾宾浩斯曲线
 - Supermemo
- 个性化记忆模型
 - 每个学生的记忆能力不一样
 - 每个学生记忆模型也不相同

- 每个学生的历史记录都不一样
- 同样的模型 → 不同的复习策略
- 不同的模型 → 不同的复习策略

当前词场采用的复习机制

1

单词1历史记录
单词2历史记录
单词3历史记录
单词4历史记录

+

记忆模型

=

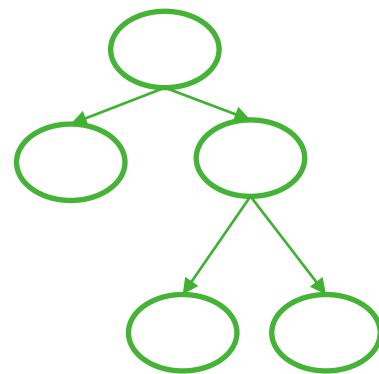
下次复习时间

下次复习时间	单词
1/7/2012 13:00	单词1
1/7/2012 13:00	单词2
1/8/2012 13:00	单词3
1/9/2012 13:00	单词4

2

对于下次复习时间接近的单词，根据以下特征的组合进行重新排序

- 上次复习时间
- 上次复习是否作对
- 上次反应时间
- 历史做错次数
- 历史作对次数
- ...



学习，成为更好的自己

正在研究的复习机制

- 个性化记忆模型

通过分析大量数据, 考察记忆和以下特征之间的关系

- 背词间隔
- 历史准确率
- 一次学习的量 (疲劳控制)
- 答题反应时间
- 以前背过的单词
- ...

个性化记忆模型

用户1记忆模型

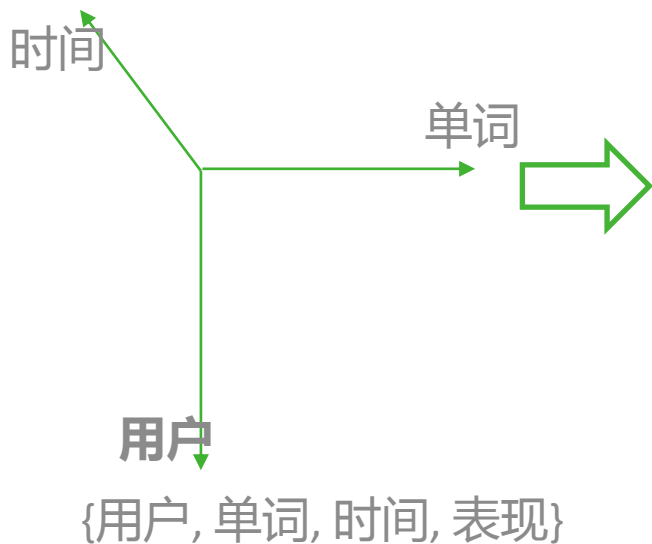
用户2记忆模型

用户3记忆模型

用户4记忆模型

...

用户背词数据



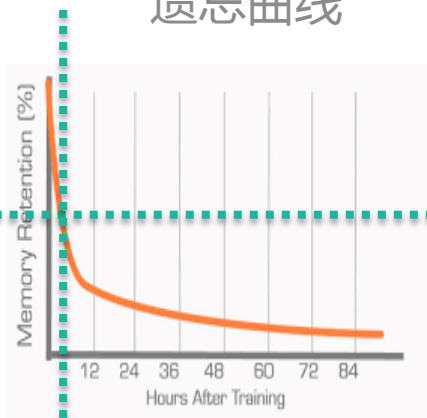
采用的技术

- Logistic Regression
- Multiple Task Learning/协同过滤

正在研究的复习机制

- 优化的复习/学习策略

遗忘曲线



下次复习时间

time	word	correct
1/7/2012 13:00	depress	1

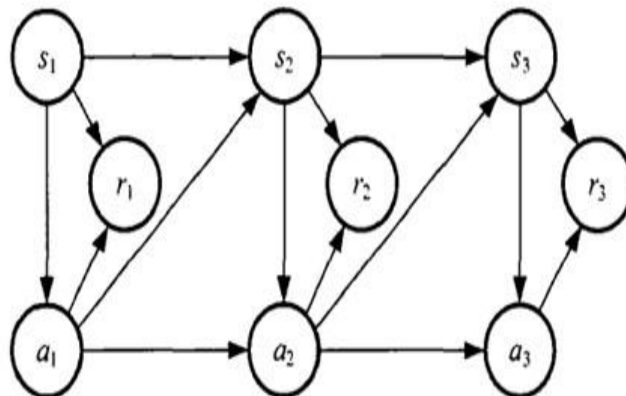
智能和个性化的
优化策略

在特定的约束条件下，寻找最优的复习方案

- 学生只愿意每天学习/复习少量的词，如何调整词的顺序？

采用的技术

- MDP (给定记忆模型)
- Reinforcement Learning (Model Free)



更好的自己



知识追踪DKT

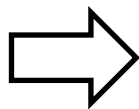
1. 知识追踪的意义和现状
2. 贝叶斯知识追踪——简单模型
3. 深度知识追踪基础——循环神经网络
4. 深度知识追踪模型
5. 深度知识追踪实践
6. 总结和展望

什么是知识追踪？

知识追踪指的是使用计算机模型，根据学生和习题之间的交互，
来自动衡量学生的知识水平。



传统教育



互联网教育

知识追踪特点：

1. 自动化
2. 个性化



知识追踪的现状和挑战

鉴于知识追踪方法对于教育上的重大意义，业界已经有很多相关的模型应用于知识追踪上。但是知识追踪还有很多问题有待解决。

1. 人脑的复杂性：不同的场景需要不同的记忆模型描述
2. 知识的复杂性：不同知识点之间有复杂的内部联系，如何描述这些联系



知识空间描述

如何描述人类掌握知识的程度？

使用知识空间，知识空间是一些概念的集合，一个人可以掌握这些概念的一部分，即构成这个人掌握的知识的集合。这个集合和外表现出来的和习题的交互有关。通过这个交互的过程和结果来描述一个人掌握的知识。

https://en.wikipedia.org/wiki/Knowledge_space

贝叶斯知识追踪模型简介

[Michael V. Yudelson et. al.](#)

贝叶斯知识追踪模型的核心是隐马尔科夫模型(Hidden Markov Model, HMM)，把学生的知识看做隐变量，通过观察学生在习题上的表现（答对或答错）来达到推测学生对知识掌握程度的目的。

为什么叫贝叶斯——通过条件概率，从结果（习题的表现）出发推断原因（知识水平），即贝叶斯推断。

模型参数

$$p(L_1)_u^k = p(L_0)^k,$$

→ 初始知识水平

$$p(L_{t+1}|obs = correct)_u^k = \frac{p(L_t)_u^k \cdot (1 - p(S)^k)}{p(L_t)_u^k \cdot (1 - p(S)^k) + (1 - p(L_t)_u^k) \cdot p(G)^k}$$

→ 做题失误的概率

→ 猜对题目的概率

$$p(L_{t+1}|obs = wrong)_u^k = \frac{p(L_t)_u^k \cdot p(S)^k}{p(L_t)_u^k \cdot p(S)^k + (1 - p(L_t)_u^k) \cdot (1 - p(G)^k)},$$

$$p(L_{t+1})_u^k = p(L_{t+1}|obs)_u^k + (1 - p(L_{t+1}|obs)_u^k) \cdot p(T)^k$$

→ 从不会到会的概率

$$p(C_{t+1})_u^k = p(L_t)_u^k \cdot (1 - p(S)^k) + (1 - p(L_t)_u^k) \cdot p(G)^k$$



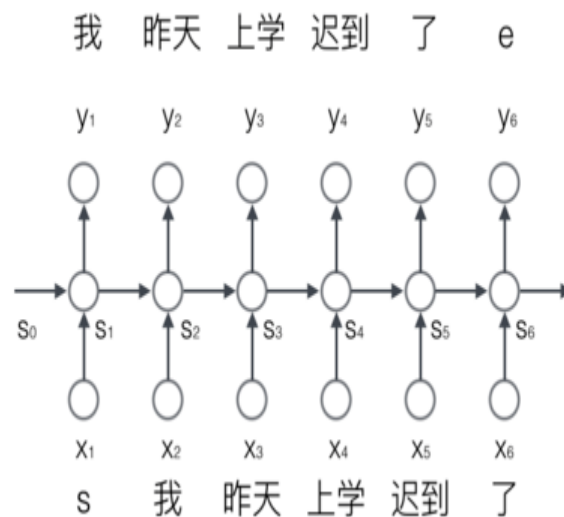
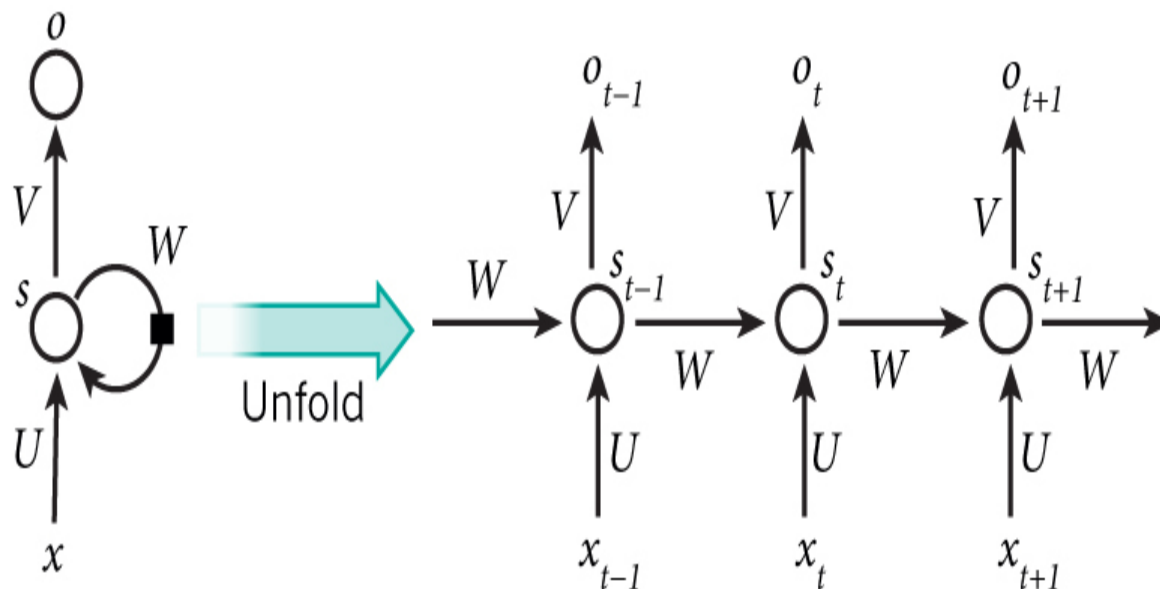
贝叶斯知识追踪模型的缺点

1. 问题和知识的映射不能严格一对一
2. 模型参数太少，难以处理比较复杂的知识体系
3. HMM假设当前题目只和上一次的状态有联系（难以模拟较长序列）

循环神经网络简介

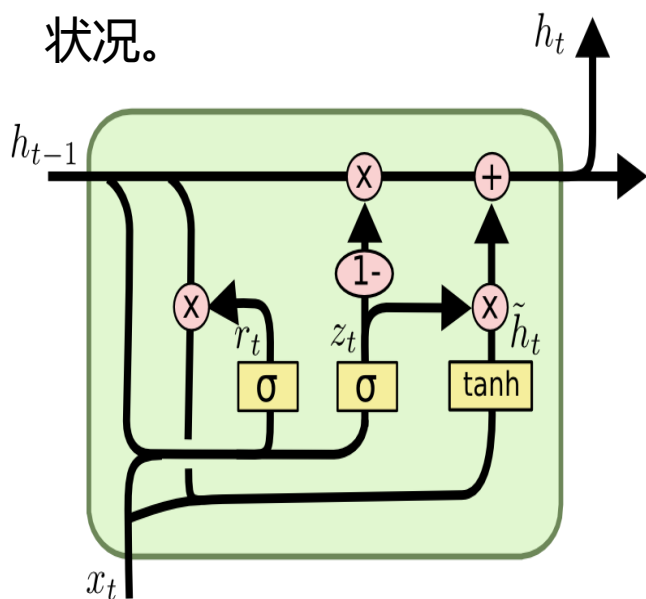
输入按照一定顺序（时间或空间）排列的数据，输出下一步

循环神经网络



长短期记忆网络 (LSTM)

一般RNN容易碰到的一个问题是梯度消失/爆炸，这个对于长序列的训练很不利。长短期记忆网络(Long Short Term Memory Network, LSTM)是RNN的一种特殊变体，能够很好的应对这种状况。



$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

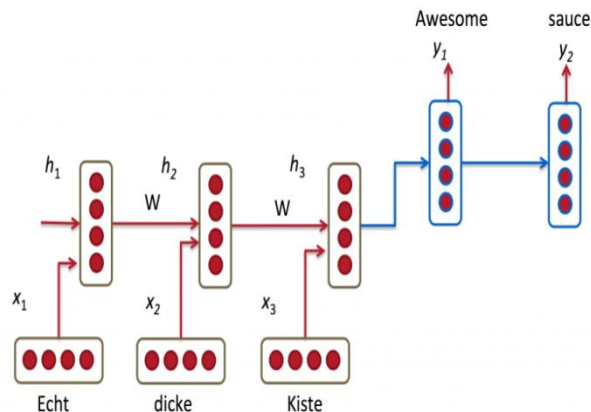
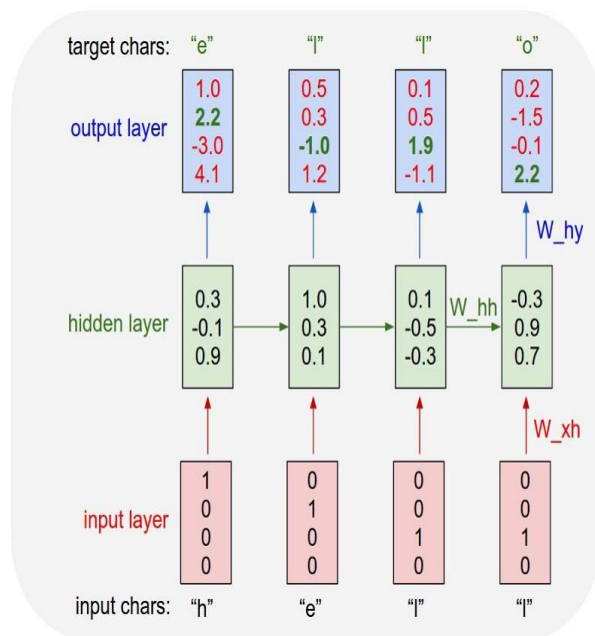
$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

RNN在深度学习中的应用实例

1. 语言模型，文本生成
2. 机器翻译
3. 语音识别

...



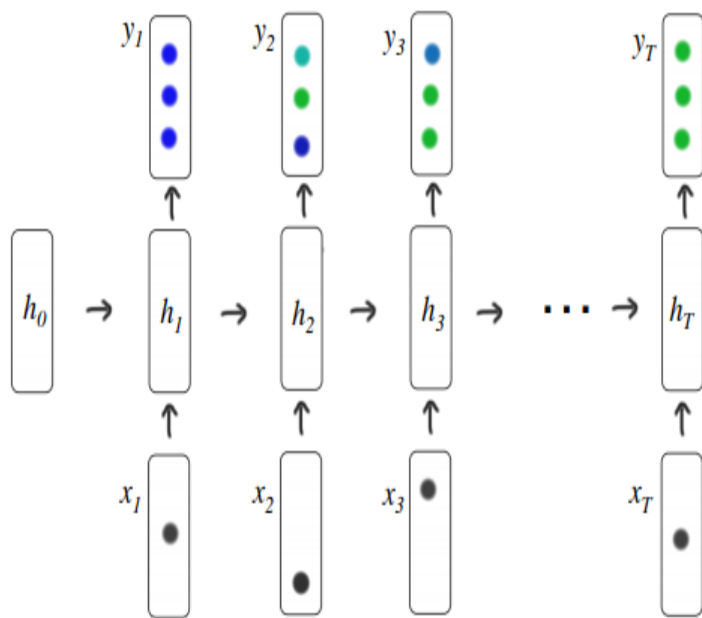
深度知识追踪模型描述

$x_1, x_2, x_3 \dots$ 代表问题是否做对的编码

$y_1, y_2, y_3 \dots$ 代表对于所有问题做对的概率

[Chris Piech et. al.](#)

$h_0, h_1, h_2 \dots$ 代表学生隐含知识水平

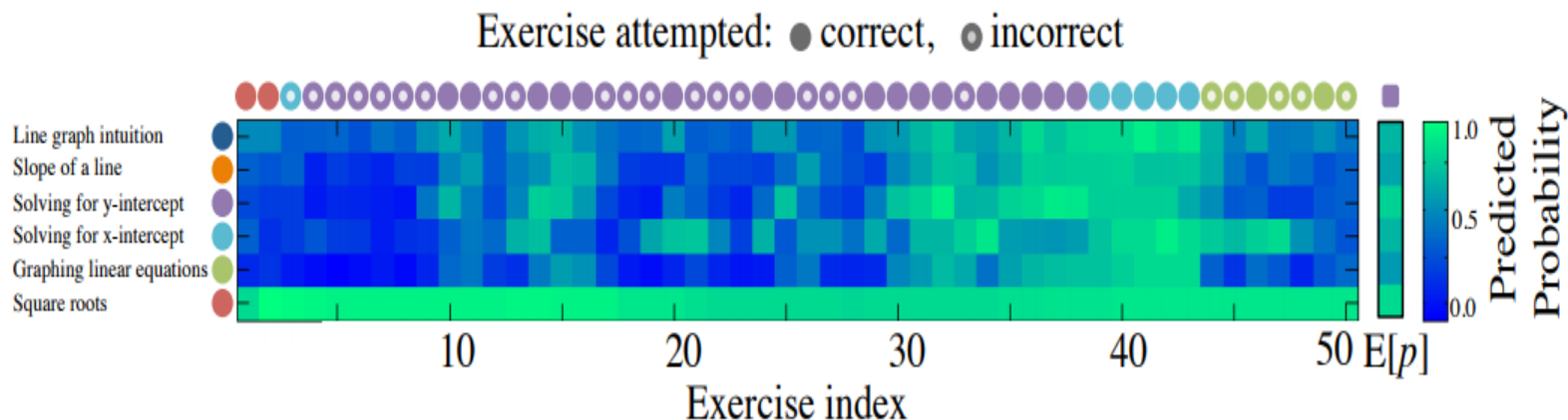


$$\mathbf{h}_t = \tanh(\mathbf{W}_{hx}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h),$$

$$\mathbf{y}_t = \sigma(\mathbf{W}_{yh}\mathbf{h}_t + \mathbf{b}_y),$$

深度知识追踪模型简单实例

1. 问题的类型和回答是否正确（输入）
2. 问题做对的概率（输出）





深度知识追踪优点

1. 能够反映长时间的知识关系（知识遗忘等）
2. 能够对复杂的知识点之间的联系进行建模
3. 能够反应连续的知识水平的变化（输出的是一个概率值）

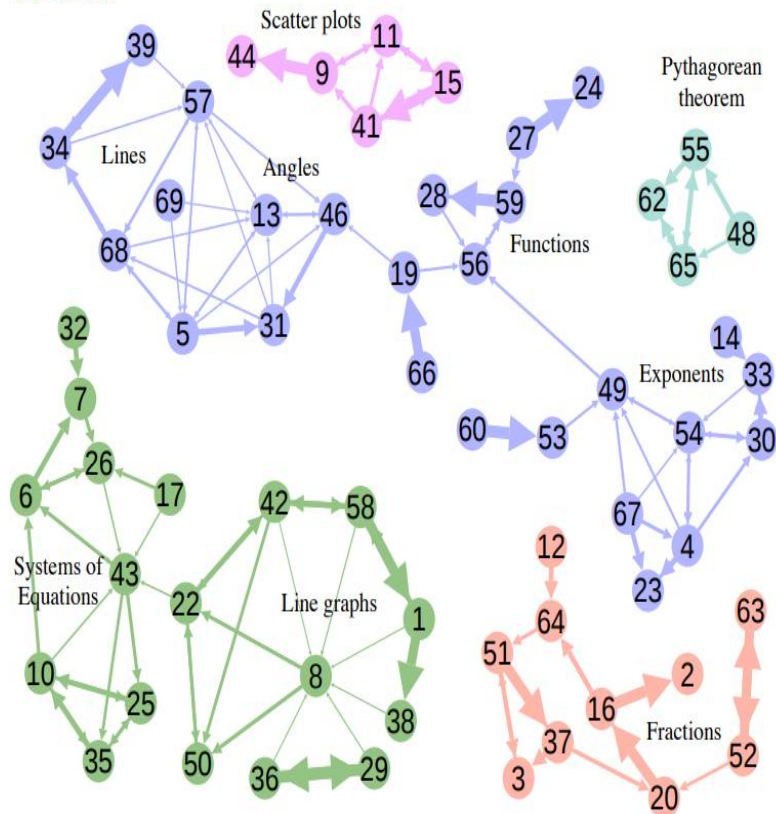
深度知识追踪描述知识点相互依赖

$$J_{ij} = \frac{y(j|i)}{\sum_k y(j|k)}$$

问题i和j的关联的权重可以通过使用深度知识追踪模型给出, 其中 J_{ij} 指的是问题i到问题j的关联权重, $y(j, i)$ 是在问题i答对的前提下问题j答对的概率。

可汗学院八年级数据集知识点关系

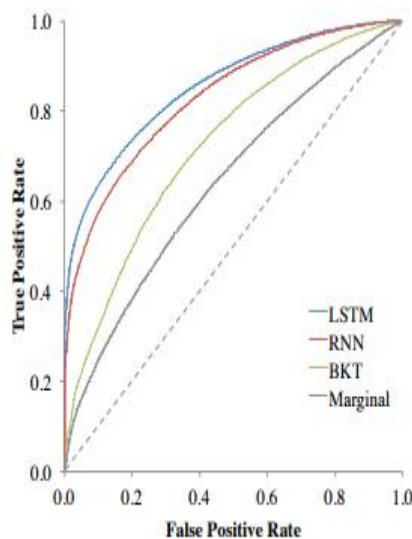
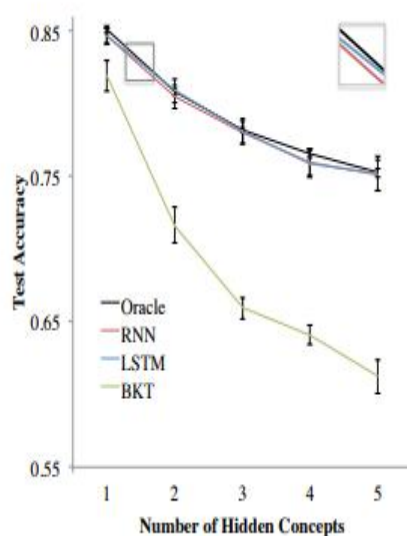
Khan Data



1 Linear function intercepts	24 Interpreting function graphs	47 Constructing inconsistent system
2 Recognizing irrational numbers	25 Systems of equations w. Elim. 0	48 Pythagorean theorem proofs
3 Linear equations 3	26 Solutions to systems of equations	49 Scientific notation intuition
4 Multiplication in scientific notation	27 Views of a function	50 Line graph intuition
5 Parallel lines 2	28 Recog func 2	51 Multistep equations w. distribution
6 Systems of equations	29 Graphing proportional relationships	52 Fractions as repeating decimals
7 Equations word problems	30 Exponent rules	53 Cube roots
8 Slope of a line	31 Angles 2	54 Scientific notation
9 Linear models of bivariate data	32 Understand equations word problems	55 Pythagorean theorem 2
10 Systems of equations with elimination	33 Exponents 2	56 Functions 1
11 Plotting the line of best fit	34 Segment addition	57 Vertical angles 2
12 Integer sums	35 Systems of equations w. substitution	58 Solving for the x intercept
13 Congruent angles	36 Comparing proportional relationships	59 Recognizing functions
14 Exponents 1	37 Solutions to linear equations	60 Square roots
15 Interpreting scatter plots	38 Finding intercepts of linear functions	61 Slope and triangle similarity
16 Repeating decimals to fractions 2	39 Midpoint of a segment	62 Distance formula
17 Graphical solutions to systems	40 Volume word problems	63 Converting decimals to fractions 2
18 Linear non linear functions	41 Constructing scatter plots	64 Age word problems
19 Interpreting features of linear functions	42 Solving for the y intercept	65 Pythagorean theorem 1
20 Repeating decimals to fractions 1	43 Graphing systems of equations	66 Comparing features of functions 0
21 Constructing linear functions	44 Frequencies of bivariate data	67 Orders of magnitude
22 Graphing linear equations	45 Comparing features of functions 1	68 Angle addition postulate
23 Computing in scientific notation	46 Angles 1	69 Parallel lines 1

在测试数据集上的表现

Dataset	Overview			AUC			
	Students	Exercise Tags	Answers	Marginal	BKT	BKT*	DKT
Simulated-5	4,000	50	200 K	0.64	0.54	-	0.75
Khan Math	47,495	69	1,435 K	0.63	0.68	-	0.85
Assistments	15,931	124	526 K	0.62	0.67	0.69	0.86



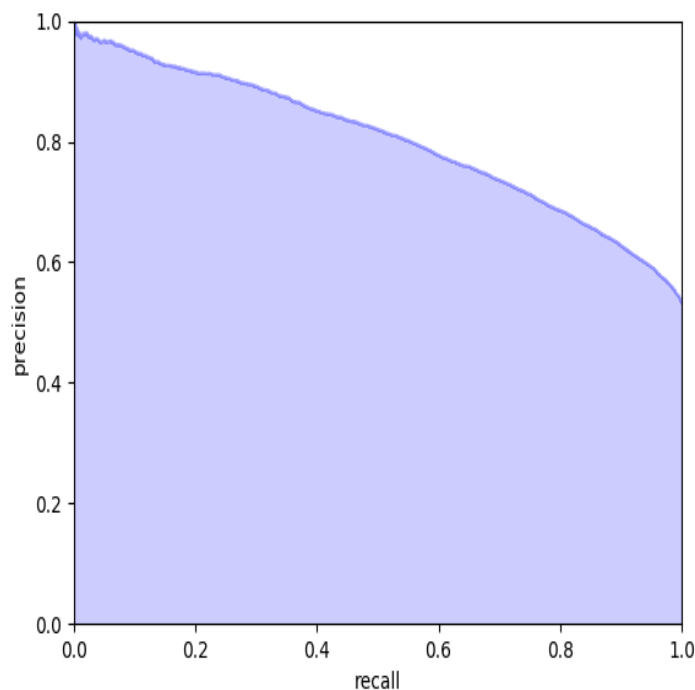
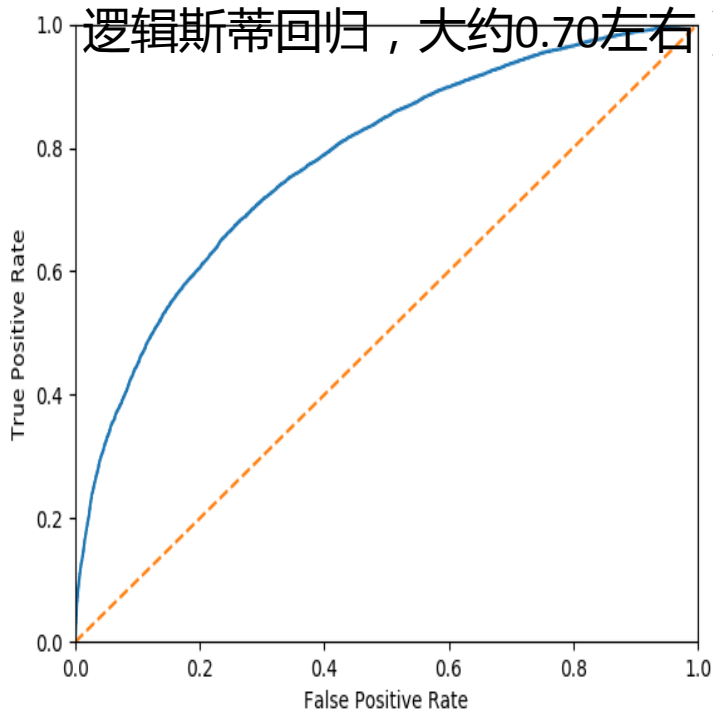
在沪江英语词汇量测试数据集上的

使用10w次沪江英语的词汇量测试的结果作为训练集，1w次

测试集，结果如下：

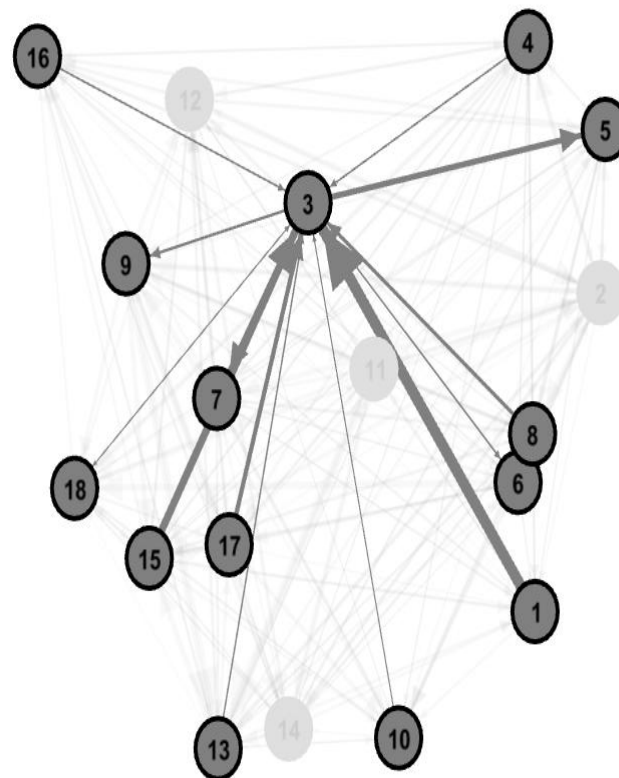
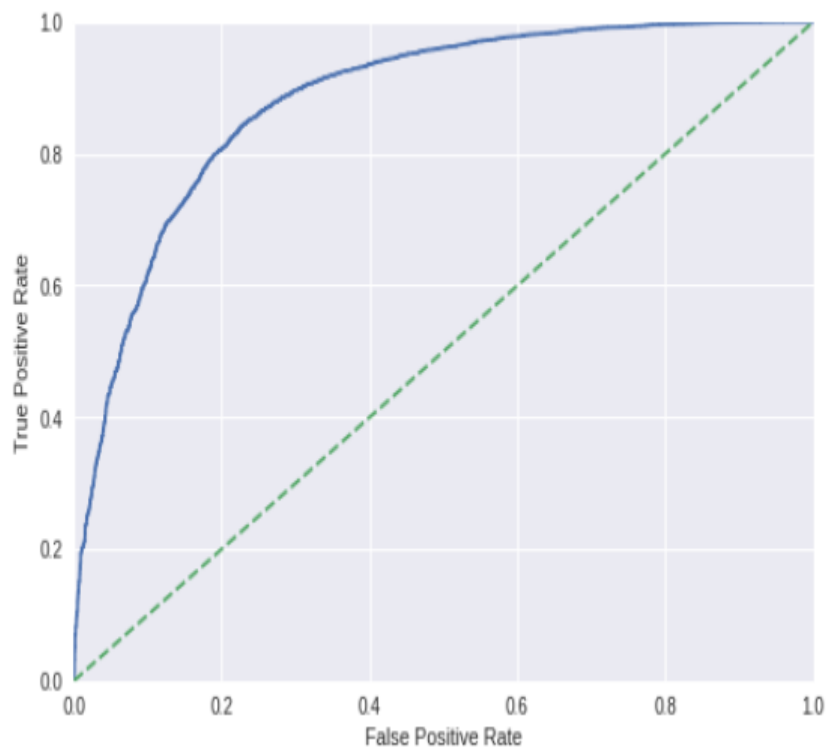
AUC: 0.78, 相比于其他词汇量测试的模型（如IRT模型的简单

逻辑斯蒂回归，大约0.70左右）：



在中考模拟考数学试题数据集上的

AUC score: 0.881





深度知识追踪的未来

1. 人工智能辅助教育的一种手段
2. 自动发现知识点间的相互联系
3. 智能设计课程，减少教育工作量

Q&A

