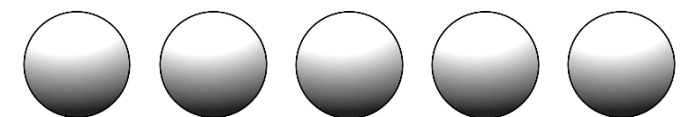


基于循环神经网络的自动语法纠错研究

畢蕊斌

主要内容

- ⊕ 研究背景
- ⊕ 自动语法纠错研究简述
- ⊕ iWrite自动语法纠错的总体原则
- ⊕ 双引擎自动语法纠错方法简介
- ⊕ 将来工作



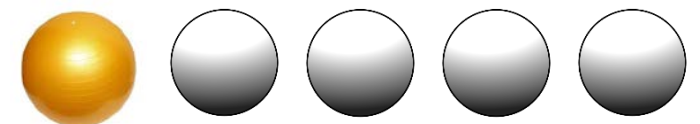
研究背景

⊕ iWrite 研究

- ⊕ iWrite的主要评价维度
- ⊕ iWrite与其他产品的比较
 - ⊕ 内容分析
 - ⊕ 语法纠错

⊕ 前期其他相关研究

- ⊕ 链语法
- ⊕ 错误模板



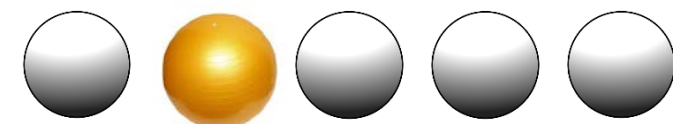
自动语法纠错研究简述

⊕ 主要方法

- ⊕ 基于规则的方法
- ⊕ 基于统计的方法
- ⊕ 基于神经网络的方法

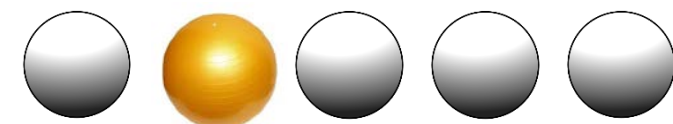
⊕ 三种方法的优势与劣势

- ⊕ 对数据的依赖
- ⊕ 准确率和召回率
- ⊕ 对错误的分类
- ⊕ 诊断性反馈



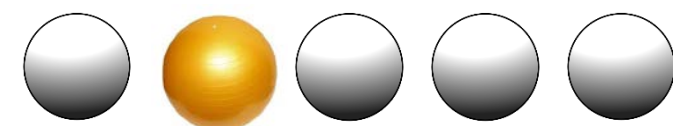
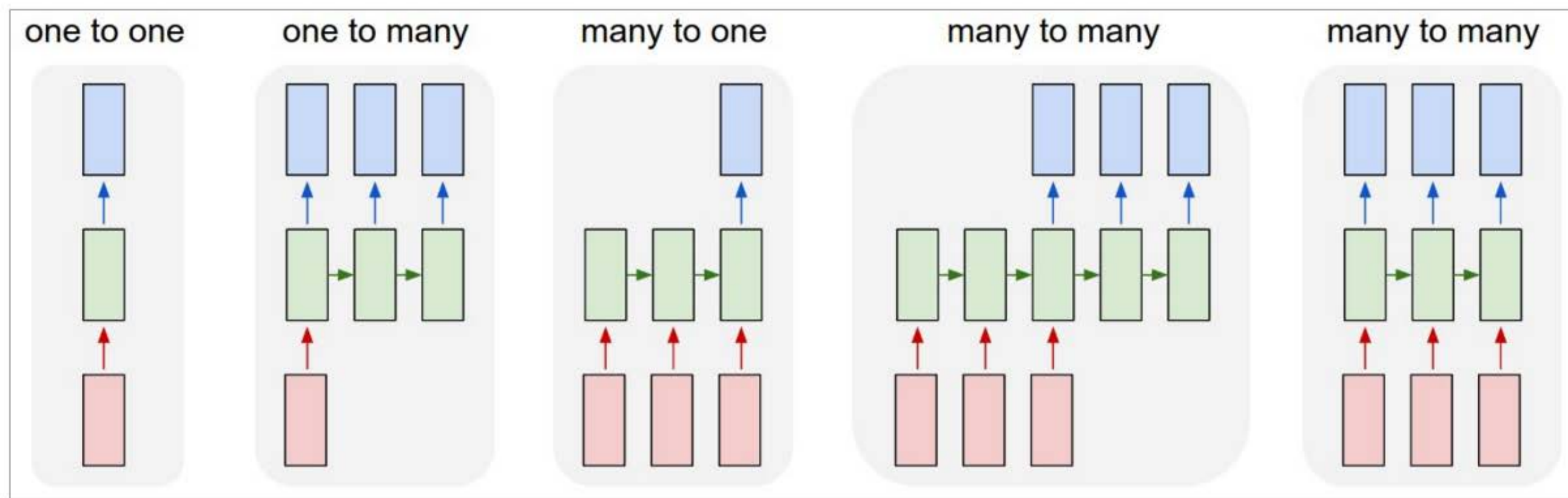
自动语法纠错研究简述

- ⊕ 基于神经网络的方法
 - ⊕ 大数据分析的智能方法
 - ⊕ 依赖大数据
 - ⊕ 标注数据: supervised learning
 - ⊕ 未标注数据: unsupervised learning
 - ⊕ 给出错误的确切类型
 - ⊕ 便于给出合适的反馈语

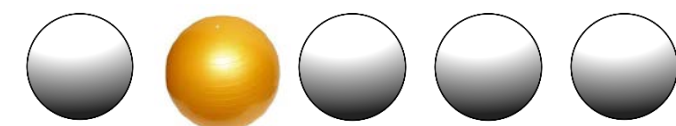
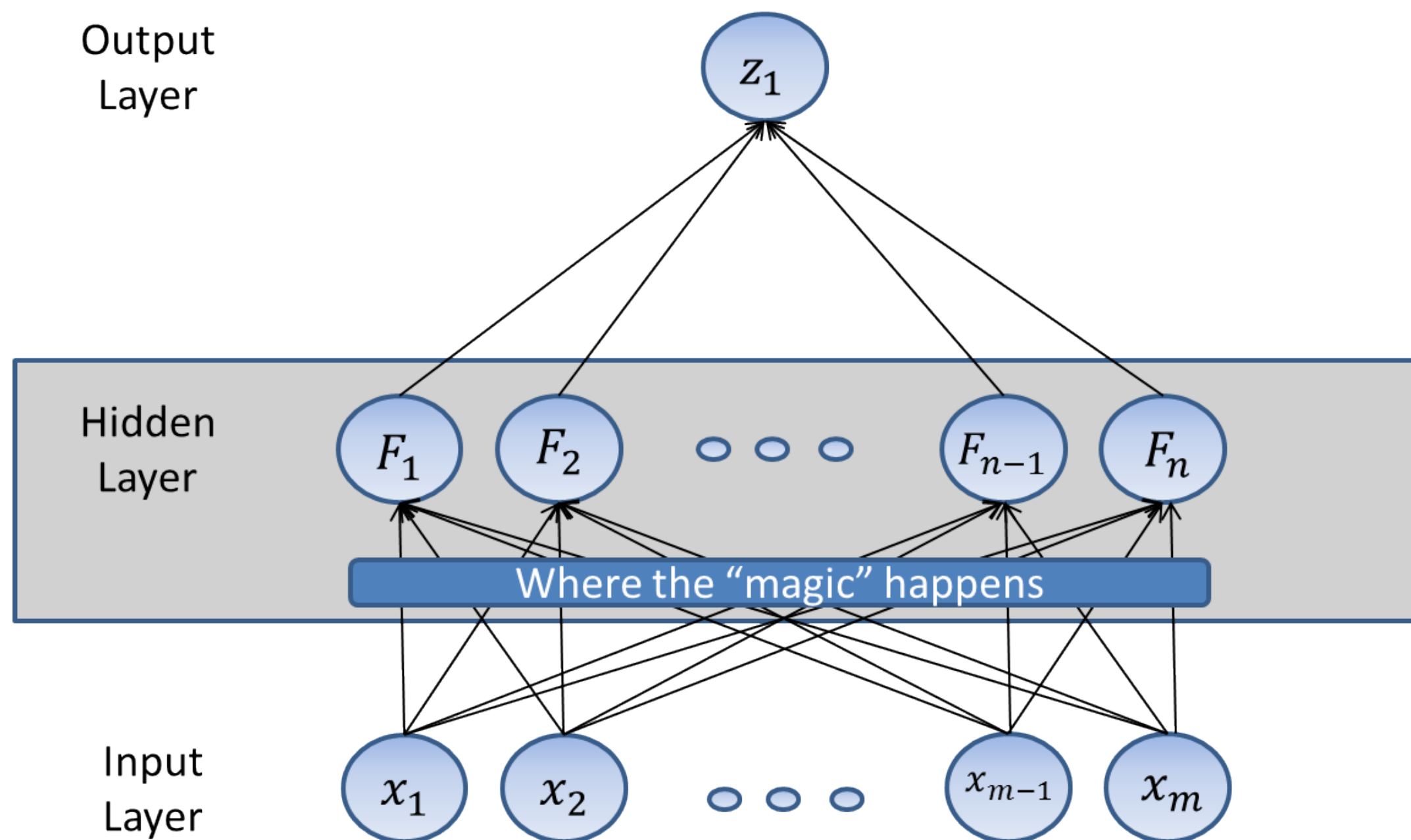


自动语法纠错研究简述

人工神经网络分析中的变量 (input, output)

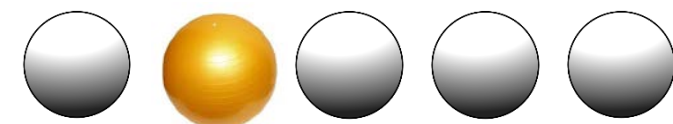


自动语法纠错研究简述

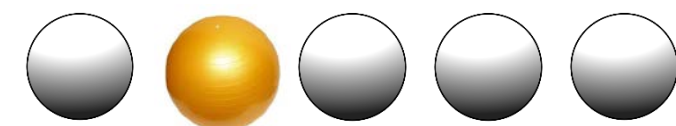
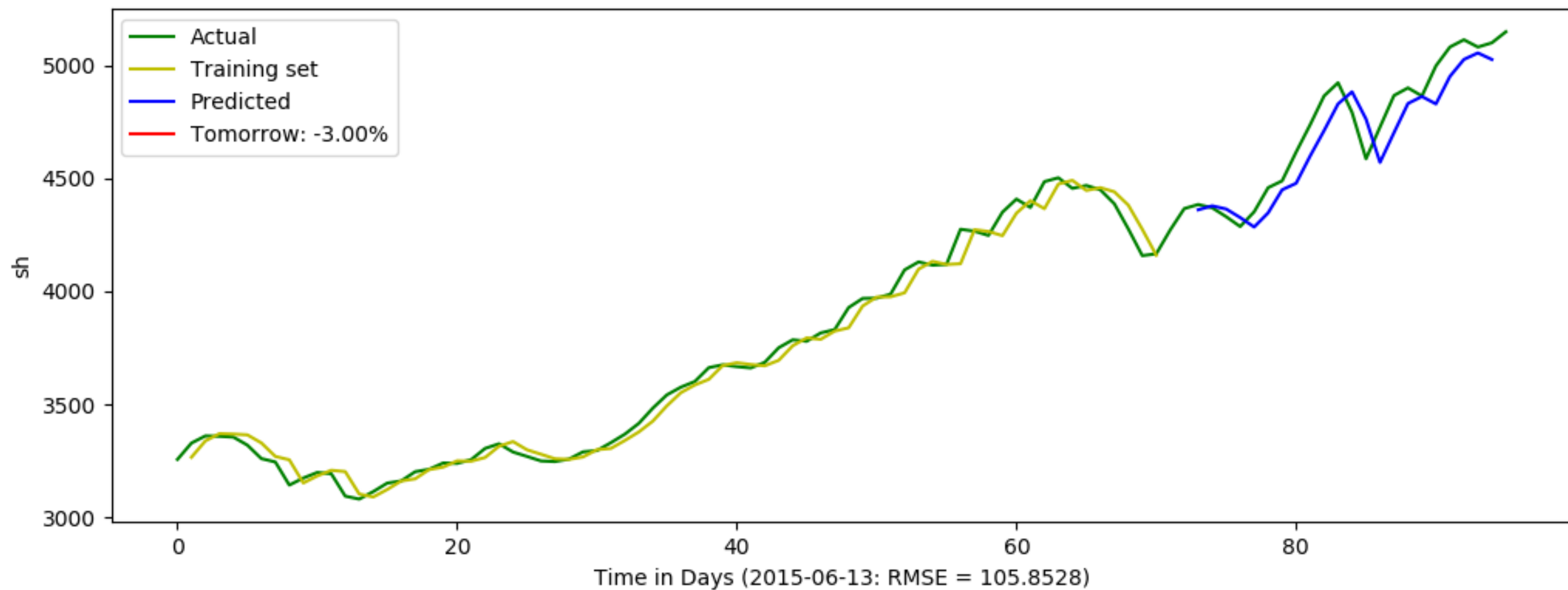


自动语法纠错研究简述

- ⊕ 循环神经网络（Recurrent Neural Network）
 - ⊕ 特别适合于时间序列数据（time series data）
 - ⊕ 特别适合于建构语言模型
 - ⊕ 已经投入大量实际应用



自动语法纠错研究简述



iWrite自动语法纠错的总体原则

- ⊕ 以研究为基础
- ⊕ 准确率与召回率孰轻孰重
- ⊕ 基于规则的方法和基于神经网络的方法相结合
- ⊕ 对学习者的错误进行科学分类
 - ⊕ 词法类
 - ⊕ 句法类
 - ⊕ 搭配类
 - ⊕ 有多少人工就有多少智能
- ⊕ 对不同类别错误的权重
- ⊕ 纠错与反馈相结合



双引擎自动语法纠错方法简介

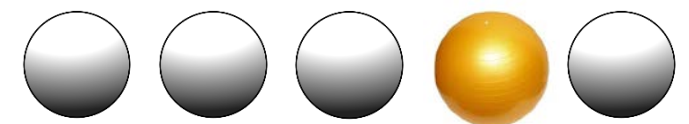
⊕ 双引擎

⊕ 基于规则的方法

- ⊕ 学习者常见错误模板

- ⊕ 面向链语法的语法规则自动学习

⊕ 基于循环神经网络的方法



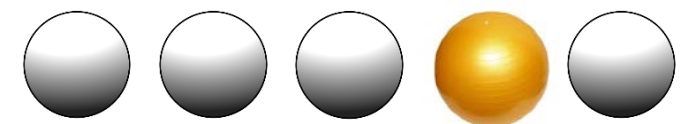
双引擎自动语法纠错方法简介

⊕ 循环神经网络分析中的两大要素

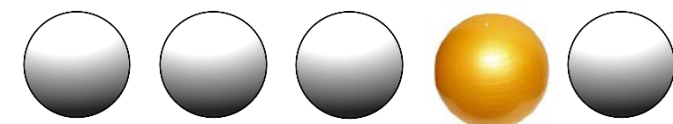
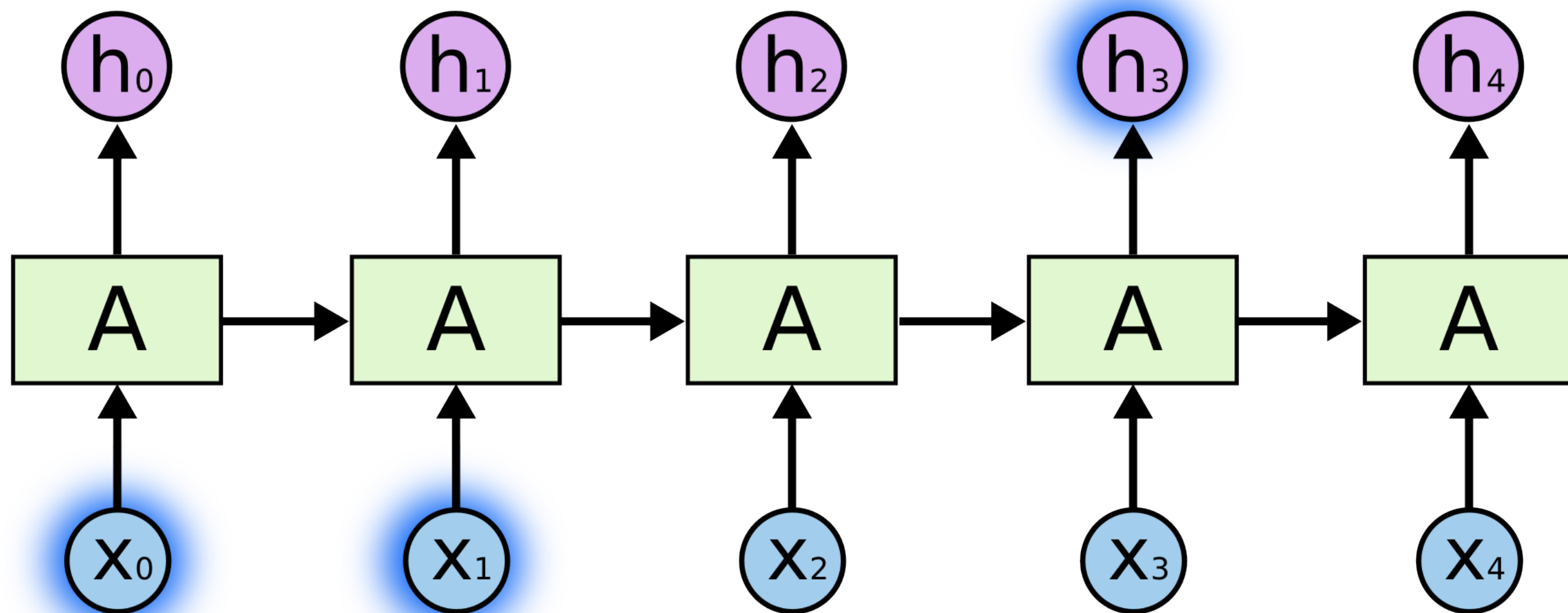
⊕ 数据

- ⊕ 英语本族语者语料
- ⊕ 中国学习者语料

⊕ 技术：深度学习（Deep Learning）



循环神经网络

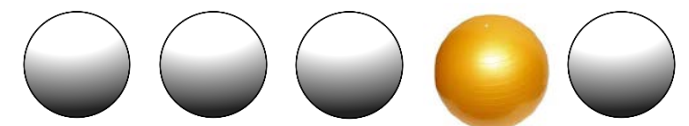




循环神经网络

基于本族语者语料的深度学习：

At the end of the week , a party was organized to
celebrate our partnership with Dell .

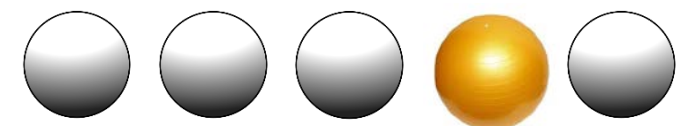




循环神经网络

基于本族语者语料的深度学习：

At the end of the week , a party was organized to
celebrate our partnership with Dell .

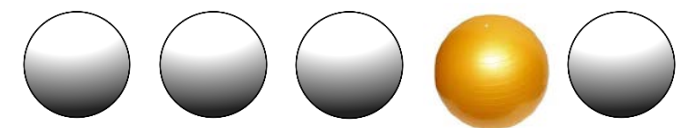




循环神经网络

基于本族语者语料的深度学习：

At the end of the week , a party was organized to
celebrate our partnership with Dell .

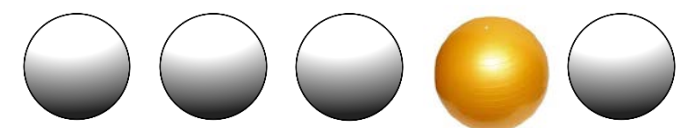




循环神经网络

基于本族语者语料的深度学习：

At the end of the week , a party was organized to
celebrate our partnership with Dell .

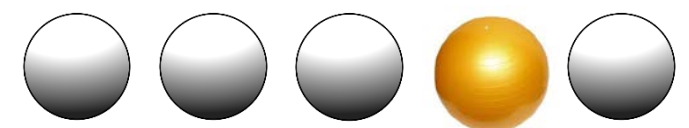




循环神经网络

基于本族语者语料的深度学习：

At the end of the week , a party was organized to
celebrate our partnership with Dell .

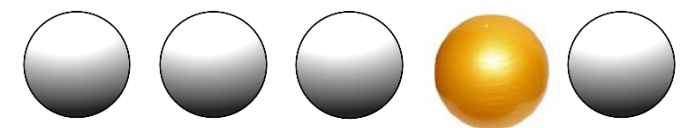




循环神经网络

基于本族语者语料的深度学习：

At the end of the week , a party was organized to
celebrate our partnership with Dell .

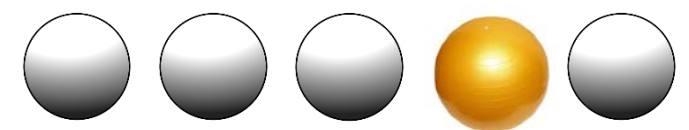




循环神经网络

基于本族语者语料的深度学习：

At the end of the week , a party was organized to
celebrate our partnership with Dell .

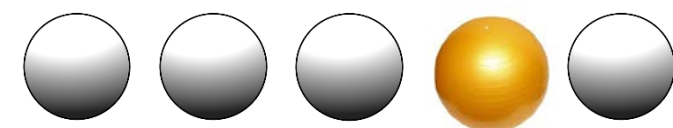




循环神经网络

基于本族语者语料的深度学习：

At the end of the week , a party was organized to
celebrate our partnership with Dell .

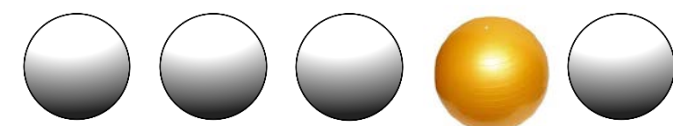




循环神经网络

基于本族语者语料的深度学习：

**At the end of the week , a party was organized to
celebrate our partnership with Dell .**



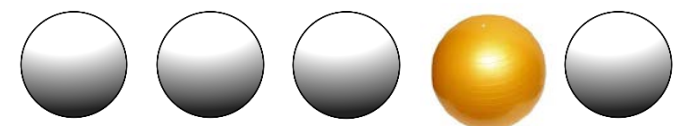
循环神经网络

基于学习者语料的深度学习：

Input: There has a tree in front of the house.

Output: There is a tree in front of the house.

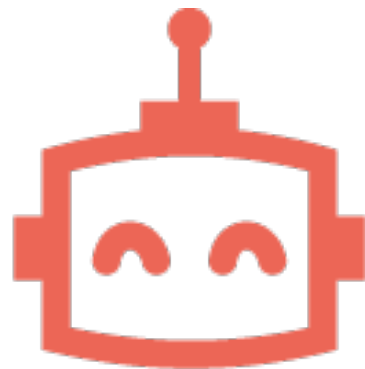
Label: 存现句错误。建议改成there is。



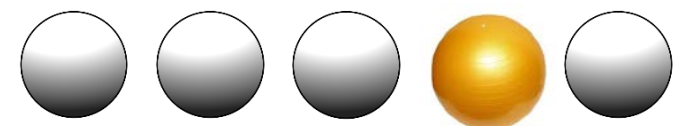
循环神经网络

基于学习者语料的深度学习：

Lewis, G. 2016. Sentence Correction using Recurrent Neural Networks. *Stanford Report*.



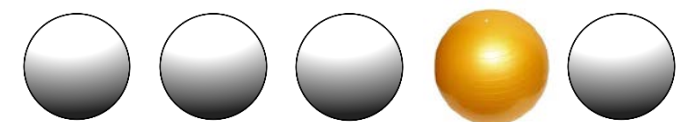
Deep Grammar



循环神经网络

- 两种深度学习方法相结合，最大限度提高纠错的准确率。

	正确率	召回率
iWrite 2.0	66.67%	57.14%
某产品	40%	28.57%





将来工作

- ⊕ **More research**
- ⊕ **More data**

