



江西师范大学
Jiangxi Normal University since 1940

机器译文质量估计的研究进展

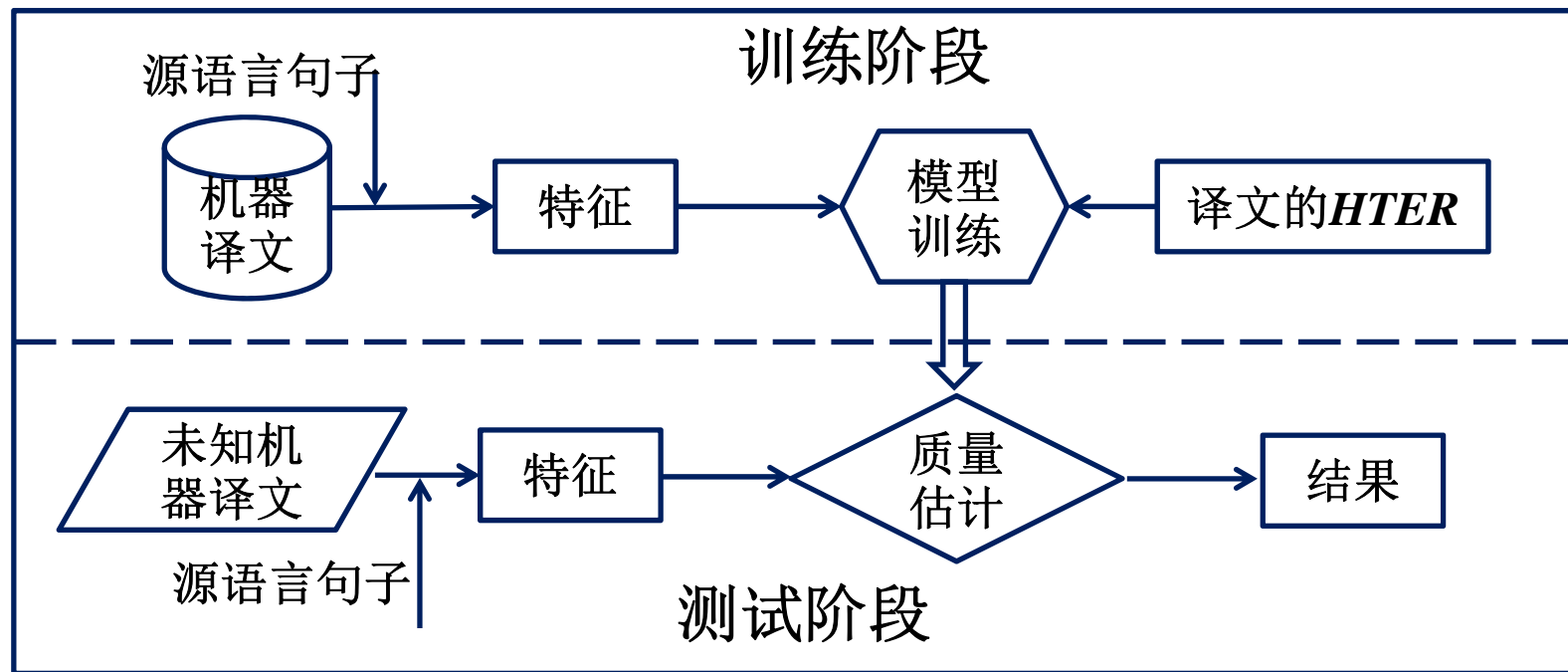
李茂西

mosesli@jxnu.edu.cn



机器译文质量估计

- 利用机器学习算法，在没有人工参考译文的情况下、实时地自动估计机器译文的翻译质量。





机器译文质量估计

- 与传统译文质量自动评价的区别
 - 只能使用源语言句子与机器译文；
 - 使用有监督机器学习的方法，而不是启发式的方法。
 - 以**HTER**为质量估计的基准，自动估计的结果能表示译文后编辑成基本正确状态的工作量；
 - ✓ **HTER** 是指人工把机器译文编辑成正确状态所需的最少编辑次数，包括插入、删除、替换以及移位。
 - ✓ **HTER VS. 流利度/充实度**
准确、量化程度更高，但是人工成本高；
Mechanical Turk: en-de, en-fr, ru-en, \$1/per hypothesis



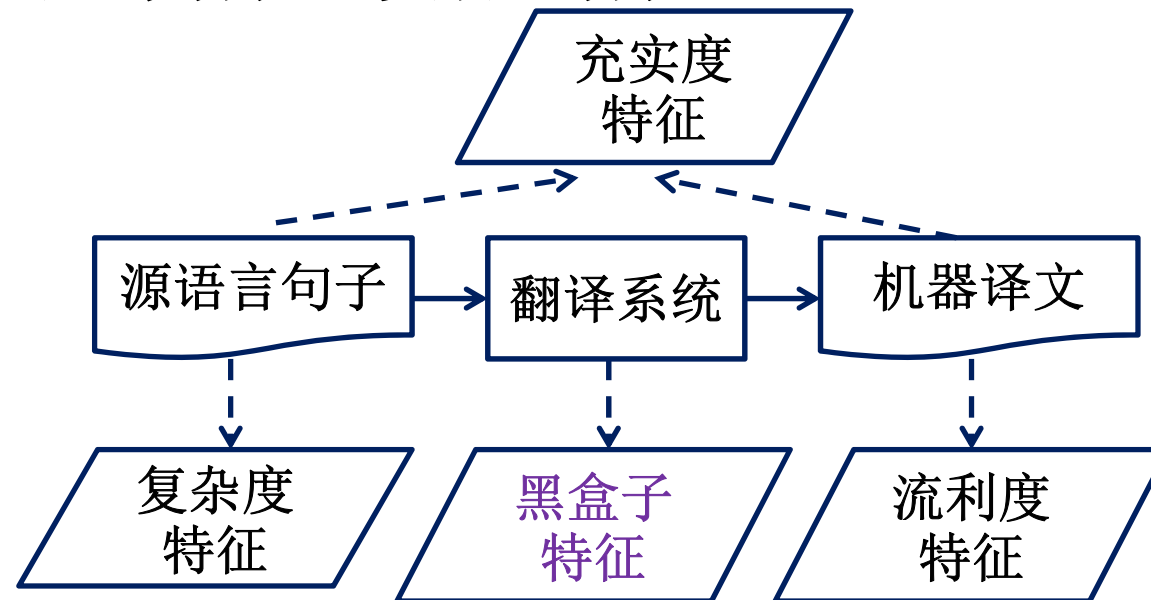
机器译文质量估计

- 与传统译文质量自动评价的区别
 - 只能使用源语言句子与机器译文；
 - 使用有监督机器学习的方法，而不是启发式的方法。
 - 以**HTER**为质量估计的基准，自动估计的结果能表示译文后编辑成基本正确状态的工作量；
- 非常接近于无参考译文的自动评价方法，区别在于：
 - 不是使用流利度/充实度，而是以**HTER**为学习基准；
 - 为了应用于统计翻译系统自训练中，对实时性要求高。



机器译文质量估计

- 方法框架Q_UE_{ST} (<http://www.quest.dcs.shef.ac.uk/>)
 - 把译文质量估计问题看成是一个回归问题，包括两个部分：特征提取和机器学习算法。
 - 提取3类基本特征+1类备选特征





机器译文质量估计

- 方法框架 **QuEST** (<http://www.quest.dcs.shef.ac.uk/>)
 - 把译文质量估计问题看成是一个回归问题，包括两个部分：特征提取和机器学习算法。
 - 学习算法：回归算法
 - ✓ 基于径向基函数核的支持向量机回归算法，
 - ✓ 利用网格搜索进行特征权重学习，
 - ✓ 提供随机拉索(**randomized lasso**)和高斯过程(**Gaussian Process**)算法进行特征的选择。



机器译文质量估计

● 评测

➤ WMT Shared Task - Quality Estimation

✓ 前身是2008-2011年 Metrcis Task,

✓ 2012年在传统任务上增加了 Quality Estimation Task,

✓ 2013年 Quality Estimation Task 完全取代了 Metric Task。

➤ 评测译文种类

✓ WMT'12 English-**Spanish**

✓ WMT'13 German-**English**, English-**Spanish**

✓ WMT'14 English-**Spanish**, English-**German**,
German-**English**, Spanish-English



机器译文质量估计

- 方法框架 **QuEst** (<http://www.quest.dcs.shef.ac.uk/>)
 - 性能
 - ✓ **QuEst** 在 **WMT'12** 评测和 **WMT'13** 评测的 **quality estimation shared task** 中被作为基线系统提供给参加评测的单位使用。
 - ✓ **WMT'12 quality estimation shared task** 中 **5/19**,
 - ✓ **WMT'13 quality estimation shared task** 中 **9/16**.



机器译文质量估计

- 方法框架QUEST (<http://www.quest.dcs.shef.ac.uk/>)

- 性能

- ✓ **WMT'12 quality estimation shared task,
English-Spanish translation direction**

System ID	Spearman Correlation
Best	0.64
QUEST	0.58
Oracle BLEU	0.62
Oracle HTER	0.70



机器译文质量估计

- 方法框架QUEST (<http://www.quest.dcs.shef.ac.uk/>)

- 性能

- ✓ WMT'13 quality estimation shared task,
German-English translation direction

System ID	Kendall's τ with ties ignored
Best	0.31
QUEST	0.08
Oracle BLEU	0.22
Oracle METEOR	0.20



机器译文质量估计

- 方法框架QUEST (<http://www.quest.dcs.shef.ac.uk/>)

- 性能

- ✓ WMT'13 quality estimation shared task,
English-Spanish translation direction

System ID	Kendall's τ with ties ignored
Best	0.23
QUEST	0.03
Oracle BLEU	0.19
Oracle METEOR	0.25



机器译文质量估计

- 在QUEST基础上对特征的研究
 - 词性标注POS (Soricut et al., 2012),
 - 概率上下文无关文法PCFG (Avramidis, 2012),
 - 组合范畴文法CCG (Almaghout and Specia, 2013),
 - 伪参考译文(Soricut et al., 2012)。
- 在QUEST基础上对学习算法的研究
 - M5P 模型学习决策树(Soricut et al., 2012),
 - 基于多项式核的SVM 回归算法(Hardmeier et al., 2012),
 - Logistic 回归(Almaghout and Specia, 2013),
 - 回归树学习进行特征的选择(Rubino et al., 2013)。



总结与展望

● 总结

- 在传统的译文自动评价方法的基础上发展起来，
- 无需人工参考译文的新的译文自动评价方法
- 在有监督机器学习的框架下估计译文质量。



总结与展望

● 展望

- 传统的“特征工程+任务建模”的学习范式是否适合机器译文质量估计任务？
 - ✓ 探索深度学习
- 加强对汉语译文质量估计的研究也是亟待解决的。
 - ✓ 使用*HTER*标注一定规模的汉语译文，
 - ✓ 针对汉语的特点展开研究。



总结与展望

● 展望

- 探索适合统计翻译系统自训练的译文质量估计方法。
 - ✓ 由于译文质量估计时无需人工参考译文，这大大扩展了它的应用范围。
 - ✓ 对于统计机器翻译系统，如果它在解码过程中，可以利用训练好的译文质量估计工具，实时的获取输出译文的质量，我们就可以实现无需开发集的统计翻译系统参数调整方式。
 - ✓ 对译文质量估计要求：集成度高、独立性强、实时性好。



江西师范大学
Jiangxi Normal University since 1940

请大家批评指正，谢谢！

