论文写作小白的成长之路

兰艳艳
中科院计算所 研究员
lanyanyan@ict.ac.cn

起始

Query-Level Stability and Generalization in Learning to Rank

Yanyan Lan* LANYANYAN@AMSS.AC.CN
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, P. R. China.

Tie-Yan Liu TYLIU@MICROSOFT.COM
Microsoft Research Asia, Sigma Center, No. 49, Zhichun Road, Haidian District, Beijing, 100190, P. R. China.

Tao Qin* QINSHITAO99@MAILS.THU.EDU.CN
Department of Electronic Engineering, Tsinghua University, Beijing, 100084, P. R. China.

Zhiming Ma MAZM@AMT.AC.CN
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, P. R. China.

Hang Li HANGLI@MICROSOFT.COM
Microsoft Research Asia, Sigma Center, No. 49, Zhichun Road, Haidian District, Beijing, 100190, P. R. China.

Abstract

This paper is concerned with the generalization ability of learning to rank algorithms for information retrieval (IR). We point out that the key for addressing the learning problem is to look at it from the viewpoint of query. We define a number of new concepts, including query-level loss, query-level risk, and query-level stability. We then analyze the generalization ability of learning to rank algorithms by giving query-level generalization bounds to them using query-level stability as a tool. Such an analysis is very helpful for us to derive more advanced algorithms for IR. We apply the proposed theory to the existing algorithms of Ranking SVM and IRSVM. Experimental results on the two algorithms verify the correctness of the theoretical analysis.

1. Introduction

Recently, learning to rank has gained increasing attention in machine learning and information retrieval (IR). When applied to IR, learning to rank is a task as follows. Given a set of training queries, their associated documents, and the corresponding relevance judgments, a ranking model is created which best represents the relevance of documents with respect to queries. When a user submits a query to the IR system, the trained model assigns a score to each document associated with the query, sorts the documents based on their scores, and presents the top ranked documents to the user. Average ranking accuracy over a large number of queries is usually used to evaluate the effectiveness of a ranking model. Therefore, from the application's perspective, both training and evaluation should be conducted at query level.

Many learning to rank algorithms have been proposed in recent years. Examples include the pointwise ranking algorithms like MCRank (Li et al., 2007), the pairwise ranking algorithms like Ranking SVM (Herbrich et al., 1999) and RankBoost (Freund et al., 2003), and the listwise ranking algorithms like ListNet (Cao et al., 2007). Analysis on the algorithms in the light of statistical learning theory, however, was not sufficient, particularly that on the generalization ability of the proposed algorithms. The pointwise and pairwise approaches transform the ranking problem to classification or regression, and thus existing theory on classification and regression can be applied. However, it deviates from the direction of enhancing ranking accuracy at query level. Furthermore, the listwise approach lacks of analysis on generalization ability.

In this paper, we investigate the generalization ability of learning to rank algorithms, in particular from the viewpoint of query-level training and evaluation.

Appearing in *Proceedings of the 25th International Conference on Machine Learning*, Helsinki, Finland, 2008. Copyright 2008 by the author(s)/owner(s).

*The work was performed when the first and the third authors were interns at Microsoft Research Asia.

导师的感觉

论文分块

- Abstract
- Introduction
- Related Work
- Main Results (Theory&Model&Algorithm)
- Experiments
- Discussion
- Conclusion

论文写作的灵魂：逻辑

3W2H
写作法

Novelty
Significance

Why为什么要做这个工作?
Motivation

→

What做什么?
Contribution

Clarity
Soundness
Reproducibility

How怎么做的?
Theorem&Model&Algorithm

→

How Much效果怎么样?
Experiments&Discussion

Significance

What Then启示、未来还能做什么?
Conclusion

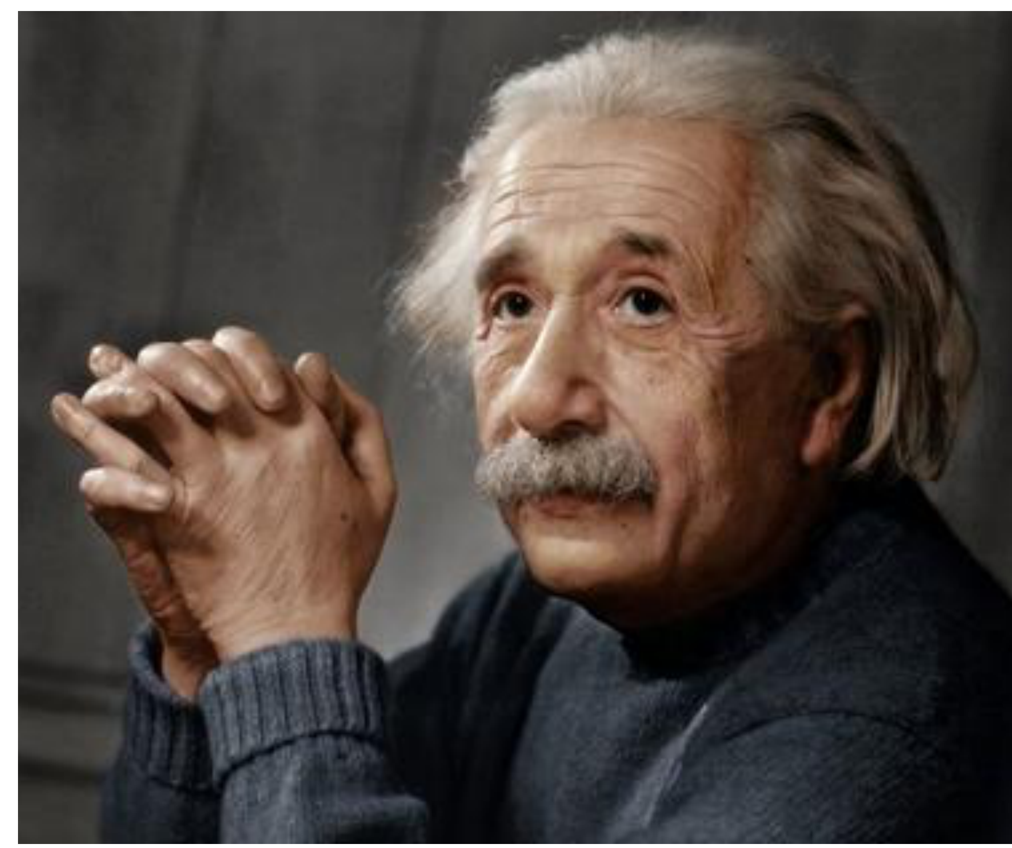
论文写作的灵魂：逻辑

Why为什么要做这个工作? Motivation

- 理论or应用需求
- 以往工作缺陷
- 解决该问题的影响

论文写作的灵魂：逻辑

每句话之间也是要有逻辑的



“Pure Mathematics is, in its way, the poetry of logical ideas.”

多问自己几个为什么?

辩

论文润色

- 善用名言

1 Introduction

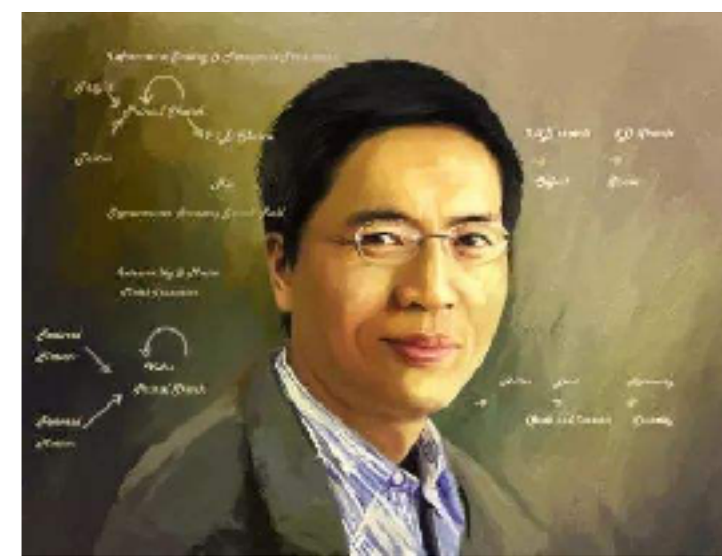
Number is the ruler of forms and ideas, and the cause of gods and demons.

— Pythagoras, c. 300 (Taylor 1818)

1. Introduction

The study of vision must therefore include not only the study of how to extract from images . . . , but also an inquiry into the nature of the *internal representations* by which we *capture* this information and thus make it available as a **basis** for *decisions about our thoughts and actions*.

— David Marr, 1982 [35]



朱松纯 教授
UCLA

论文润色

- 举例子

It has been widely recognized that making a good matching decision requires to take into account the rich interaction structures in the text matching process, starting from the interactions between words, to various matching patterns in the phrases and the whole sentences. Taking the aforementioned two sentences as an example, the interaction structures are of different levels, as illustrated in Figure 1.

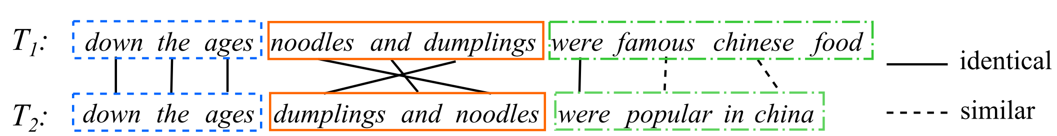


Figure 1: An example of interaction structures in paraphrase identification.

Word Level Matching Signals refer to matchings between words in the two texts, including not only identical word matchings, such as "*down–down*", "*the–the*", "*ages–ages*", "*noodles–noodles*", "*and–and*","*dumplings–dumplings*" and "*were–were*", but also similar word matchings, such as "*famous–popular*" and "*chinese–china*".

Phrase Level Matching Signals refer to matchings between phrases, including n-gram and n-term. N-gram matching occurs with n exactly matched successive words, e.g. "(*down the ages*)–(*down the ages*)". While n-term matching allows for order or semantic alternatives, e.g. "(*noodles and dumplings*)–(*dumplings and noodles*)", and "(*were famous chinese food*)–(*were popular in china*)".

Sentence Level Matching Signals refer to matchings between sentences, which are composed of multiple lower level matching signals, e.g. the three successive phrase level matchings mentioned above. When we consider matchings between paragraphs that contain multiple sentences, the whole paragraph will be viewed as a long sentence and the same composition strategy would generate paragraph level matching signals.

Text Matching as Image Recognition, Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, and Xueqi Cheng, AAAI2016.

论文润色

- 对比

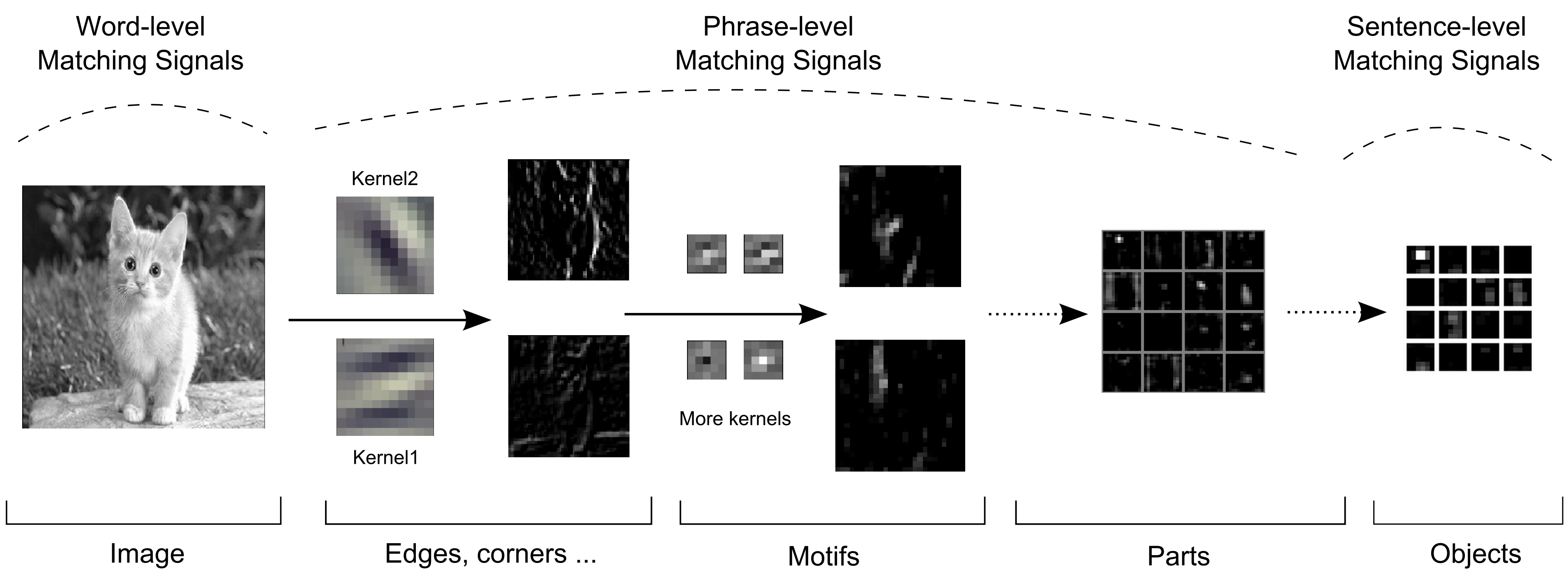


Figure 2: Relationships between text matching and image recognition.

Text Matching as Image Recognition, Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, and Xueqi Cheng, AAAI2016.

论文润色

- 善用图表

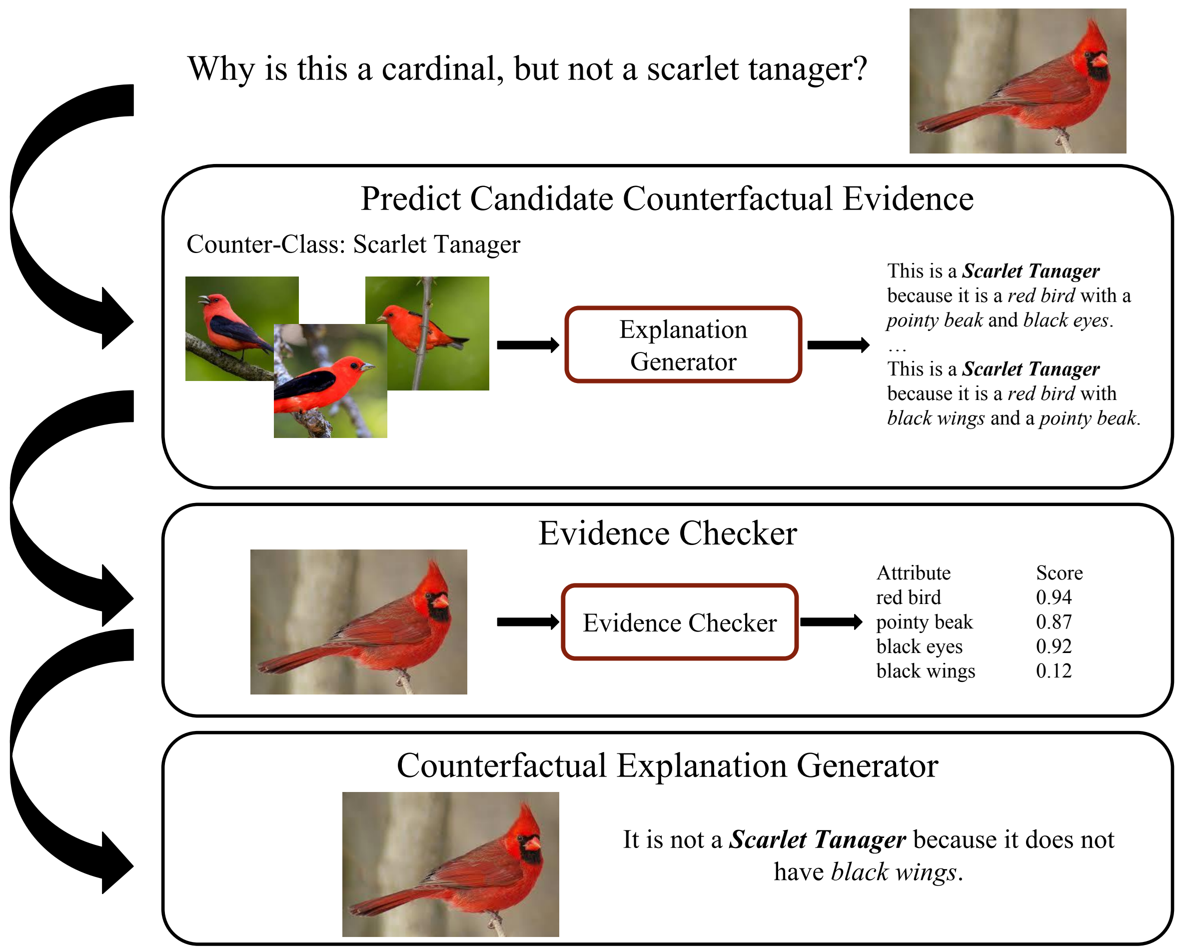


Figure 1. Outline of our counterfactual explanation pipeline. We first predict candidate counterfactual evidence, then determine if counterfactual evidence is in the image, then finally generates a cohesive sentence which mentions the counterfactual evidence which cannot be found in the image.

Prof. Trevor Darrell
UC Berkeley

润色后的道理

Recently *pairwise preference judgment* has been investigated as a good alternative [20, 26]. Instead of assigning a relevance grade to a document, an assessor looks at two pages and judges which one is better. Compared with absolute relevance judgment, the advantages lie in that: (1) There is no need to determine the gradation specifications as it is a binary decision. (2) It is easier for an assessor to express a preference for one document over the other than to assign a pre-defined grade to each of them [7]. (3) Most state-of-the-art learning to rank models, pairwise or listwise, are trained over preferences. As noted by Carterette et al. [7], "by collecting preferences directly, some of the noise associated with difficulty in distinguishing between different levels of relevance may be reduced." Although preference judgment likely produce more reliable labeled data, it is often criticized for increasing the complexity of judgment (e.g. from $\mathcal{O}(n)$ to $\mathcal{O}(n \log n)$ [20]), which poses a big challenge in wide use. Do we actually need to judge so many pairs for real search systems? If not, which pairs do we choose? How to choose? These questions become the original motivation of this paper.

→ 具体解释

→ 对比

→ 名言

→ 数据

→ 问题

论文润色

- 承前启后 Smooth

First, Second, Third, Finally

As a result, Consequently

更高级一点：每句话都是水到渠成的，Natural

语言?

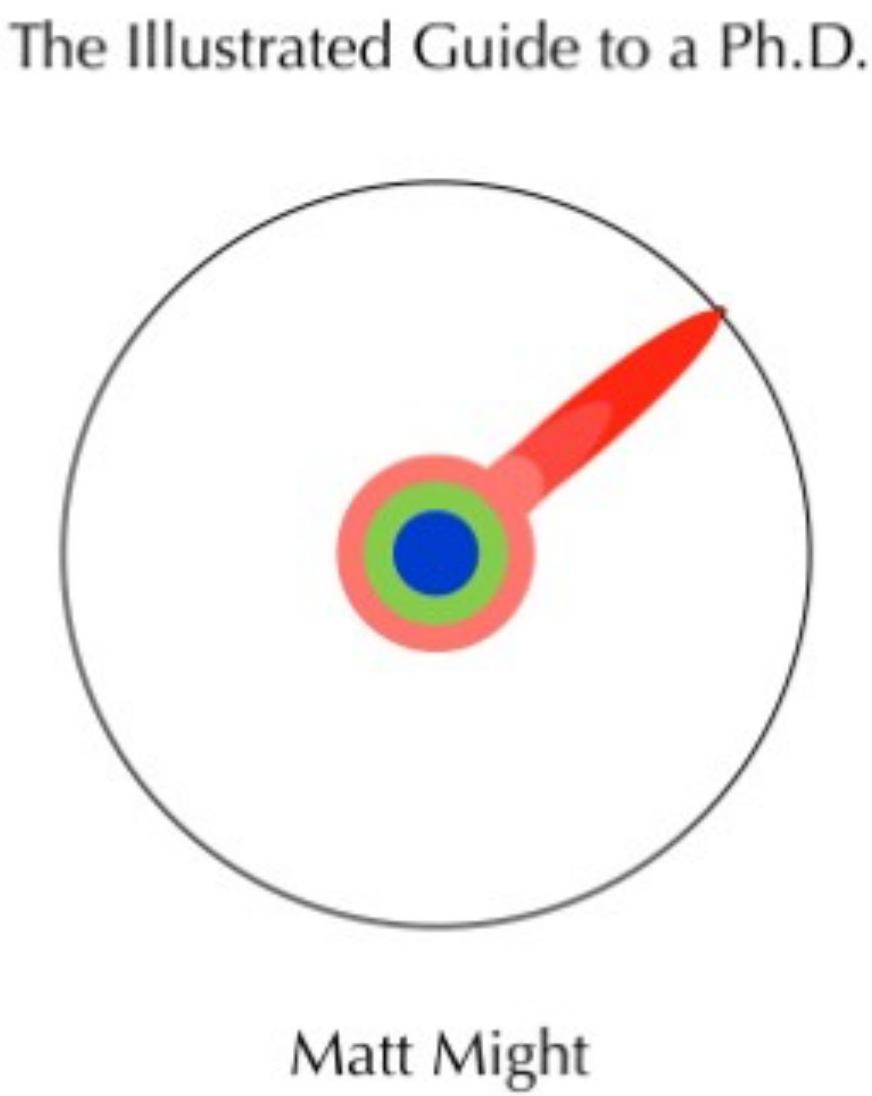

Fluency or pronunciation
for a better conversation?
Panglossity

思想

逻辑

文化

提升



Framework

努力探寻你的工作边界

努力找寻你和其他工作的关系

Learn from Others

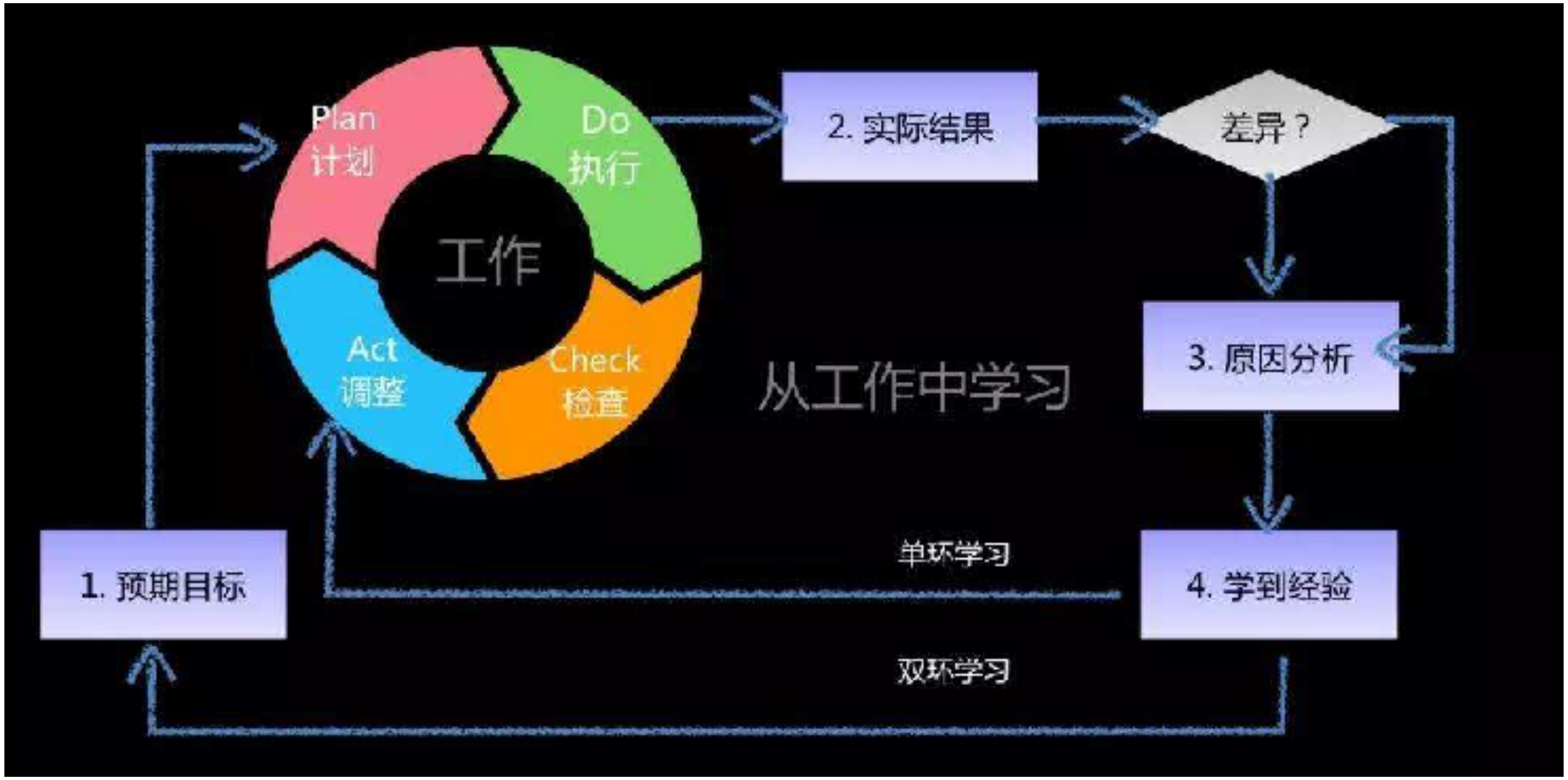
Self Learning

Less is More 舍得

隔而不离　　密斯·凡德罗　　流动空间

复盘：把经验转化为能力


02
03
复盘


Plan
计划
Do
执行
工作
Act
调整
Check
检查
2. 实际结果
差异？
3. 原因分析
从工作中学习
单环学习
1. 预期目标
4. 学到经验
双环学习

Practice! Practice! Practice!

“Think like a man of action, act like a man of thought.”

–Henri Bergson