


(/apps/redirect?utm\_source=side-banner-click)

# FM因式分解机

 ying\_zhang (/u/21a5b951bbb3) [+ 关注](#)  
🔥 0.7 2018.09.04 22:02 字数 1074 阅读 285 评论 0 喜欢 5  
(/u/21a5b951bbb3)

## 写在前面

FM全称为factorization machine, 可以用解决回归、二分类问题  
目的：解决高维**稀疏**数据中特征组合问题，适用于**categorical feature**。

## 参考文献

- 1、[http://www.cnblogs.com/Matrix\\_Yao/p/4773221.html](http://www.cnblogs.com/Matrix_Yao/p/4773221.html)  
([http://www.cnblogs.com/Matrix\\_Yao/p/4773221.html](http://www.cnblogs.com/Matrix_Yao/p/4773221.html)) 梳理了ctr估计问题的大致流程，给出了一些工业界的方法，可以当做入门资料
- 2、<https://blog.csdn.net/google19890102/article/details/45532745>  
(<https://blog.csdn.net/google19890102/article/details/45532745>) 针对于fm讲的很透彻，有些点没提到，程序很好懂
- 3、<https://blog.csdn.net/g11d111/article/details/77430095>  
(<https://blog.csdn.net/g11d111/article/details/77430095>) 背景交代很清楚，我的背景介绍也是copy他的
- 4、关于fm与LR的比较，可以参看  
<https://www.zhihu.com/question/27043630/answer/151138052>  
(<https://www.zhihu.com/question/27043630/answer/151138052>)最高赞的回答
- 5、<https://blog.csdn.net/itplus/article/details/40536025>  
(<https://blog.csdn.net/itplus/article/details/40536025>) 理论进阶

([https://dsp-click.youdao.com/clk\\_slot=30edd91dd86374e20-437c-b7ef-b3f9bcc76249&iid=%](https://dsp-click.youdao.com/clk_slot=30edd91dd86374e20-437c-b7ef-b3f9bcc76249&iid=%)

## 背景

### 1、稀疏数据

**强调一点，FM的适用对象是稀疏数据。**这一点之后会有更深入的介绍。  
实际中，很多特征类型是categorical型，比如性别特征，有男、女两个选项，如果将男性标记为1，女性标记为2是不太合理的，因为数字是具有意义的，2是1的2倍，而不能说女性是男性的2倍，所以对于categorical feature都会使用独热编码one-hot encoding，将男性标记为[1,0]，女性标记为[0,1]。其他的categorical feature还有很多，比如文章类型，娱乐、运动、军事、科技等等，这些类别之间不具有数值意义的关系，同理需要使用one-hot encoding，关于one-hot encoding，如果你还不太了解，请看<https://www.imooc.com/article/35900> (<https://www.imooc.com/article/35900>)  
categorical feature做完one-hot encoding之后是非常稀疏的，这在实际中十分常见，而许多方法对于稀疏数据都束手无策，比如SVM，它无法在非常稀疏的数据下学习复杂的非线性内核空间中的参数。

### 2、特征组合

在进行数据分析的过程中，特征工程是非常重要的一步，在特征工程这部分处理的好的话可以让模型的效果事半功倍。实际中有很多特征是相关联的，比如一般女性用户看化妆品服装之类的广告比较多，而男性更青睐各种球类装备。那很明显，女性这个特征与化妆品类服装类商品有很大的关联性，男性这个特征与球类装备的关联性更为密切。如果我们能将这些有关联的特征找出来，显然是很有意义的。FM就提供了一种这样特征组合的思路。



原理

我们还是先从线性模型说起好啦：)

一般的线性模型为

$$y = \omega_0 + \sum_{i=1}^n \omega_i x_i$$

线性模型.png

n表示n维特征

如果在线性模型中加入二阶特征的组合，那么会是这个样子的

$$\hat{y}(x) = \omega_0 + \sum_{i=1}^n \omega_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \omega_{ij} x_i x_j$$

加入二阶特征组合的线性模型.png

这里存在一个问题，对于稀疏数据来说，xi和xj同时不为0的情况非常少，这样会导致Wij无法通过训练获得。为了解决这个问题，FM诞生了，我们看一下FM是如何解决这个问题的：

对每一个特征分量xi引入辅助向量 $\mathbf{v}_i = (v_{i,1}, v_{i,2}, \dots, v_{i,k})$ ，利用 $\mathbf{v}_i \mathbf{v}_j^T$ 对交叉项的系数 $w_{ij}$ 进行估计，即 $\hat{w}_{ij} := \mathbf{v}_i \mathbf{v}_j^T$

令

$$\mathbf{V} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1k} \\ v_{21} & v_{22} & \cdots & v_{2k} \\ \vdots & \vdots & & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nk} \end{pmatrix}_{n \times k} = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_n \end{pmatrix}$$

则

$$\hat{\mathbf{W}} = \mathbf{V} \mathbf{V}^T = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_n \end{pmatrix} (\mathbf{v}_1^T \quad \mathbf{v}_2^T \quad \cdots \quad \mathbf{v}_n^T)$$

FM.png

这也解释了FM因式分解机名字的由来，它是将Wij进行了拆解。FM的模型为

对于度为2的因子分解机FM的模型为：

$$\hat{y} := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$

其中，参数 $w_0 \in \mathbb{R}$ ， $\mathbf{w} \in \mathbb{R}^n$ ， $\mathbf{V} \in \mathbb{R}^{n \times k}$ 。 $\langle \mathbf{v}_i, \mathbf{v}_j \rangle$ 表示的是两个大小为k的向量 $\mathbf{v}_i$ 和向量 $\mathbf{v}_j$ 的点积：

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle := \sum_{f=1}^k v_{i,f} \cdot v_{j,f}$$

其中， $\mathbf{v}_i$ 表示的是系数矩阵 $\mathbf{V}$ 的第i维向量，且 $\mathbf{v}_i = (v_{i,1}, v_{i,2}, \dots, v_{i,k})$ ， $k \in \mathbb{N}^+$ 称为超参数。在因子分解机FM模型中，前面两部分是传统的线性模型，最后一部分将两个互异特征分量之间的相互关系考虑进来。

因子分解机FM也可以推广到高阶的形式，即将更多互异特征分量之间的相互关系考虑进来。

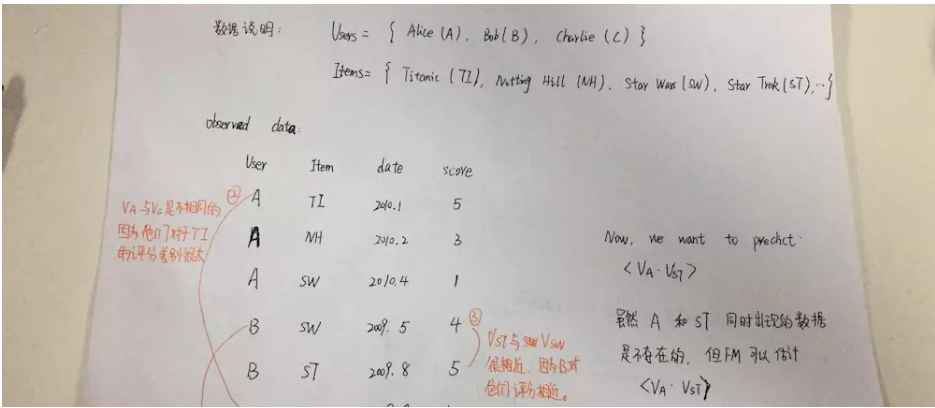
FM.png

求解

先不急着想如何求解的，这部分想解释下为什么将Wij拆解成vi和vj就能够求解了呢？这部分在作者的论文中有提到，下面的图片看着不太舒服的话，可以去看论文的第三部分：)

(/apps/redirect?utm\_source=side-banner-click)

(https://dsp-click.youdao.com/clkslot=30edd91dd86374e20-437c-b7ef-b3f9bcc76249&iid=%



(/apps/redirect?utm\_source=side-banner-click)

解释如何求解vi、vj.jpg

求解的话肯定需要一个优化目标，就是使损失函数最小

前面我们提到, FM 的优化目标是整体损失函数

$$L = \sum_{i=1}^N \text{loss}(\hat{y}(\mathbf{x}^{(i)}), y^{(i)}),$$

其中, 对于回归问题, loss 取最小平方误差

$$\text{loss}^R(\hat{y}, y) = (\hat{y} - y)^2,$$

对于二分类问题, loss 取 logit 函数

$$\text{loss}^C(\hat{y}, y) = -\ln \sigma(\hat{y}y).$$

于是, FM 的最优化问题就变成了

$$\Theta^* = \arg \min_{\Theta} \sum_{i=1}^N \text{loss}(\hat{y}(\mathbf{x}^{(i)}), y^{(i)}),$$

当然, 我们通常还考虑 L2 正则, 因此, 最优化问题就变成

$$\Theta^* = \arg \min_{\Theta} \sum_{i=1}^N \left( \text{loss}(\hat{y}(\mathbf{x}^{(i)}), y^{(i)}) + \sum_{\theta \in \Theta} \lambda_{\theta} \theta^2 \right),$$

其中,  $\lambda_{\theta}$  表示参数  $\theta$  的正则化系数.

损失函数.png

其中,

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

sigmoid.png

表示的是阶跃函数Sigmoid  
基于随机梯度下降方式的求解：



§4.3 随机梯度下降法 (SGD)

首先给出利用 SGD 训练 FM 模型的算法流程.

算法 4.1 (SGD)

- 输入: 训练集  $D$ , 正则化系数集  $\lambda$ , 学习率  $\eta$ , 正态分布方差参数  $\sigma$ .
- 输出: 模型参数  $\Theta = (w_0, \mathbf{w}, V)$ .

Initialization:  $w_0 := 0; \mathbf{w} := \mathbf{0}; V \sim \mathcal{N}(0, \sigma)$ .

Repeat

FOR  $(\mathbf{x}, y) \in D$  DO

{

$w_0 = w_0 - \eta \left( \frac{\partial \text{loss}(\hat{y}(\mathbf{x}), y)}{\partial w_0} + 2\lambda^0 w_0 \right);$

FOR  $i \in \{1, 2, \dots, n\}$  DO

{

IF  $x_i \neq 0$

{

$w_i = w_i - \eta \left( \frac{\partial \text{loss}(\hat{y}(\mathbf{x}), y)}{\partial w_i} + 2\lambda_{\pi(i)}^w w_i \right);$

FOR  $j \in \{1, 2, \dots, k\}$  DO

{

$v_{ij} = v_{ij} - \eta \left( \frac{\partial \text{loss}(\hat{y}(\mathbf{x}), y)}{\partial v_{ij}} + 2\lambda_{\pi(i), j}^v v_{ij} \right);$

}

}

}

}

}

Until stopping criterion is met.

(/apps/redirect?utm\_source=side-banner-click)

(https://dsp-click.youdao.com/clkslot=30edd91dd86374e20-437c-b7ef-b3f9bcc76249&iid=%

算法 4.1 中涉及损失函数的偏导数, 具体计算公式如下

- 当  $\text{loss} = \text{loss}^R$  时

$$\frac{\partial \text{loss}^R(\hat{y}(\mathbf{x}), y)}{\partial \theta} = 2(\hat{y}(\mathbf{x}) - y) \frac{\partial \hat{y}(\mathbf{x})}{\partial \theta}$$

- 当  $\text{loss} = \text{loss}^C$  时

sgd求解.png

$$\begin{aligned} & \sum_{i=1}^{n-1} \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j - \frac{1}{2} \sum_{i=1}^n \langle \mathbf{v}_i, \mathbf{v}_i \rangle x_i x_i \\ &= \frac{1}{2} \left( \sum_{i=1}^n \sum_{j=1}^n \sum_{f=1}^k v_{i,f} v_{j,f} x_i x_j - \sum_{i=1}^n \sum_{f=1}^k v_{i,f} v_{i,f} x_i x_i \right) \\ &= \frac{1}{2} \sum_{f=1}^k \left( \left( \sum_{i=1}^n v_{i,f} x_i \right) \left( \sum_{j=1}^n v_{j,f} x_j \right) - \sum_{i=1}^n v_{i,f}^2 x_i^2 \right) \\ &= \frac{1}{2} \sum_{f=1}^k \left( \left( \sum_{i=1}^n v_{i,f} x_i \right)^2 - \sum_{i=1}^n v_{i,f}^2 x_i^2 \right) \end{aligned}$$

image.png

这也解释了为什么FM的计算复杂度为O ( kn )

$$\frac{\partial \hat{y}}{\partial \theta} = \begin{cases} 1, & \text{if } \theta = w_0 \\ x_i, & \text{if } \theta = w_i \\ x_i \sum_{j=1}^n v_{j,f} x_j - v_{i,f} x_i^2, & \text{if } \theta = v_{i,f} \end{cases}$$

image.png



FM优点

- 1、可以对稀疏数据中的特征进行组合
- 2、计算时间复杂度为O ( kn )
- 3、FM是一种可以与任何实值特征向量一起使用的通用预测器。

(/apps/redirect?utm\_source=side-banner-click)

缺点

你发现了没？在做特征组合的时候，我们不确定是同一域内的特征相组合（这不太合理，比如男性【1,0】两个维度数据，如果组合的是性别本身这两个维度，不太有意义），还是组合的是域间维度，比如性别和商品类别之间的组合是有意义的，对FM的一个改进是FFM, FFM是使得特征有自己的归属域，比如男性【1,0】是性别域，这两维数据不能拆开，下次我们再详细介绍一下FFM.

paper (/nb/22406224) 举报文章 © 著作权归作者所有



ying\_zhang (/u/21a5b951bbb3)

写了 64654 字，被 7 人关注，获得了 10 个喜欢

(/u/21a5b951bbb3)

+ 关注

精致的沉默，凌驾于一切之上。

(https://dsp-click.youdao.com/clkslot=30edd91dd86374e20-437c-b7ef-b3f9bcc76249&iid=%

喜欢 | 5



更多分享



下载简书 App ▶  
随时随地发现和创作内容



(/apps/redirect?utm\_source=note-bottom-click)




登录后发表评论 (/sign\_in?utm\_source=desktop&utm\_medium=not-signed-in-comment-form)

评论

智慧如你，不想发表一点想法 (/sign\_in?utm\_source=desktop&utm\_medium=not-signed-in-nocomments-text)详~

被以下专题收入，发现更多相似内容



机器学习 (/c/53c50020e436?utm\_source=desktop&utm\_medium=notes-included-collection)



(/p/83ca2a3499c8?

utm\_campaign=maleskine&utm\_content=note&utm\_source=side-banner-click)

月宸 (/u/5aa13414a08c?)

utm\_campaign=maleskine&utm\_content=user&utm\_medium=pc\_all\_hots&utm\_source=recommendation)

不背单词和语法，高效学英语→

(<https://dsp-click.youdao.com/clk/request.s?>

slot=f2ac00aef0eb6b673f4e4639046bc6f8&k=%2FmJtG5ky5A%2FYF3r2o02Q%2Fub72ogckckcpoVJruO1LsrXx3975-42d5-bdf2-fb8a98610771&iid=%7B%226329144451546090023%22%3A1%7D&sid=17836)

转自：许莫私人音乐厅作品 没有谁的人生，能十全十美 没有谁的生活，能一帆风顺 没有一个人，不会经历坎坷 没有一个人，不会被情折磨 真心没有错，给错了人 才会让你一再伤心 认真没有错，爱得太满 才会落...

月宸 (/u/5aa13414a08c?)

utm\_campaign=maleskine&utm\_content=user&utm\_medium=pc\_all\_hots&utm\_source=recommendation) slot=30edd91dd86374e20-437c-b7ef-b3f9bcc76249&iid=%

utm\_campaign=maleskine&utm\_content=note&utm

新年将至，千家万户沉浸在喜庆之中。今年过年，你除了像以往一样，为孩子准备了新衣服，新玩具等礼物，是否想到，在这个特别的日子里，送份特殊的新...

暖暖育儿家 (/u/6300ea954632?)

utm\_campaign=maleskine&utm\_content=user&utm\_medium=pc\_all\_hots&utm\_source=recommendation)

(/p/4083d9ade644?

来时车马两匆匆，去日侯门类转蓬。此次欠卿同一醉，引尊遥拜九天风。（平水韵平起 首句入韵 一东）

夏朝辉\_b46e (/u/73917f126014?)

utm\_campaign=maleskine&utm\_content=user&utm\_medium=pc\_all\_hots&utm\_source=recommendation)

(/p/fd8e37f781c5?

- 1.最好的爱情，不是因为我们彼此需要在一起，而是因为我们彼此想要在一起。
- 2.爱，从来就是一件千回百转的事。不曾被离弃，不曾受伤害，怎懂得爱人？...

苏若儿 (/u/9b487134d84b?)


utm\_campaign=maleskine&utm\_content=user&utm\_medium=pc\_all\_hots&utm\_source=recommendation)

(/p/7fcce74d90f2?



utm\_campaign=maleskine&utm\_content=note&utm\_medium=seo\_notes&utm\_source=recommendation)

大数据文摘投稿作品 \*\*作者：龙心尘、寒小阳\*\* 多年以后，当资深算法专家们看着无缝对接用户需求的广告收入节节攀升时，他们可能会想起自己之前痛苦推导FM与深度学习公式的某个夜晚…… ——题记 1.引言 ...

 大数据文摘 (/u/5303a7d15da1?)

utm\_campaign=maleskine&utm\_content=user&utm\_medium=seo\_notes&utm\_source=recommendation)

(/p/dce9f1af7bc9?






utm\_campaign=maleskine&utm\_content=note&utm\_medium=seo\_notes&utm\_source=recommendation) (/apps/redirect?utm\_source=side-banner-click)

逻辑回归（LR）个人学习总结篇 (/p/dce9f1af7bc9?utm\_campaign=male...

写作计划：线性模型LR(没有考虑特征间的关联)——>LR + 多项式模型（特征组合，不适用于特征稀疏场景，泛化能力弱）——>FM（适用于稀疏特征场景\*，泛化能力强）——>FFM【省去零值特征，提高FFM...

 流川枫AI (/u/c87c7667648a?)

utm\_campaign=maleskine&utm\_content=user&utm\_medium=seo\_notes&utm\_source=recommendation)

(/p/f71aad7b5e5d?)

- 从超参数为 $\alpha$ 的 Dirichlet 分布中抽样生成文档 $d_n$ 的主题分布 $\theta_n$
- 对文档 $d_n$ 中的每一个词 $w_{nm}$ :
  - 从代表主题的多项式分布 $\theta_n$ 中抽样生成它所对应的主题 $z_{nm}$
  - 从超参数为 $\beta$ 的 Dirichlet 分布中抽样生成主题 $z_{nm}$ 对应的分布 $\phi_{z_{nm}}$
  - 从代表词的多项式分布 $\phi_{z_{nm}}$ 中抽样生成词 $w_{nm}$

utm\_campaign=maleskine&utm\_content=note&utm\_medium=seo\_notes&utm\_source=recommendation)

Hulu面试题(三) (/p/f71aad7b5e5d?utm\_campaign=maleskine&utm\_con...

https://mp.weixin.qq.com/s/BuHiG9FjX-OiSNWx3KquQQ 17.随机梯度下降算法之经典变种 场景描述 提到 Deep Learning中的优化方法，人们都会想到Stochastic Gradient Descent (SGD)，但是S...

 龙雀 (/u/beacac8e35d1?)

utm\_campaign=maleskine&utm\_content=user&utm\_medium=seo\_notes&utm\_source=recommendation) (https://dsp-click.youdao.com/clk?slot=30edd91dd86374e20-437c-b7ef-b3f9bcc76249&iid=...


(/p/36b7fcd1591f?)

feature column		feature column		feature column		label
0.1	0.3	0	1	0.01	0.02	1

utm\_campaign=maleskine&utm\_content=note&utm\_medium=seo\_notes&utm\_source=recommendation)

<转载>TensorFlow Wide And Deep 模型详解与应用 (/p/36b7fcd1591f?ut...

该文章为转载文章，作者简介：汪剑，现在在出门问问负责推荐与个性化。曾在微软雅虎工作，从事过搜索和推荐相关工作。 TensorFlow Wide And Deep 模型详解与应用（一）TensorFlow Wide And Deep 模型详...

 名字真的不重要 (/u/65a751648086?)

utm\_campaign=maleskine&utm\_content=user&utm\_medium=seo\_notes&utm\_source=recommendation)

总结一点面试问题 - - 算法工程师（机器学习） (/p/4a3c5e34d0f8?utm\_ca...


~~~~~个人整理，如需转载，请说明并备注，不甚感激~~~~~ csdn博客文章地址 需要内推三七互娱的盆友萌，（9月5号截止）可以参考另一篇文章，或者内推QQ群：175423207 BAT机器学习面试系列 1.清简要...

 suxuer (/u/ff138b35b8ed?)

utm\_campaign=maleskine&utm\_content=user&utm\_medium=seo\_notes&utm\_source=recommendation)

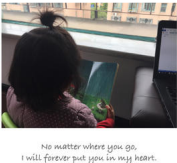
佛与心 (/p/204a77ad7796?utm\_campaign=maleskine&utm\_content=no...

过年烧香是持续了不知多少年的传统活动，从小到大二十多年来年年跟着父母后面看着他们买香烛；看着他们将香点起，青烟袅袅；看着他们虔诚参拜，闭目合十。其实小的时候也会学着大人模样对着无数菩萨从...

 你好达达尼安 (/u/73009fbb83ff?)

utm\_campaign=maleskine&utm\_content=user&utm\_medium=seo\_notes&utm\_source=recommendation)


(/p/6e5ee62ad5e3?)



utm\_campaign=maleskine&utm\_content=note&utm\_medium=seo\_notes&utm\_source=recommendation)

雨后 (/p/6e5ee62ad5e3?utm\_campaign=maleskine&utm\_content=note...

下午出门前，天空下起了雷雨，带着大孩上完课回来，雨已停，外面的空气里全是潮湿，屋里显得闷热，随即即将窗户打开，顿觉四面来风，心里一阵舒畅。 小不点，穿了一件长T，中午出门前给她解散睡觉的头...

 木子石榴 (/u/ecaad9d6244c?utm\_campaign=maleskine&utm\_content=user&utm\_medium=seo\_notes&utm\_source=recommendation)

(/p/ab8d3541487c?




(/apps/redirect?utm\_source=side-banner-click)

utm\_campaign=maleskine&utm\_content=note&utm\_medium=seo\_notes&utm\_source=recommendation)


**如何让阅读为写作服务 | 英语学习 (/p/ab8d3541487c?utm\_campaign=ma...**

好的英语阅读材料有三大要求，分别是： Read for pleasure. Read for information. Read for language improvement. 但遗憾的是，这样的资源少之又少。因此，若想让阅读为写作服务，动机就要进行调整。早...

 梁洋Earnest (/u/cd59228d0593?utm\_campaign=maleskine&utm\_content=user&utm\_medium=seo\_notes&utm\_source=recommendation)

**《格列佛游记》第三章概括 (/p/18f86dddcdb1?utm\_campaign=maleskin...**


格列佛的善良和大度赢得了小人国对他的好感，上至国王下至臣民渐渐对他消除了敌意。格列佛在皇宫里看到了小人国的几种娱乐活动，了解了政府的可笑，同时格列佛也发明了一种游戏供皇帝消遣。在格列佛多...

 顾晓雨妈妈 (/u/28ca7632cc6c?utm\_campaign=maleskine&utm\_content=user&utm\_medium=seo\_notes&utm\_source=recommendation)

(https://click.youdao.com/click\_slot=30edd91dd86374e20-437c-b7ef-b3f9bcc76249&iid=%

**CentOS下如何删除或重命名乱码文件 (/p/5c5b28751da2?utm\_campaign=...**

第一步：获取文件的id号ls -li第二步：执行删除find . -inum 32983551 -exec rm {} ;二、重命名find . -inum 32983542 -exec mv {} test.php ;实际操作：

 wsf535 (/u/e00d16255326?utm\_campaign=maleskine&utm\_content=user&utm\_medium=seo\_notes&utm\_source=recommendation)

