

网站用户访问特征评估体系

iResearch Website User Behavior Evaluation System

2010 年

目 录

I. 评估体系简介	3
II. 评估体系指标定义	3
III. 评估体系详细说明.....	4
1. 什么是好的评估指标体系.....	4
2. 4 大基本评估指标，以及和其他日均指标间的关系	5
3. 网民回访模型和 1 个回访评估参数.....	7
4. 目标用户集中度指数 TGI.....	10
5. 完整的网站用户访问特征评估体系	12
5.1. 六大指标的统计和计算方法定义	12
5.2. 媒体类型评估实例	13
法律声明	14

图 表 目 录

图 2- 1 iResearch-艾瑞风向标媒体评估日均指标的因素关系	5
图 2- 2 iResearch-艾瑞风向标媒体评估日均指标的非因果关系.....	6
图 3- 1 iResearch-艾瑞风向标媒体访客累加模式图	7
图 3- 2 iResearch-艾瑞风向标媒体访客累加指数的数据验证	8
图 5- 1 iResearch-艾瑞风向标媒体类别雷达图	13
表 3- 1 iUserTracker-2009 年 47 周主要网站回访比例	9
表 3- 2 iUserTracker-2009 年 47 周主要网站目标用户 TGI.....	10

I. 评估体系简介

如何衡量一个网站的用户访问特性，一直是行业内热烈讨论的问题，而由于指标众多；媒体，代理商，各家用法不同，所以至今尚无一套完整的评估体系。本文将从第三方的角度，提出一套相对完整和科学的网站用户访问特征评估体系。上一期网民风向标中已经运用了一些标准话的评估指标，这里会将这套体系更加完善，全面化。

考虑到评估方法的科学和公正，该体系主要处理网站覆盖人数，流量，停留时间等可由第三方样本监测得到，并且可以标准化，指标化的网站用户访问属性。另外一些网站分析中使用的来源去向，路径分析等数据和指标并不在该评估体系范围内。

本文最后的结论是一个包含 6 个指标的网站特征评估指标体系。包括 4 个已有的基本指标，1 个回访模型参数，1 个目标人群集中度参数。还有基于这个指标体系的应用。

II. 评估体系指标定义

- **日均覆盖人数 (Unique Visitors Per Day)**
该月中，该网站平均每天的独立访问用户数，用户重复访问不重复统计
$$\text{日均覆盖人数} = \text{当周每日的网站覆盖人数的平均值}$$
- **日均回访比例 (Revisit rate)**
该天访客中，平均属于本周老访客的比例（也等与 1-新访客比例）
$$\text{日均回访比例} = (7\text{-周覆盖人数} / \text{日均覆盖人数}) / 6$$
- **人均单日访问次数 (Visits Per User Per Visiting Day)**
平均每个用户，平均每日访问该网站日访问该网站的次数
$$\text{人均单日访问次数} = \text{Average}(\text{该用户该周访问该网站的次数} / \text{该用户该周访问该网站的天数})$$
- **人均单次浏览页面数 (PV Per User Per Visit)**
平均每个用户，平均每次访问该网站的浏览页面数量
$$\text{人均单次浏览页面数} = \text{Average}(\text{该用户该周该网站浏览页面数} / \text{该用户该周访问该网站的次数})$$
- **人均单页有效浏览时间 (Effective Viewing Duration Per User Per page)**
平均每个用户在浏览该网站每个页面时的有效浏览时间
$$\text{人均单页有效浏览时间} = \text{Average}(\text{该用户某时间段内该网站总浏览时间} / \text{该用户该时间段内该网站总浏览页面数})$$

注：单页打开时间小于 3 秒不计，单页停留 5 分钟以上，则视为用户脱机，不计其后的时间，直到新的网络行为发生
- **目标受众集中度 (Target group index)**
$$\text{TGI} = \text{某网站所有访问者中目标用户的比例} / \text{总体网民中目标用户的比例}$$

III. 评估体系详细说明

1. 什么是好的评估指标体系

从数学的角度看，要用最少的参数，建立最准确有效的模型；从统计学的角度看，要用最少的变量，反映样本最多的数据特征；所以这里，我们也将用最少的数据指标，准确和全面的反应网站的用户访问特征。通俗的将，就是定义各个网站的“德智体美劳”。

因此衍生出以下两个具体原则需要遵守：

1. 体系中的各指标间不相关，或弱相关
2. 体系中的各指标组合，换算，可以得出现有常用体系中的所有指标

以上两个原则和主成份分析，因子分析的原则是一致的。这些通用的统计方法，从纯粹的数据角度出发，找出几个不相关的统计指标，并以此来解释原有数据。这种方法存在下述一些缺陷：

1. 人为创造出因子间的高度不相关，而忽视了实际指标中的模糊性
2. 得出的数据模型不稳定，可能因数据变化而变化
3. 忽视了实际中这些指标间的内在联系

所以本文所属的评估体系，并不是从数学解释出发，而是从指标间的因果关系，和网民行为的内在模式出发。建立稳定，准确的，有效的，反应网民访问各个网站不同特征的指标体系。证明过程中用到的都是基本的统计学和数学知识，大家可以在标准的统计学教材内找到解释，这里不另作标注。证明和举例使用的都是 iUserTracker 2009 年 47 周，11 月 23 日-11 月 29 日的的数据。

2.4 大基本评估指标，以及和其他日均指标间的关系

目前市场上常用指标的有月度指标、周度指标和日度指标。日度指标则用一段时间的日均指标来代替。作为基本的评估指标，首先选取 4 个日均和人均的网站访问指标：

1. 网站日均覆盖人数
2. 人均单日访问次数
3. 人均单次浏览页面
4. 人均单页浏览时间

这 4 个基本指标和其他常用指标间存在因果关系，如下图：

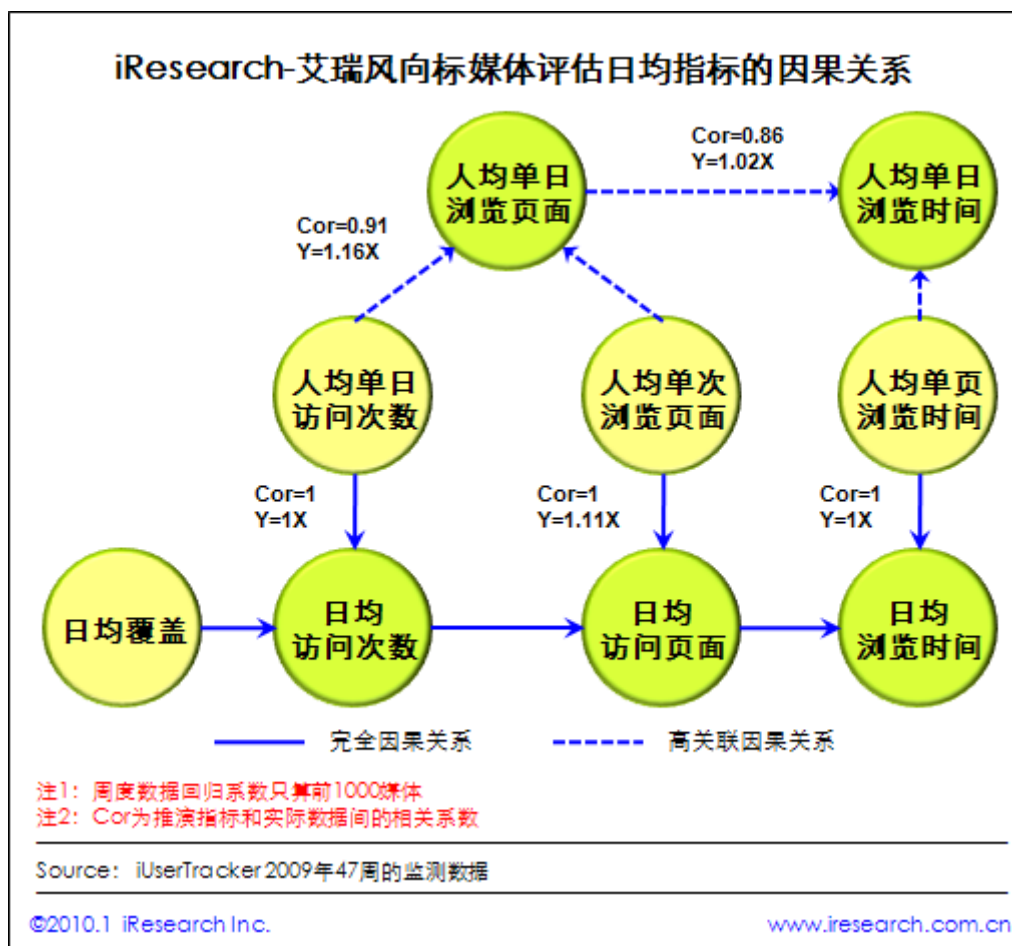


图 2-1 iResearch-艾瑞风向标媒体评估日均指标的因素关系

如图，一共有 9 个常用的日均和人均指标，其中选出的 4 个相对基本的指标用黄色圈表示。这 4 个指标间交叉相乘，恰能很好的得出其他的 5 个日均指标。而且这 4 个指标间互不相关。

- 1) 日均覆盖 × 人均单日访问次数 => 日均访问次数 cor=1
- 2) 人均单次浏览页面 × 日均访问次数 => 日均访问页面 cor=1
- 3) 人均单页浏览时间 × 日均访问页面 => 日均浏览时间 cor=1

以上 3 个推演的逻辑非常简单，而且很多第三方网络数据提供商实际就是根据这样的公式来计算人均指标的，所以他们的相关系数都接近 1。

- 4) 人均单日访问次数 × 人均单次浏览页面 => 人均单日浏览页面 cor=0.91
- 5) 人均单日浏览页面 × 人均单页浏览时间 => 人均单日浏览时间 cor=0.86

以上 2 个推演的逻辑也很简单，且从极高的相关系数看，也可认为逻辑成立。如果将该模型作为事件背后的规律，那么我们来看一下这个规律在现实中如何表现

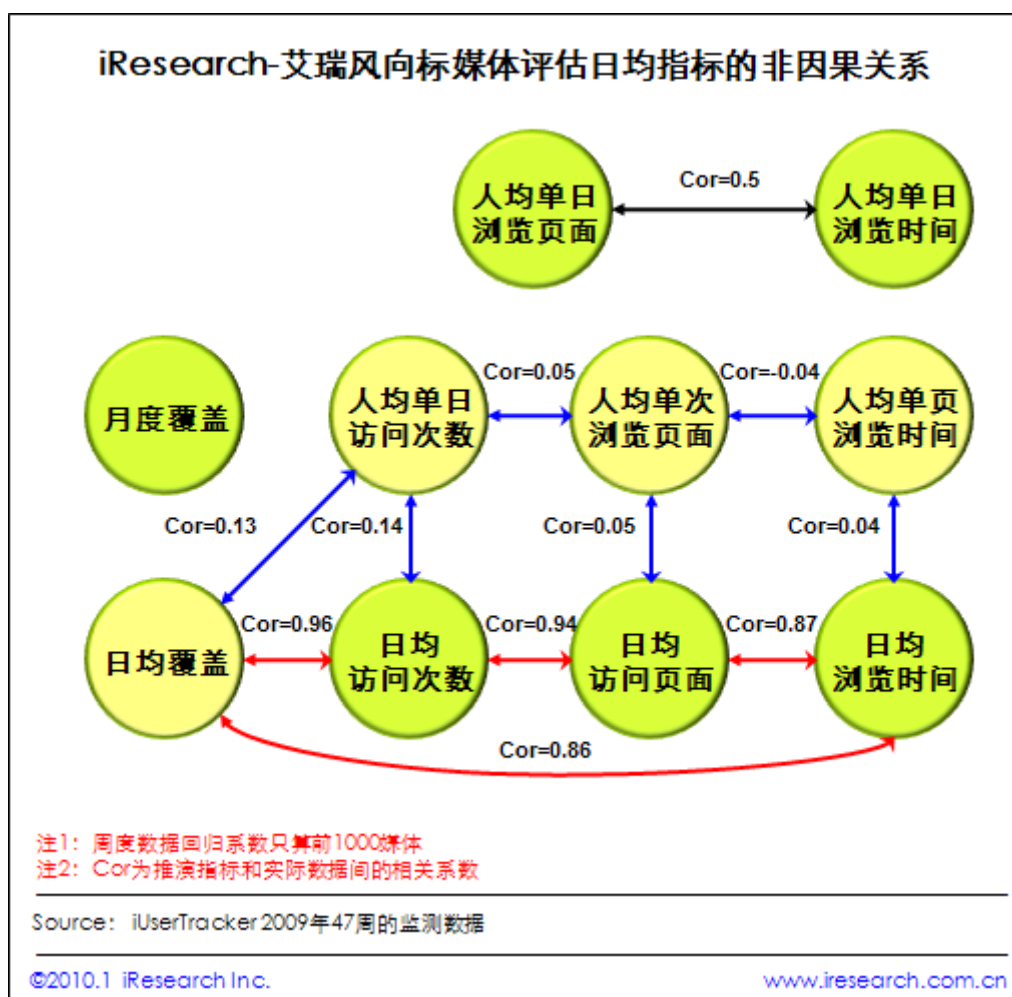


图 2- 2 iResearch-艾瑞风向标媒体评估日均指标的非因果关系

如图，我们可以看到 4 个基本指标间（黄色的圈）高度不相关，而最下面的 3 个日均指标都高度相关。原因就是每个日均指标都严重受到日均覆盖的影响。所以在评估媒体特征的时候，如果同时选择了多个日均指标，将对结果没有太大的意义。而只有选择 4 个基本指标（黄色的圈），才能有效的区分出不同媒体的特性。

这里有兴趣的读者可以从累计方差的角度进一步验证这 4 个指标。本文篇幅有限，且因为指标间的因果关系已经很明确，就不附加证明了。

3. 网民回访模型和 1 个回访评估参数

上面的部分已经把日均和人均指标间的问题解决了,剩下的就是如何从日均指标推演出周度,月度指标。因为目前大多的媒体行为都以周为单位,所以这里主要解决周度数据的推演。关于月度数据,有兴趣的读者可以自行扩展。

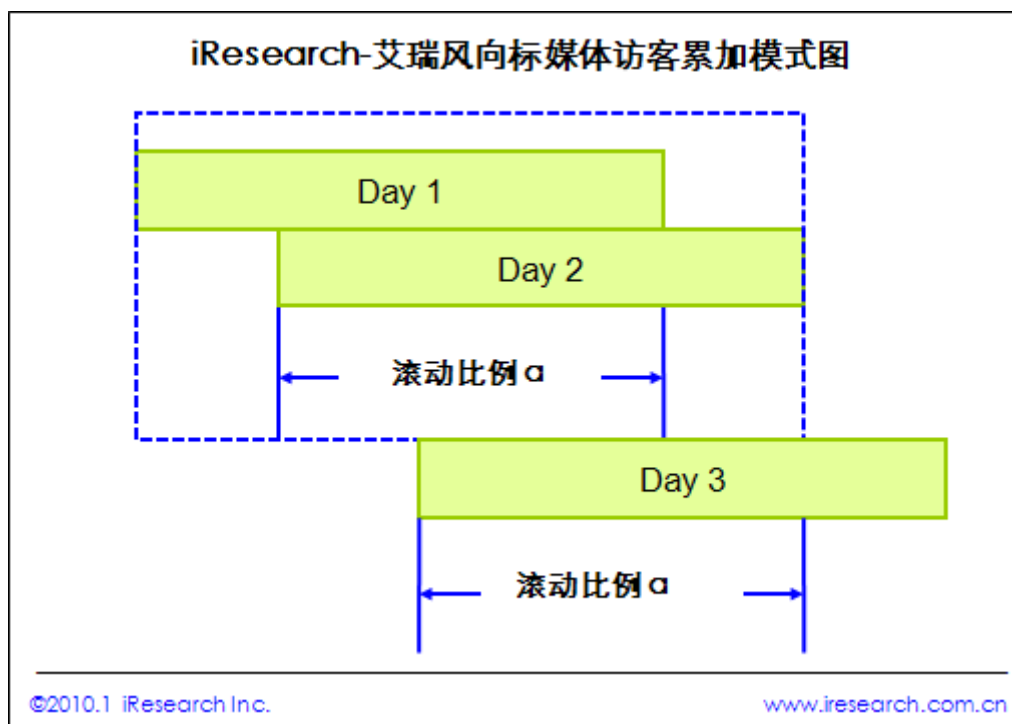


图 3-1 iResearch-艾瑞风向标媒体访客累加模式图

如图 3-1, 是一周内网民连续访问同一个网站的滚动模型。第一天访问了该网站的网民中, 有一部分第二天继续回访, 也有一部分不再访问。而第二天访问该网站的网民中, 有一部分来自第一天的老网民, 也一部分是本周还没来过的新网民。所以图中用参数 α 表示后来的网站访问中, 本周老访客的比例。

以一周 7 天计, 且假设每天的用户行为一致, 设以下变量:

每一天的: 网民覆盖人数 n , 人均访问次数 c , 人均浏览页面 p , 人均浏览时间 t

每一周的: 网民覆盖人数 N , 人均访问次数 C , 人均浏览页面 P , 人均浏览时间 T

在图中滚动比例 α 确定的情况下, 经过 7 天的迭代, 得到:

$$6) \quad N = (7 - 6\alpha)n$$

$$7) \quad C = 7nc / N = 7c / (7 - 6\alpha)$$

$$8) \quad P = 7p / (7 - 6\alpha)$$

$$9) \quad T = 7t / (7 - 6\alpha)$$

其中公式 7) 也等效为人均周访问天数

我们将 N/n , C/c , P/p , T/t 称为他们各自的周累加指数, 即:

10) 单日的指标 $\times$ 周累加指数 = 周度指标

可以看出，除了周覆盖累加指数和其他的累加指数呈反比关系，其他各累加指数间都是线性关系。iUserTracker 的数据证明了这一现象，从而也证明了这个模型的有效性。见下图：

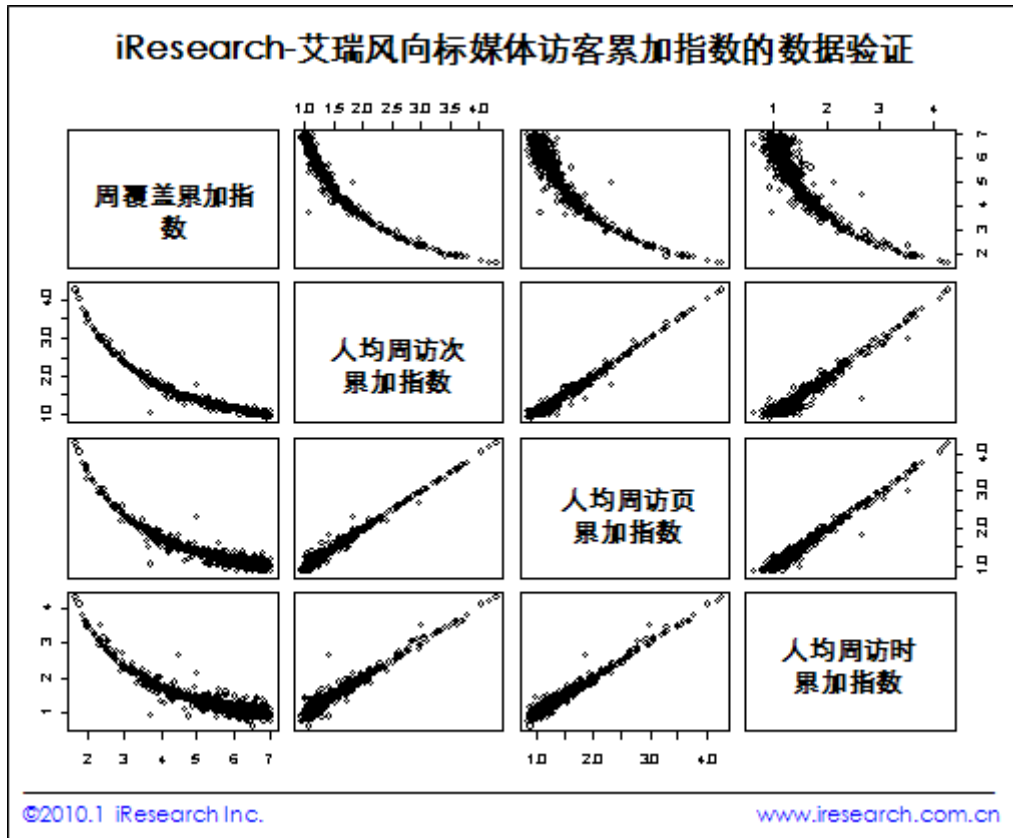


图 3- 2 iResearch-艾瑞风向标媒体访客累加指数的数据验证

在模型得以初步证明以后，我们可以根据公式 6) -9) 获得 4 个滚动系数。经检验，这 4 种方法获得系数基本一致。为了方便处理，我们以周覆盖累加指数计算出的滚动系数 α 为准，得到以下数据表格：（其中日回访比例就是模型中的滚动比例）

iUserTracker-2009 年 47 周主要网站回访比例

媒体			日回访比例	周度覆盖	周度访问次数
网站	域名	类别	%	万人	万次
腾讯	qq.com	综合门户	89.50%	18009.24	411992.00
百度	baidu.com	搜索引擎	84.35%	17920.63	288293.51
新浪	sina.com.cn	综合门户	71.03%	12374.72	98825.72
网易	163.com	综合门户	70.96%	9894.52	100511.30
搜狐	sohu.com	综合门户	64.16%	8584.87	71453.65
Google 中文	google.cn	搜索引擎	66.34%	7963.61	40336.35
搜搜	soso.com	搜索引擎	48.37%	7763.50	19000.60
优酷	youku.com	在线视频	56.08%	6529.30	35258.47
淘宝网	taobao.com	网上购物	72.55%	5757.51	134202.14
网址之家	hao123.com	网站导航	85.74%	5533.55	53378.08
土豆网	tudou.com	在线视频	49.72%	5459.76	24605.84
360 安全中心	360.cn	其他网站	65.77%	5303.64	22032.85
迅雷看看	xunlei.com	在线视频	53.49%	4695.50	23338.58
天涯虚拟社区	tianya.cn	社区博客	40.23%	4531.33	11584.34
中国雅虎	yahoo.com.cn	综合门户	50.95%	3610.95	14329.22
人人网	renren.com	交友社区	77.17%	3586.74	65889.11
狗狗影视	gougou.com	搜索引擎	41.54%	3444.64	10091.41
搜狗	sogou.com	搜索引擎	43.42%	3238.72	8843.54
Source: iUserTracker 2009 年 47 周的监测数据					
©2010.1 iResearch Inc.			www.iresearch.com.cn		

表 3-1 iUserTracker-2009 年 47 周主要网站回访比例

表中覆盖人数最多的几个大网站，他们的日回访比例都呈现显著的差异，所以这个日回访比例（滚动比例）能够有效的描述以前未被发现的一个网站特征。将这个根据周覆盖累加指数计算的日回访比例回代到其他的累加指数中，结果同样非常准确。

4. 目标用户集中度指数 TGI

反映目标网民用户在某一特定网站集中情况的参数就是目标用户集中度指数。业内常用的缩写为 TGI = Target group index。

11) $TGI = \text{某网站所有访问者中目标用户的比例} / \text{总体网民中目标用户的比例}$

要计算 TGI, 首先要确定目标用户, 这里以常见的人口属性来定义目标用户, 比如设: 25-40 岁的女性为目标用户。然后根据目标用户定义, 计算总体网民中这个人群的比例和不同媒体中这个人群的比例。如下表:

iUserTracker-2009 年 47 周主要网站目标用户 TGI

网站	网站类别	目标用户一周覆盖人数(万人)	总一周覆盖人数(万人)	目标用户一周覆盖 TGI
合计		4858.8	22173.8	-
qq.com [腾讯]	综合门户	3899.9	18009.2	98.8
baidu.com [百度]	搜索引擎	3870.4	17920.6	98.6
sina.com.cn [新浪]	综合门户	2675.3	12374.7	98.7
163.com [网易]	综合门户	2118.6	9894.5	97.7
sohu.com [搜狐]	综合门户	1850.7	8584.9	98.4
google.cn [Google 中文]	搜索引擎	1665.6	7963.6	95.4
soso.com [搜搜]	搜索引擎	1539.1	7763.5	90.5
taobao.com [淘宝网]	网上购物	1457.4	5757.5	115.5
hao123.com [网址之家]	网站导航	1228.3	5533.5	101.3
youku.com [优酷]	在线视频	1173.8	6529.3	82
tudou.com [土豆网]	在线视频	1027.5	5459.8	85.9
360.cn [360 安全中心]	其他网站	979.1	5303.6	84.3
tianya.cn [天涯虚拟社区]	社区博客	954.8	4531.3	96.2
msn.com.cn [MSN 中国]	综合门户	843.3	3171.5	121.4
yahoo.com.cn [中国雅虎]	综合门户	841	3611.0	106.3
kaixin001.com [开心网]	交友社区	824.8	2802.6	134.3
xunlei.com [迅雷看看]	在线视频	799.4	4695.5	77.7
ifeng.com [凤凰网]	综合门户	623.1	3097.0	91.8

Source: iUserTracker 2009 年 47 周的监测数据

©2010.1 iResearch Inc.

www.iresearch.com.cn

表 3-2 iUserTracker-2009 年 47 周主要网站目标用户 TGI

其中，目标受众一周覆盖 =
(网站目标受众一周覆盖 / 网站一周总覆盖) / (合计总目标受众数 / 合计总网民数)

5. 完整的网站用户访问特征评估体系

5.1. 六大指标的统计和计算方法定义

- 1) 网站日均覆盖人数
- 2) 网站日均回访比例
- 3) 网站人均单日访问次数
- 4) 网站人均单次浏览页面
- 5) 网站人均单页浏览时间
- 6) 目标用户集中度

当没有特别定义目标受众的时候，就使用 1-5 的指标，但针对目标受众分析的时候，需要将 1-5 的这些指标根据目标受众的定义重新计算，并加入第六个目标受众集中度指标。

以艾瑞的 iUserTracker 为例，对这些指标的具体统计解释和计算方法为：

- **日均覆盖人数 (Unique Visitors Per Day)**
该月中，该网站平均每天的独立访问用户数，用户重复访问不重复统计
日均覆盖人数=当周每日的网站覆盖人数的平均值
- **日均回访比例 (Revisit rate)**
该天访客中，属于本周老访客的比例（也等与 1-新访客比例）
日均回访比例=(7-周覆盖人数 / 日均覆盖人数) / 6
- **人均单日访问次数 (Visits Per User Per Visiting Day)**
平均每个用户，平均每个访问该网站日访问该网站的次数
人均单日访问次数=Average(该用户该周访问该网站的次数/该用户该周访问该网站的天数)
- **人均单次浏览页面数 (PV Per User Per Visit)**
平均每个用户，平均每次访问该网站的浏览页面数量
人均单次浏览页面数=Average(该用户该周该网站浏览页面数/该用户该周访问该网站的次数)
- **人均单页有效浏览时间 (Effective Viewing Duration Per User Per page)**
平均每个用户在浏览该网站每个页面时的有效浏览时间
注：单页打开时间小于 3 秒不计，单页停留 5 分钟以上，则视为用户脱机，不计其后的时间，直到新的网络行为发生
人均单页有效浏览时间=Average(该用户某时间段内该网站总浏览时间/该用户该时间段内该网站总浏览页面数)
- **目标受众集中度 (Target group index)**
TGI = 某网站所有访问者中目标用户的比例 / 总体网民中目标用户的比例

5.2. 媒体类型评估实例

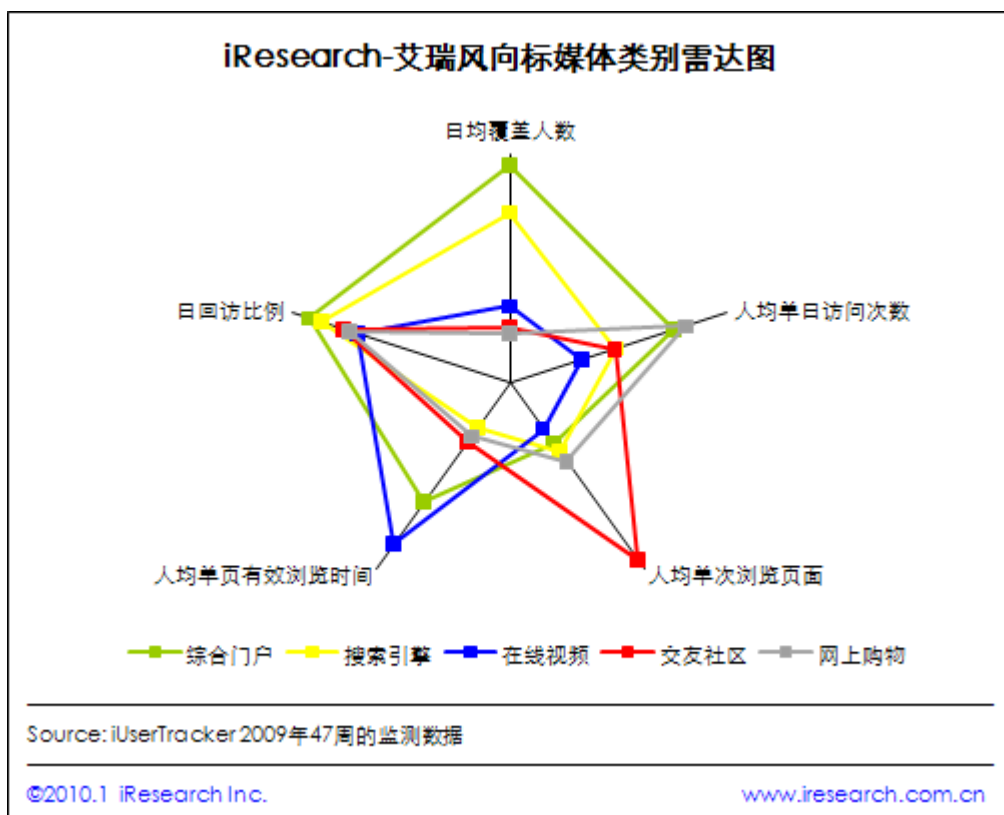


图 5-1 iResearch-艾瑞风向标媒体类别雷达图

如图，将 5 个指标放在雷达图里能够有效的发现他们的特征：

综合门户

除人均单次浏览页面外各指标均表现良好，单次浏览深度有限

搜索引擎

日回访比例高，单页有效浏览时间低

在线视频

单页有效浏览时间最高，但网站普及程度，用户浏览深度较弱

交友社区

回访比例和用户浏览深度都较为突出

网上购物

人均单日访问次数和回访比例都较高，但日均覆盖人数和单页浏览时间处于低水平

目标受众 TGI 的指标也可用同样方法加入分析。但因目标受众定义太广泛，这里不做举例。

法律声明

本报告为艾瑞市场咨询有限公司制作，报告中所有的文字、图片、表格均受到中国法律知识产权相关条例的版权保护。没有经过本公司书面许可，任何组织和个人，不得使用本报告中的信息用于其它商业目的。本报告中部分文字和数据采集于公开信息，所有权为原著者所有。没有经过原著者和本公司许可，任何组织和个人不得使用本报告中的信息用于其他商业目的。

本报告中运营商收入及相关市场预测主要为公司研究员采用行业访谈、市场调查、二手数据及其他研究方法分析获得，部分数据未经运营商直接认可。本报告中发布的调研数据部分采用样本调研方法，其数据结果受到样本的影响。由于调研方法及样本的限制，调查资料收集范围的限制，部分数据不能够完全反映真实市场情况。本报告只提供给购买报告的客户作为市场参考资料，本公司对该报告的数据准确性不承担法律责任。

公司服务介绍：

艾瑞咨询集团：<http://www.iresearch.com.cn>

公司资讯网站：

艾瑞连续数据服务 www.iresearch.com.cn/product/

艾瑞专项研究咨询服务 www.iresearch.com.cn/research/

网络广告监测系统 iAdTracker <http://www.iadtracker.com>

网民上网行为研究系统 iUserTracker <http://www.iusertracker.com>

艾瑞网 www.iresearch.cn

艾瑞会议服务 www.iresearch.com.cn/events/

报告购买/咨询：

艾瑞市场咨询有限公司 北京

地址：北京市朝阳区建国路 88 号 SOHO 现代城 C 座 1509 室，100022

电话：010-51283899-843 或 821

艾瑞市场咨询有限公司 上海

地址：上海市徐汇区漕溪北路 333 号中金国际广场 B 栋 701 室，200030

电话：021-51082699 -业务部

艾瑞市场咨询有限公司 广州

地址：广州市天河区林和西路 3-15 号耀中广场 707 室，510620

电话：020 -38010229

报告总顾问：

杨伟庆 iResearch 总裁

Email: henry@iresearch.com.cn