

We Know What You Want to Buy: A Demographic-based System for Product Recommendation On Microblogs

Wayne Xin Zhao¹, Yanwei Guo², Yulan He³,
Han Jiang², Yuexin Wu² and Xiaoming Li²

¹School of Information, Renmin University of China, Beijing,

²Peking University, Beijing, P.R.China

³School of Engineering and Applied Science, Aston University, UK



Background

- Recent years have witnessed a great success of e-commerce companies.



Product recommender systems play an important role to improve the sale of these companies

Typical challenges

- Challenge I:
 - Can only provide **onsite recommendation service**
 - Cannot capture users' instantaneous purchase intents outside those websites



A short status update (e.g. a tweet):

“Need your recommendation for my new phone!”

Current e-commerce recommender systems cannot capture such business opportunities.



Typical challenges

- Challenge II:
 - Know very little about “new users” (cold start)
 - Especially, it is hard to obtain important **demographic** information for recommendation



New users



Recommender
systems

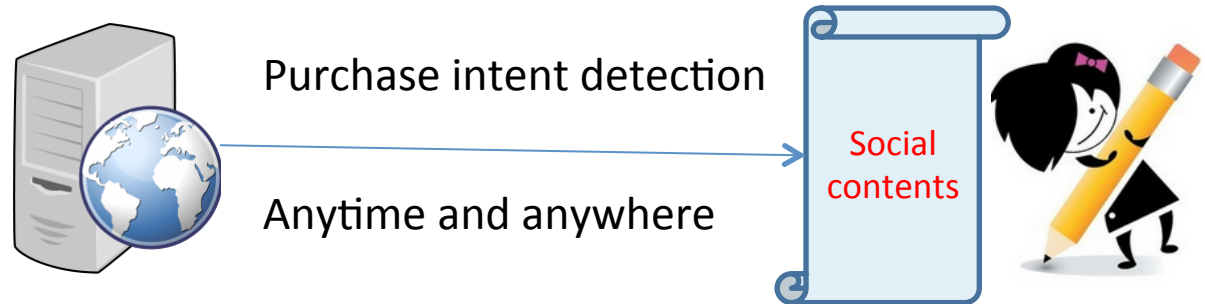


Age = ?
Sex = ?
Work = ?
Interests = ?
...

Our idea

- Online social media come to help
 - Important platforms where users **discuss their needs and desires**, and even disclose their purchase information

Users' purchase intents can be discovered from real-time status messages in online social media



Our idea

- Online social media come to help
 - Usually containing **profile information** of the users

Contain public profiles of users, especially rich demographic data



Age = 15
Sex = Female
Work = Student
Interests = Music
...



A new paradigm for e-commerce recommender system

- Social media based recommender systems
 - Embedded recommender system → Recommender systems on online social media



Social media accounts → Users' online identity

- Status update data : Monitor the status update messages for purchase intent detection
- Profile text data: Using profile information for user demographic modeling

Our system METIS

(A MERchanT Intent based recommendation System)



1.7 billion tweets from
5 million active users
within a half-year time
span

Social media platform

A linking bridge



Three product types: laptop,
camera and phone, containing
3,155 products and 1.13 million
user reviews.

E-commerce platform

We recommend products from e-commerce platform to users (in need) on microblogs

System demonstration



一个人的旅行陈少:

现在的世界没有, 没有公平, 没有公正, 没有信用可言.. 有钱有权就是王道.... 自己一定要告诉自己.. 凡事都要不择手段



无所谓2285663667:

【大连51个科技小微企业乐享项目扶持金】3月18日, 从市科技局获悉: 近日, 我市对51个科技小微企业项目发放了扶持资金。正如这一项目的名字“金种子工程”一样, 市科技局希望这些初创期的科技企业像一粒粒“金种子”, 在政策、资金、服务等“阳光、雨露”的滋润下成长为参天大树。



姓长名弓Zireal:

想买台苹果笔记本有什么推荐吗??

有商业意图



mathyYuan:

想入手佳能100D迷你单反了@爬楼梯的胖蜗牛 @

有商业意图



Laichuu-Szee:

不错, 你是行的! 你终究会滚开了。早点嘛[挖鼻屎][挖鼻屎][挖鼻屎]

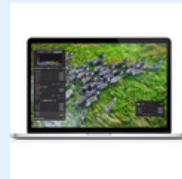


失眠怪物孙大莹:

iphone4s, 三星galaxy note2, 三星galaxy s3, 三星galaxy s4 换哪个好~ [疑问]

有商业意图

用户需求: 品牌: 苹果



苹果 (Apple) MacBook Pro ME665CH/A 15.4英寸宽屏笔记本电脑 (配备 Retina 显示屏)

商品评分: ★★★★★

商品价格: ¥20688 销量: 23

推荐理由: 男



苹果 (Apple) MacBook Pro MD212CH/A 13.3英寸宽屏笔记本电脑 (配备 Retina 显示屏)

商品评分: ★★★★★

商品价格: ¥10388 销量: 49

推荐理由: 男

高清屏就是爽 硬盘开机快



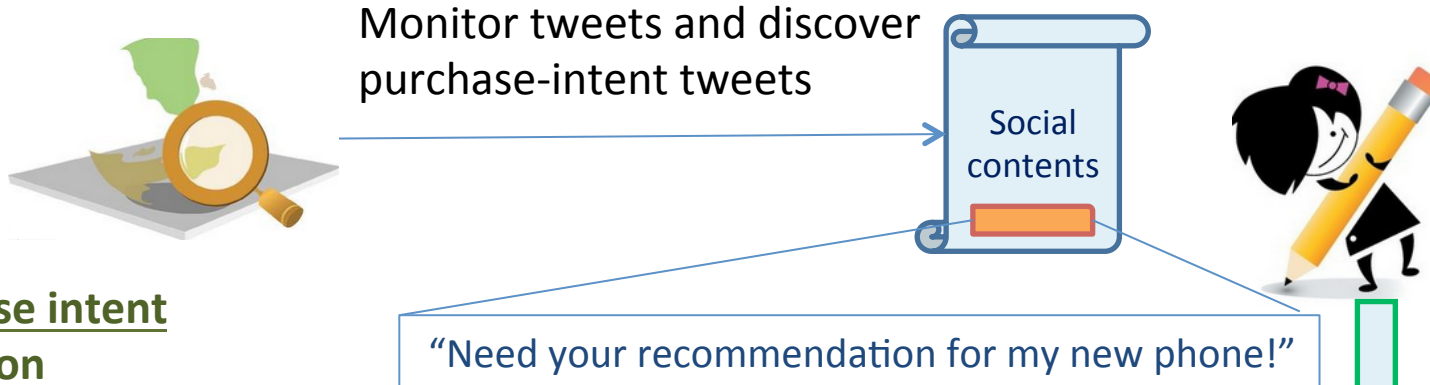
苹果 (Apple) MacBook Pro MD102CH/A 13.3英寸宽屏笔记本电脑

商品评分: ★★★★★

商品价格: ¥10388 销量: 242

速度快 东西不错 质量好

Overview of our system METIS



Purchase intent detection

The key idea is to represent users and products in the same dimensions of demographic attributes.

Need=A phone
Age = 15
Sex = Female
Interests=Music

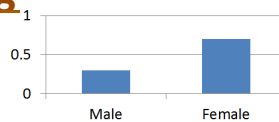
Product demographic learning



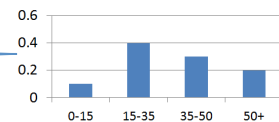
MICROBLOG



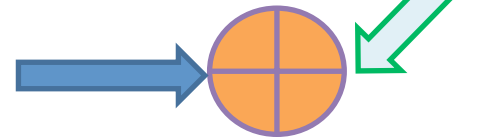
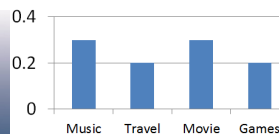
SEX



AGE



INTERESTS



Product Recommendation

Product Demographic Learning

- The product demographics, sometimes called the target audience, of a product or service is a collection of the characteristics of the people who buy that product or service.
 - Online review

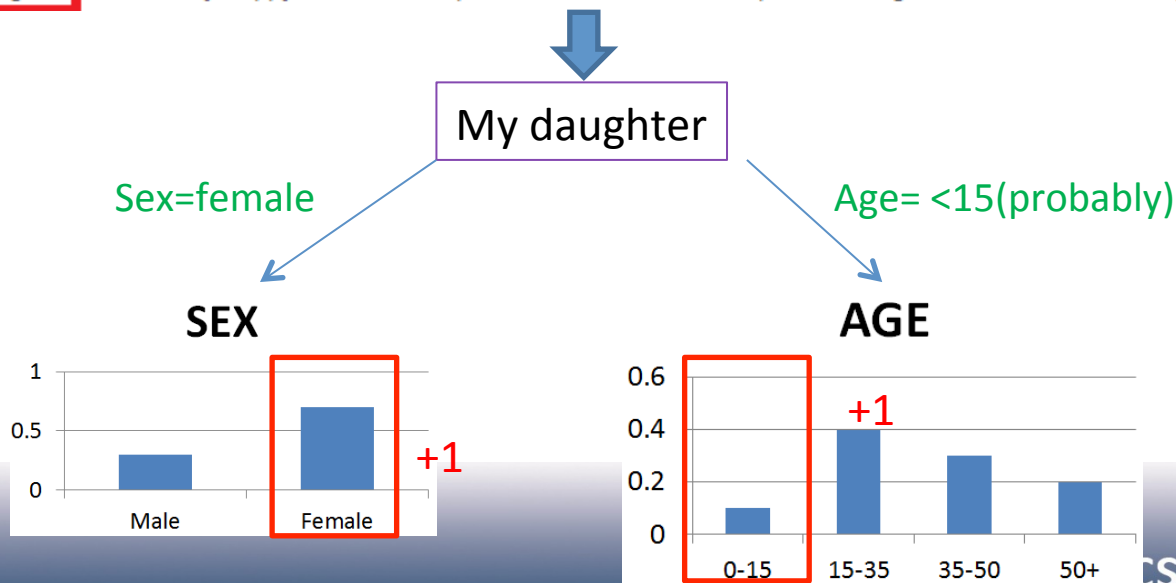
21 of 24 people found the following review helpful

★★★★★ Love it!

By [yourkm](#) on December 28, 2012

Verified Purchase

My daughter is extremely happy with her new phone. As advertised, the phone is in great condition exceeding my expectations; thank you!



Product Demographic Learning

- The product demographics, sometimes called the target audience, of a product or service is a collection of the characteristics of the people who buy that product or service.
 - Microblogs



We collect the demographics of the followers

Product Demographic Learning

- The product demographics, sometimes called the target audience, of a product or service is a collection of the characteristics of the people who buy that product or service.
 - **Microblogs**



小米之友 V: 三星Note4前期推广的第二条广告，从第一条表白到第二条结婚后，8月22号的第三条广告岂不是两人老了的故事，三星手机这两年的广告拍的还不错 @三星手机官网 <http://t.cn/RPYSYUU>

8月15日23:02 来自微博 weibo.com

👍 (10) | 转发(4) | 收藏 | 评论(2)

We collect the demographics of the microblogging users who have published the **positive comments** on a product

Product Recommendation Framework

- Learning to Rank
 - A user's need is viewed as a query (together with users' profile data)
 - A candidate product is viewed as a candidate document
 - The recommendation score can be reflected by the "relevance score"
 - Given a user's need, we rank products by their recommendation scores

Highlights:

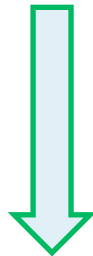
- Demographic based features + traditional features (e.g. ratings and sale)
- Semi-automatic acquisition of training data

Feature vector construction

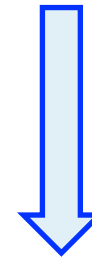
- Given a user and a candidate product



	Male	Female		
SEX	0.3	0.7		
	0-15	15-35	35-50	50+
AGE	0.1	0.4	0.3	0.2
	Music	Travel	Movie	Games
INTERESTS	0.3	0.2	0.3	0.2



[(SEX:0.7), (AGE:0.1), (INTERESTS:0.3)]



Profiles

Age = 15
Sex = Female
Interests = Music

Training data

- Semi-automatic acquisition of training data
 - Capturing the evidence of the user's self-disclosure

Oct 1, 2012 Please recommend! I want to buy my son a Samsung phone.

Oct 4, 2012 Done! I have bought Galaxy II for my lovely son!

Deployment

- A demonstration system (in Chinese)
 - With real SINA Weibo and Jingdong data
 - <http://sewm.pku.edu.cn/metis>

Case I

- A young girl wants to buy a new phone & does not have special requirements
 - White iPhone and Sumsung phones

 **妍霁Abby:** 有商业意图
[黑线]越来越觉得我把盖3当平板电脑用了~~现在连电脑都扔在单位不带回来了~~我又想买手机了!!!

 **苹果 (APPLE) iPhone 5 16G版 3G手机 (白色) WCDMA/GSM**
商品评分: ★★★★★
商品价格: **¥4648** 销量: **21035**
推荐理由: **18-29岁 女**

速度快 东西很好 价格便宜 价格贵 速度慢 电池不耐用

 **三星 S7562I 3G手机 (纯白) WCDMA/GSM 双卡双待**
商品评分: ★★★★★
商品价格: **¥748** 销量: **9574**
推荐理由: **18-29岁 女** 教育背景: **经济社科** 从事公共服务业

性价比高 速度快 外观漂亮 软件太多 电池不耐用 像素低

Case II

- A young boy wants to buy a new phone & does not have special requirements
 - Black iPhone and Lenovo phones

Color changes to black from white with sex changing from female to male.

 爱永辉的永晖:
疯了 换手机 求推荐[抓狂]

有商业意图

 苹果 (APPLE) iPhone 5 16G版 3G手机 (黑色) WCDMA/GSM

商品评分: ★★★★★

商品价格: ¥4648 销量: 13653

推荐理由: 18-29岁

速度快 东西很好 东西不错 价格贵 电池不耐用 速度慢

 联想 (Lenovo) A820T 3G手机 (深篷黑) TD-SCDMA/GSM 双卡双待

商品评分: ★★★★★

商品价格: ¥768 销量: 12845

推荐理由: 18-29岁 男

性价比高 速度快 屏幕大 软件太多 电池不耐用 姐夫说不好

Case III

- A young boy wants to buy his girl-friend a phone as the birthday gift
 - While phones specially designed for high-quality photographing

 xinda2004: 有商业意图
想给女朋友买个手机当生日礼物，送女朋友手机送哪款好？

 美图 MeituKiss 3G手机 (贝壳白) WCDMA/GSM
商品评分: ★★★★★
商品价格: ¥2199 销量: 98
推荐理由: 18-29岁 女
效果好

 vivo X3t 3G手机 (极光白) TD-SCDMA/GSM 双卡双待
商品评分: ★★★★★
商品价格: ¥2498 销量: 608
推荐理由: 18-29岁 女
速度快 外观漂亮 手感很好

 OPPO U705T 3G手机 (白色) TD-SCDMA/GSM
商品评分: ★★★★★
商品价格: ¥1998 销量: 230
推荐理由: 18-29岁 女
机壳就更好 方面都还不错 样子很漂亮

Translations from Chinese: “I would like to buy my gf a new phone as the birthday gift.”

Accurate identification of the real buying target!
Not the boy but his gf.

Case IV

- A young girl who wants to buy cheap laptops (dislike Apple products)

 **漫踏紫霜:** 有商业意图
想换个笔记本，又不想要苹果这么贵的，只想实惠一点便于携带的，大家有什么好的推荐吗？

 **华硕 (ASUS) X45EI237VD 14英寸笔记本 (i3-2370M 2G 500G 1G独显 D刻 USB3.0 黑色)**
商品评分: ★★★★★
商品价格: ¥2899 销量: 753
推荐理由: ✖
速度快 性价比高 京东越做越大

 **东芝 (TOSHIBA) L800-C57W1 14英寸笔记本电脑 (i3-3120M 4G 500G HD7670 1G独显 2*USB3.0 DOS) 白色**
商品评分: ★★★★★
商品价格: ¥2999 销量: 835
推荐理由: ✖
性价比高 外观漂亮 性价比不错

 **索尼 (SONY) SVF1421AYCW 14.0英寸笔记本电脑 (i5-3337U 4G 500G GT740 2G独显 D刻 Linux 白)**
商品评分: ★★★★★
商品价格: ¥4499 销量: 1023
推荐理由: ✖
外观漂亮 速度快 性价比高

In Chinese yuan

Summary

- The first public microblogging based product recommendation system in universities
- Learn product demographics from online reviews and microblogging data
- Automatic acquisition of the training data
- Demo systems and real service systems

Experiments

- Test collection

Table 4: Statistics of the dataset for product recommendation.

Types	#brands	#models	#query-decision pairs
phone	57	1,584	170
camera	25	724	496
laptop	25	829	437

- Evaluation metrics
 - NDCG
 - Precision @ k

Experiments

- Methods to compare
 - Simple baselines: sale, rating and polarity
 - Pointwise: MART , RandomForest (RF)
 - Pairwise: Ranksvm, RankBoost
 - Listwise: Listnet, AdaRank

Results

- Pointwise methods work very well, especially MART

Types	Metrics	Baselines			Pointwise		Pairwise		Listwise	
		sale	polarity	rating	MART	RF	RankSVM	RankBoost	Listnet	AdaRank
PHONE	p@1	0.006	0.006	0.006	0.106	0.041	0.018	0.012	0.006	0.006
	p@5	0.041	0.018	0.029	0.282	0.253	0.141	0.112	0.124	0.047
	NDCG@1	0.094	0.094	0.015	0.226	0.126	0.085	0.097	0.129	0.165
	NDCG@5	0.170	0.093	0.099	0.256	0.239	0.168	0.165	0.176	0.176
CAMERA	p@1	0.000	0.000	0.004	0.606	0.466	0.041	0.062	0.033	0.004
	p@5	0.238	0.004	0.018	0.688	0.688	0.183	0.310	0.152	0.244
	NDCG@1	0.128	0.128	0.008	0.641	0.515	0.092	0.140	0.073	0.127
	NDCG@5	0.254	0.094	0.061	0.495	0.466	0.146	0.232	0.115	0.253
LAPTOP	p@1	0.019	0.000	0.000	0.608	0.599	0.140	0.055	0.133	0.017
	p@5	0.026	0.014	0.012	0.653	0.665	0.439	0.102	0.302	0.026
	NDCG@1	0.239	0.002	0.002	0.695	0.686	0.145	0.228	0.138	0.220
	NDCG@5	0.212	0.073	0.059	0.573	0.559	0.223	0.219	0.172	0.202

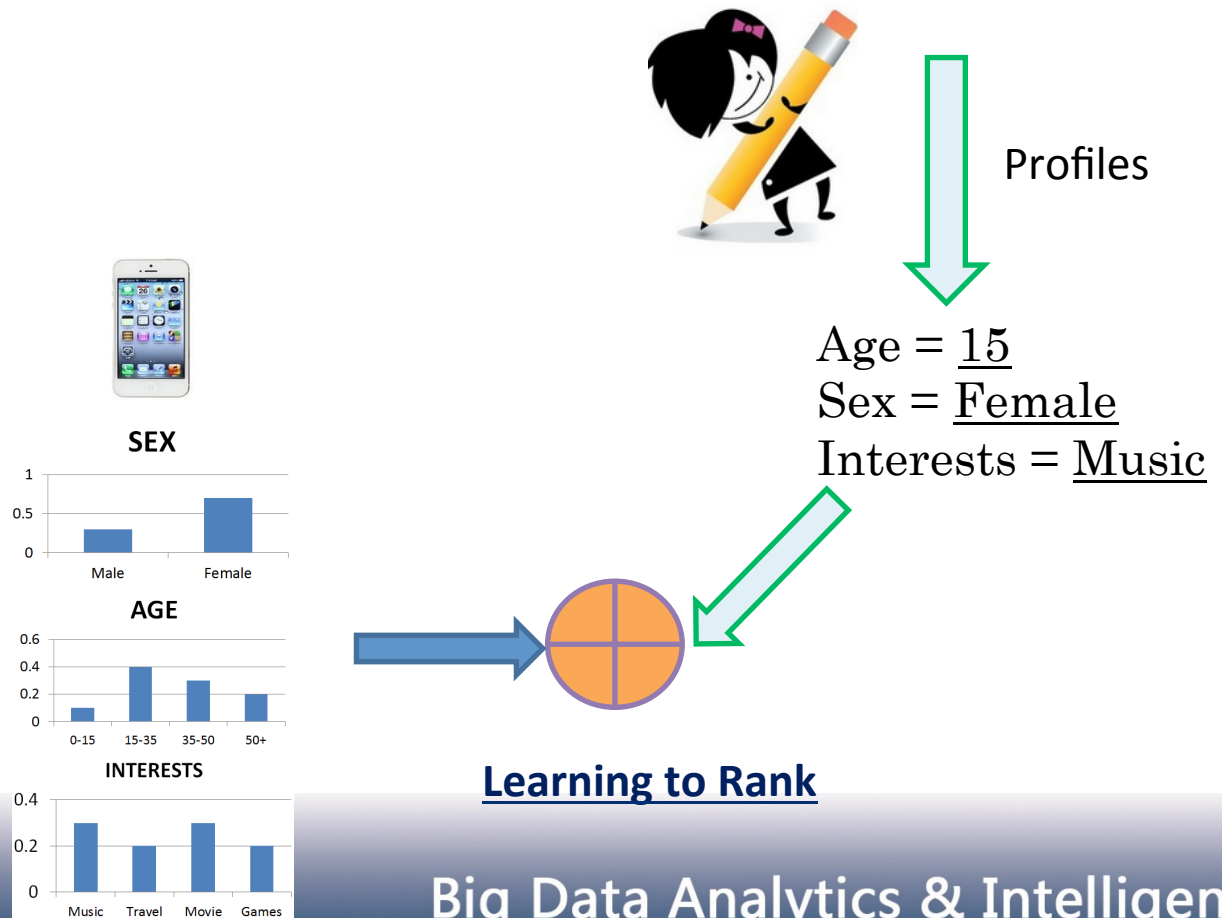
Product Recommendation Framework

- Candidate product generation
 - Manually generated rules
 - e.g. preferred price interval, preferred colors
 - Candidate product list pruning
 - At most 30 best-sale products

- Learning to Rank

Feature vector construction

- Given a user and a candidate product

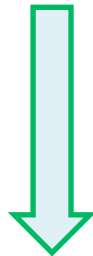


Feature vector construction

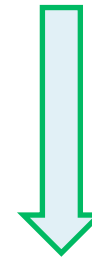
- Given a user and a candidate product



	Male	Female		
SEX	0.3	0.7		
	0-15	15-35	35-50	50+
AGE	0.1	0.4	0.3	0.2
	Music	Travel	Movie	Games
INTERESTS	0.3	0.2	0.3	0.2



[(SEX:0.7), (AGE:0.1), (INTERESTS:0.3)]



Profiles

Age = 15
Sex = Female
Interests = Music

Training data

- Semi-automatic acquisition of training data
 - Capturing the evidence of the user's self-disclosure

Oct 1, 2012 Please recommend! I want to buy my son a Samsung phone.

Oct 4, 2012 Done! I have bought Galaxy II for my lovely son!

Purchase Intent Detection



Purchase intent detection

Tweet streams

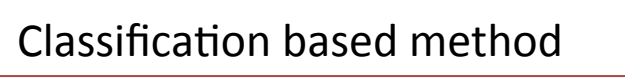
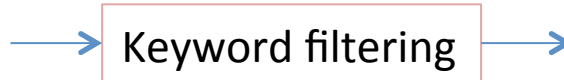


Table 2: List of purchase indicator keywords and their English translations.

buy (买), recommend (推荐), change (换), which is better (哪个更好), cheap (便宜), cost (价值), auction (拍), on sale (降价), price (价格), need (需要), shopping (购物)

Table 3: Performance comparison for purchase intent detection with SVM.

Kernel	Methods	Precision	Recall	F-value
Linear	<i>Baseline_T</i>	0.738	0.771	0.754
	<i>Baseline_{T+D}</i>	0.743	0.788	0.765
	<i>Ours_T</i>	0.762	0.802	0.782
	<i>Ours_{T+D}</i>	0.77	0.806	0.788
RBF	<i>Baseline_T</i>	0.81	0.582	0.678
	<i>Baseline_{T+D}</i>	0.725	0.726	0.725
	<i>Ours_T</i>	0.796	0.829	0.812
	<i>Ours_{T+D}</i>	0.805	0.832	0.818

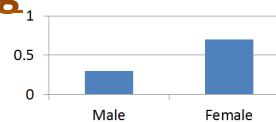
Product Demographic Learning

- The **product demographics**, sometimes called **the target audience**, of a product or service is a collection of the characteristics of the people who buy that product or service.
 - Modeled as distributions over attribute values

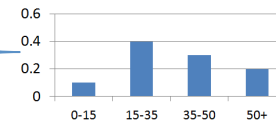
Product demographic learning



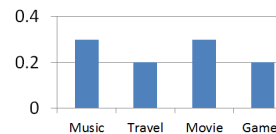
SEX



AGE



INTERESTS



Demographic attributes

Table 1: List of demographic attributes.

Attribute	Values
Gender	male, female
Age	1-11, 12-17, 18-30, 31-45, 46-59, 60+
Marital Status	single, engaged, loving secretly, married, relationship seeking, bereft of one's spouse, separated, divorced, ambiguous, loving
Education	literature, natural science, engineering, social sciences, medical science, art, others
Career	internet technology, designing, media, service industry, manufacturing, medicine, scientific research, management, others
Interests (Weibo tags)	travel, photographing, music and movie, computer games, Internet surfing, other

Features

- Query dependent features
 - Demographic based features
- Query independent features
 - Sale
 - Rating
 - Opinion polarity score based on text

Samples of the learnt product demographics based on online reviews

Table 7: Samples of the learnt product demographics based on online reviews. Real numbers denote the learned weights for the corresponding attribute values.

Galaxy S4 (White)	$(SEX, [“male”, 0.271], [“female”, 0.729])$
Galaxy S4 (Blue)	$(SEX, [“male”, 0.688], [“female”, 0.312])$
Galaxy S4 (Black)	$(SEX, [“male”, 0.852], [“female”, 0.148])$
Galaxy S4 (White)	$(Age, [“ < 45”, 0.931], [“ ≥ 45”, 0.069])$
Galaxy S4 (Blue)	$(Age, [“ < 45”, 0.755], [“ ≥ 45”, 0.245])$
Galaxy S4 (Black)	$(Age, [“ < 45”, 0.650], [“ ≥ 45”, 0.350])$

Young females prefer white phones while young males like black phones more.

Samples of the learnt product demographic based on microblogs

Table 8: Samples of the learnt product demographic based on microblogs.

Apple	(<i>SEX</i> , [<i>“male”</i> , 0.593], [<i>“female”</i> , 0.407])
	(<i>CAREER</i> , [<i>“IT”</i> , 0.28], [<i>“management”</i> , 0.219], [<i>“media”</i> , 0.172], [<i>“industry”</i> , 0.139])
	(<i>TAG</i> , [<i>“music&movie”</i> , 0.316], [<i>“travel”</i> , 0.249], [<i>“Internet surfing”</i> , 0.163], [<i>“computer games”</i> , 0.161])
Samsung	(<i>SEX</i> , [<i>“male”</i> , 0.503], [<i>“female”</i> , 0.497])
	(<i>CAREER</i> , [<i>“management”</i> , 0.252], [<i>“IT”</i> , 0.223], [<i>“industry”</i> , 0.252], [<i>“media”</i> , 0.223])
	(<i>TAG</i> , [<i>“computer games”</i> , 0.281], [<i>“travel”</i> , 0.27], [<i>“music&movie”</i> , 0.209], [<i>“Internet surfing”</i> , 0.188])

- 1) Samsung has a more balanced sex distribution;
- 2) Apple products are more preferred by the consumers in the IT field.