TASK目标：通过本案例联系掌握两个重要的技巧：

（1）以分词为核心的中文文本处理能力；

（2）以word-2-vector模型为核心的文本量化能力。并在此基础上形成任意模型。用50%数据做训练，50%数据做验证，汇报完整建模过程，以及最终的外样本预测精度。

具体要求：

（1）所有代码详细注释；

（2）如果（不是必须）涉及深度学习模型，需要清点模型参数个数；

（3）外样本预测精度不用Mean Sqared Error（因为可能有大量异常值），而用Median Squared Error (MedSE)；

（4）实验重复100次，产生100个MedSE值，对其做箱线图和直方图，并汇报中位数；

（5）最后得分为：50%过程分+50%预测精度得分。

Word Embedding：文本语义的数学表达

如何理解词与词的相关性？

> 词与词的相关性可以有多种不同的理解，比如：（1）两个词经常同时出现，说明它们之间的相关性较强，像“周末+加班”；（2）两个词可能不会同时出现，但是在“语义”上非常相似，比如“苹果”和“橘子”，都是水果的一种，“酒店”和“宾馆”表示的同一个事物

如何判断两个词在语义上相似呢？

> 语义上的相关性可以通过词语在“上下文的位置”来判断：如果两个词经常出现在相似的上下文中，那两个词的语义比较相似

语义相似性的“直观”表示

> 假设某种方法可以将词语的语义直观的展示在坐标轴中，那语义相似的词语会靠的比较近

从自然语言到数学语言

> 为了根据语义刻画词语在坐标轴的位置，就需要将“自然语言”转化为“数学语言”，即给每个词语一组向量表示
>
> 这种用“向量”表示“语义”的方法，就是Word Embedding

数据介绍

我们使用的数据是携程网上的7000多条酒店评论数据

下载地址：https://github.com/SophonPlus/ChineseNlpCorpus/blob/master/datasets/ChnSentiCorp_htl (https://github.com/SophonPlus/ChineseNlpCorpus/blob/master/datasets/ChnSentiCorp_htl_all/

数据读入

In [1]:

```
import pandas as pd    # 导入pandas库
data = pd.read_csv('data/Comment.csv')  # 导入酒店评论数据
data.head()   # 查看前5行
```

Out[1]:

	label	review
0	1	距离川沙公路较近,但是公交指示不对,如果是"蔡陆线"的话,会非常麻烦.建议用别的路线.房间较...
1	1	商务大床房，房间很大，床有2M宽，整体感觉经济实惠不错!
2	1	早餐太差，无论去多少人，那边也不加食品的。酒店应该重视一下这个问题了。房间本身很好。
3	1	宾馆在小街道上，不大好找，但还好北京热心同胞很多～宾馆设施跟介绍的差不多，房间很小，确实挺小...
4	1	CBD中心,周围没什么店铺,说5星有点勉强.不知道为什么卫生间没有电吹风

中文分词

有可能您的计算环境中，没有安装jieba这个包包。

这会让您在下面的代码执行错误。

如果发生该现象，请在终端（小黑窗）输入执行下面两句代码

conda activate py37_tensorflow

pip install jieba

In [2]:

```
import jieba   # 导入分词软件
train_data = []  # 建立空列表存数据
for line in data.review:    # 取某一行
    line_fenci = jieba.lcut(line)  # 分词
    train_data.append(line_fenci)  # 附加空列表中
```

```
Building prefix dict from the default dictionary ...
Loading model from cache /tmp/jieba.cache
```

Loading model cost 0.662 seconds.
Prefix dict has been built successfully.

分词之后的结果长什么样？

In [3]:

```
print(train_data[0])   # 打印第一个列表
print(train_data[1])   # 打印第二个列表
print(train_data[2])   # 打印第三个列表
```

['距离', '川沙', '公路', '较近', ',', '但是', '公交', '指示', '不', '对', ',', '如果', '是', '"', '蔡陆线', '"', '的话', ',', '会', '非常', '麻烦', '.', '建议', '用', '别的', '路线', '.', '房间', '较为简单', '.']
['商务', '大床', '房', '，', '房间', '很大', '，', '床有', '2M', '宽', '，', '整体', '感觉', '经济', '实惠', '不错', '!']
['早餐', '太', '差', '，', '无论', '去', '多少', '人', '，', '那边', '也', '不加', '食品', '的', '。', '酒店', '应该', '重视', '一下', '这个', '问题', '了', '。', '房间', '本身', '很', '好', '。']

Word Embedding - 基本原理

> Word Embedding （词嵌入），是指将词语“嵌入”到一个共同的空间中进行比较
>
> 为了将词语嵌入到某个空间中，就需要为每个词语“分配”一个“词向量”
>
> 根据上下文的词来预测目标词出现的概率

Word Embedding 程序实现

In [4]:

```
from gensim.models import Word2Vec   # 导入函数
# size表示词向量的大小，min_count表示最少的词频，低于此词频的词会被忽略掉
model = Word2Vec(train_data, size=100, min_count=1)  # 模型输入
```

展示学习好的词向量

In [5]:

```
print(len(model['酒店']))   # 打印拟合元素为“酒店”的模型长度
model['酒店']               # 输出该元素内容
```

100

/anaconda/envs/py37_tensorflow/lib/python3.7/site-packages/ipykernel_launcher.py:1: DeprecationWarning: Call to deprecated `__getitem__` (Method will be removed in 4.0.0, use self.wv.__getitem__() instead).
  """Entry point for launching an IPython kernel.
/anaconda/envs/py37_tensorflow/lib/python3.7/site-packages/ipykernel_launcher.py:2: DeprecationWarning: Call to deprecated `__getitem__` (Method will be removed in 4.0.0, use self.wv.__getitem__() instead).

Out[5]:

```
array([ 0.4779667 , -0.55352837,  0.5858705 , -0.02616989, -0.85960567,
        0.63875896, -0.45202884,  0.15306237, -0.65110713, -0.21434826,
       -0.6711334 ,  0.2794752 ,  2.1497152 , -1.1701213 ,  0.23793352,
```

```
 0.36066642,  0.3559573 ,  0.8259308 ,  0.47765446,  0.35968092,
-1.15584   ,  0.37555584,  0.60960597,  0.06753001,  0.3175832 ,
-0.98576957,  0.8956984 ,  1.0523698 , -1.3979478 , -0.24714908,
 0.994285  , -1.2568508 , -0.23182367,  1.6792676 , -0.06917895,
 0.4414267 , -1.4733431 ,  1.0075239 ,  1.5337375 ,  1.447736  ,
-0.30351737, -0.4038807 , -1.2758311 ,  0.98617756,  0.4804772 ,
 0.08010791,  0.6266053 ,  1.4158839 ,  0.7996716 ,  0.31318828,
-0.26011762, -2.0014722 , -0.57217383,  0.28269264,  0.8041437 ,
-0.3136659 ,  0.6074194 , -1.4109305 , -0.33846793,  0.71626896,
 0.6720079 ,  0.9789441 , -0.04262506, -0.8094707 ,  0.5430507 ,
-0.6425928 , -0.46381074,  0.24475147, -1.4258884 ,  1.5469421 ,
-0.18152206,  0.68057823,  0.524647  ,  0.79342824,  0.87329197,
-0.19430718, -0.2277586 , -0.5902611 ,  1.4448286 , -0.2622007 ,
 0.41104245,  0.8634458 ,  0.42193526, -0.5343201 , -0.95811987,
 0.69789547, -0.41252366, -0.1377625 , -0.3477388 , -1.1970644 ,
-0.43051538, -0.40487313,  1.8991055 ,  0.4832525 ,  0.535456  ,
-0.5014756 ,  0.21552503,  1.4380987 , -0.76125795,  1.299436  ],
dtype=float32)
```

Word Embedding应用：判断任意两个词的相似度

In [6]:

```
from scipy import stats  # 导入库
import numpy as np

x1=model['酒店']  # 模型拟合
x2=model['饭店']  # 模型拟合
x3=model['不错']  # 模型拟合
x4=model['挺不错'] # 模型拟合

print(model.similarity('酒店', '饭店'))  # 计算2词直接的相似性
print(model.similarity('酒店', '不错'))  # 计算2词直接的相似性
print(model.similarity('挺不错', '不错'))  # 计算2词直接的相似性
r=np.mean(x3*x4)/np.sqrt(np.mean(x3**2)*np.mean(x4**2))  # 计算

model.similar_by_word("不错")  # 输出与“不错”最相似的结果
```

```
0.7939849
0.53584844
0.8970798

/anaconda/envs/py37_tensorflow/lib/python3.7/site-packages/ipykernel_launcher.py:4: DeprecationWarning: Call to deprecated `__getitem__` (Method will be removed in 4.0.0, use self.wv.__getitem__() instead).
  after removing the cwd from sys.path.
/anaconda/envs/py37_tensorflow/lib/python3.7/site-packages/ipykernel_launcher.py:5: DeprecationWarning: Call to deprecated `__getitem__` (Method will be removed in 4.0.0, use self.wv.__getitem__() instead).
  """
/anaconda/envs/py37_tensorflow/lib/python3.7/site-packages/ipykernel_launcher.py:6: DeprecationWarning: Call to deprecated `__getitem__` (Method will be removed in 4.0.0, use self.wv.__getitem__() instead).

/anaconda/envs/py37_tensorflow/lib/python3.7/site-packages/ipykernel_launcher.py:7: DeprecationWarning: Call to deprecated `__getitem__` (Method will be removed in 4.0.0, use self.wv.__getitem__() instead).
  import sys
/anaconda/envs/py37_tensorflow/lib/python3.7/site-packages/ipykernel_launcher.py:9: DeprecationWarning: Call to deprecated `similarity` (Method will be removed in 4.0.0, use self.wv.similarity() instead).
  if __name__ == '__main__':
/anaconda/envs/py37_tensorflow/lib/python3.7/site-packages/ipykernel_launcher.py:10: DeprecationWarning: Call to deprecated `similarity` (Method will be removed in 4.0.0, use self.wv.similarity() instead).
  # Remove the CWD from sys.path while we load stuff.
/anaconda/envs/py37_tensorflow/lib/python3.7/site-packages/ipykernel_launcher.py:11: DeprecationWarning: C
```

all to deprecated `similarity` (Method will be removed in 4.0.0, use self.wv.similarity() instead).
  # This is added back by InteractiveShellApp.init_path()
/anaconda/envs/py37_tensorflow/lib/python3.7/site-packages/ipykernel_launcher.py:14: DeprecationWarning: C
all to deprecated `similar_by_word` (Method will be removed in 4.0.0, use self.wv.similar_by_word() instead).

Out[6]:

```
[('挺不错', 0.8970798254013062),
 ('比较满意', 0.8858372569084167),
 ('行', 0.885407567024231),
 ('一般', 0.869199812412262),
 ('不怎么样', 0.8678579330444336),
 ('稍差', 0.8634871244430542),
 ('快餐', 0.86090087890625),
 ('合理', 0.8560066223144531),
 ('说得过去', 0.8538315296173096),
 ('还好', 0.8505702018737793)]
```

NLP - Personalized Recommendation

In [1]:

```
import pandas as pd
from gensim.models import Word2Vec
import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
```

读入数据retail.csv，其中有以下注意事项：

（1）有不少的销售记录没有customerid，因此前后购买记录没法建立联系，需要删除（建议尝试：sum(records.customerid.isnull())）

（2）每一个customerid的数据类型是float65，不好处理，因此需要强制编程整数（建议尝试：type(records.customerid[0])）

In [2]:

```
records = pd.read_csv("F:/大三（上）/深度学习/TASK6.1：Wod-2-Vector - SEM关键词展现量预测/retail.csv")
keepid=records.customerid.notnull()    # 删除空值
records=records[keepid]            # 提取元素
records['customerid'] = records['customerid'].map(int)  # 转化数据类型
records
```

Out[2]:

	customerid	itemid	description	price
0	13085	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	6.95
1	13085	79323P	PINK CHERRY LIGHTS	6.75
2	13085	79323W	WHITE CHERRY LIGHTS	6.75
3	13085	22041	RECORD FRAME 7" SINGLE SIZE	2.10
4	13085	21232	STRAWBERRY CERAMIC TRINKET BOX	1.25
...	...	...	...	...

	customerid	itemid	description	price
525456	17530	22271	FELTCRAFT DOLL ROSIE	2.95
525457	17530	22750	FELTCRAFT PRINCESS LOLA DOLL	3.75
525458	17530	22751	FELTCRAFT PRINCESS OLIVIA DOLL	3.75
525459	17530	20970	PINK FLORAL FELTCRAFT SHOULDER BAG	3.75
525460	17530	21931	JUMBO STORAGE BAG SUKI	1.95

417534 rows × 4 columns

创建一个关于itemid和description的字典

In [3]:

```
product_dictionary=dict(zip(records.itemid,records.description))  # 打包字典
M=len(product_dictionary)  # 计算长度
product_dictionary     # 打印字典
```

Out[3]:

{'85048': '15CM CHRISTMAS GLASS BALL 20 LIGHTS',
 '79323P': 'PINK CHERRY LIGHTS',
 '79323W': 'WHITE CHERRY LIGHTS',
 '22041': 'RECORD FRAME 7" SINGLE SIZE ',
 '21232': 'STRAWBERRY CERAMIC TRINKET BOX',
 '22064': 'PINK DOUGHNUT TRINKET POT ',
 '21871': 'SAVE THE PLANET MUG',
 '21523': 'DOORMAT FANCY FONT HOME SWEET HOME',
 '22350': 'ILLUSTRATED CAT BOWL ',
 '22349': 'DOG BOWL CHASING BALL DESIGN',
 '22195': 'LARGE HEART MEASURING SPOONS',
 '22353': 'LUNCH BOX WITH CUTLERY FAIRY CAKES ',
 '48173C': 'DOORMAT BLACK FLOCK ',
 '21755': 'LOVE BUILDING BLOCK WORD',
 '21754': 'HOME BUILDING BLOCK WORD',
 '84879': 'ASSORTED COLOUR BIRD ORNAMENT',
 '22119': 'PEACE WOODEN BLOCK LETTERS',
 '22142': 'CHRISTMAS CRAFT WHITE FAIRY ',
 '22296': 'HEART IVORY TRELLIS LARGE',
 '22295': 'HEART FILIGREE DOVE LARGE',
 '22109': 'FULL ENGLISH BREAKFAST PLATE',
 '22107': 'PIZZA PLATE IN BOX',
 '22194': 'BLACK DINER WALL CLOCK',
 '35004B': 'SET OF 3 BLACK FLYING DUCKS',
 '82582': 'AREA PATROLLED METAL SIGN',
 '21181': 'PLEASE ONE PERSON METAL SIGN',
 '21756': 'BATH BUILDING BLOCK WORD',
 '21333': 'CLASSIC WHITE FRAME',
 '84596F': 'SMALL MARSHMALLOWS PINK BOWL',
 '84596L': 'BISCUITS SMALL BOWL LIGHT BLUE',
 '22111': 'SCOTTIE DOG HOT WATER BOTTLE',
 '22143': 'CHRISTMAS CRAFT HEART DECORATIONS',
 '22145': 'CHRISTMAS CRAFT HEART STOCKING ',
 '22130': 'PARTY CONE CHRISTMAS DECORATION ',
 '21364': 'PEACE SMALL WOOD LETTERS',
 '21360': 'JOY LARGE WOOD LETTERS',
 '21351': 'CINAMMON & ORANGE WREATH',
 '21352': 'EUCALYPTUS & PINECONE WREATH',
 '35400': 'WOODEN BOX ADVENT CALENDAR ',
 '20695': 'FLORAL BLUE MONSTER',
 '37370': 'RETRO COFFEE MUGS ASSORTED',

'10002': 'INFLATABLE POLITICAL GLOBE ',
'84507B': 'STRIPES DESIGN MONKEY DOLL',
'20703': 'BLUE PADDED SOFT MOBILE',
'21987': 'PACK OF 6 SKULL PAPER CUPS',
'21989': 'PACK OF 20 SKULL PAPER NAPKINS',
'84970S': 'HANGING HEART ZINC T-LIGHT HOLDER',
'20971': 'PINK BLUE FELT CRAFT TRINKET BOX',
'22271': 'FELTCRAFT DOLL ROSIE',
'22272': 'FELTCRAFT DOLL MARIA',
'22274': 'FELTCRAFT DOLL EMILY',
'21912': 'VINTAGE SNAKES & LADDERS',
'22112': 'CHOCOLATE HOT WATER BOTTLE',
'21329': 'DINOSAURS WRITING SET ',
'21252': 'SET OF MEADOW FLOWER STICKERS',
'21100': 'CHARLIE AND LOLA CHARLOTTE BAG',
'21033': 'JUMBO BAG CHARLIE AND LOLA TOYS',
'20711': 'JUMBO BAG TOYS ',
'21410': 'COUNTRY COTTAGE DOORSTOP GREEN',
'21411': 'GINGHAM HEART DOORSTOP RED',
'84031A': 'CHARLIE+LOLA RED HOT WATER BOTTLE ',
'84031B': 'CHARLIE LOLA BLUE HOT WATER BOTTLE ',
'84032A': 'CHARLIE+LOLA PINK HOT WATER BOTTLE',
'84032B': 'CHARLIE + LOLA RED HOT WATER BOTTLE',
'84519A': 'TOMATO CHARLIE+LOLA COASTER SET',
'84519B': 'CARROT CHARLIE+LOLA COASTER SET',
'85132A': 'CHARLIE + LOLA BISCUITS TINS',
'85132C': 'CHARLIE AND LOLA FIGURES TINS',
'85183A': 'CHARLIE & LOLA WASTEPAPER BIN BLUE',
'85183B': 'CHARLIE & LOLA WASTEPAPER BIN FLORA',
'22065': 'CHRISTMAS PUDDING TRINKET POT ',
'22138': 'BAKING SET 9 PIECE RETROSPOT ',
'22139': 'RETROSPOT TEA SET CERAMIC 11 PC ',
'22352': 'LUNCH BOX WITH CUTLERY RETROSPOT ',
'85014A': 'BLACK/BLUE POLKADOT UMBRELLA',
'85014B': 'RED RETROSPOT UMBRELLA',
'16161P': 'WRAP ENGLISH ROSE ',
'16169N': 'WRAP BLUE RUSSIAN FOLKART',
'21491': 'SET OF THREE VINTAGE GIFT WRAPS',
'22333': 'RETROSPOT PARTY BAG + STICKER SET',
'85216': 'ASSORTED CAKES FRIDGE MAGNETS',
'21493': 'VINTAGE DESIGN GIFT TAGS',
'21731': 'RED TOADSTOOL LED NIGHT LIGHT',
'85232B': 'SET OF 3 BABUSHKA STACKING TINS',
'84691': 'PACK 20 DOLLY PEGS',
'20749': 'ASSORTED COLOUR MINI CASES',
'POST': 'POSTAGE',
'22321': 'BIRD DECORATION RED RETROSPOT',
'84029E': 'RED WOOLLY HOTTIE WHITE HEART.',
'21955': 'DOORMAT UNION JACK GUNS AND ROSES',
'84899E': 'YELLOW BROWN BEAR FELT PURSE KIT',
'85123A': 'WHITE HANGING HEART T-LIGHT HOLDER',
'21582': 'KINGS CHOICE SMALL TUBE MATCHES',
'21586': 'KINGS CHOICE GIANT TUBE MATCHES',
'21590': 'KINGS CHOICE CIGAR BOX MATCHES ',
'22091': 'EMPIRE TISSUE BOX',
'84251J': 'GREETING CARD, THE BLACK CAT',
'22024': 'RAINY LADIES BIRTHDAY CARD',
'22026': 'BANQUET BIRTHDAY CARD ',
'22029': 'SPACEBOY BIRTHDAY CARD',
'22025': 'RING OF ROSES BIRTHDAY CARD',
'22031': 'BOTANICAL LAVENDER BIRTHDAY CARD',
'21428': 'SET3 BOOK BOX GREEN GINGHAM FLOWER ',

'21430': 'SET/3 RED GINGHAM ROSE STORAGE BOX',
'21916': 'SET 12 RETRO WHITE CHALK STICKS',
'21790': 'VINTAGE SNAP CARDS',
'21888': 'BINGO SET',
'85178': 'VICTORIAN SEWING KIT',
'20754': 'RETROSPOT RED WASHING UP GLOVES',
'21035': 'SET/2 RED RETROSPOT TEA TOWELS ',
'82001S': 'VINYL RECORD FRAME SILVER',
'82580': 'BATHROOM METAL SIGN',
'85150': 'LADIES & GENTLEMEN METAL SIGN',
'21041': 'RED RETROSPOT OVEN GLOVE DOUBLE',
'35916C': 'PINK FELT HANGING HEART W FLOWER',
'35916B': 'BLUE FELT HANGING HEART W FLOWER',
'22180': 'RETROSPOT LAMP',
'22128': 'PARTY CONES CANDY ASSORTED',
'22163': 'HEART STRING MEMO HOLDER HANGING',
'35951': 'FOLKART BAUBLE CHRISTMAS DECORATION',
'22253': '6 CROCHET STRAWBERRIES',
'84988': 'SET OF 72 PINK HEART PAPER DOILIES',
'22219': 'LOVEBIRD HANGING DECORATION WHITE ',
'18097C': 'WHITE TALL PORCELAIN T-LIGHT HOLDER',
'21299': 'EDWARDIAN TOILET ROLL UNIT ',
'21099': 'SET/6 STRAWBERRY PAPER CUPS',
'21085': 'SET/6 WOODLAND PAPER CUPS',
'21929': 'JUMBO BAG PINK VINTAGE PAISLEY',
'22241': 'GARLAND WOODEN HAPPY EASTER',
'21091': 'SET/6 WOODLAND PAPER PLATES',
'21733': 'RED HANGING HEART T-LIGHT HOLDER',
'22178': 'VICTORIAN GLASS HANGING T-LIGHT',
'22125': 'UNION JACK HOT WATER BOTTLE ',
'21671': 'RED SPOT CERAMIC DRAWER KNOB',
'22086': "PAPER CHAIN KIT 50'S CHRISTMAS ",
'22294': 'HEART FILIGREE DOVE SMALL',
'22199': 'FRYING PAN RED RETROSPOT',
'22083': 'PAPER CHAIN KIT RETROSPOT',
'22114': 'HOT WATER BOTTLE TEA AND SYMPATHY',
'21485': 'RETROSPOT HEART HOT WATER BOTTLE',
'21928': 'JUMBO BAG SCANDINAVIAN PAISLEY',
'85099B': 'JUMBO BAG RED RETROSPOT',
'22356': 'CHARLOTTE BAG PINK POLKADOT',
'20724': 'RED RETROSPOT CHARLOTTE BAG',
'21592': 'RETROSPOT CIGAR BOX MATCHES ',
'20725': 'LUNCH BAG RED RETROSPOT',
'20728': 'LUNCH BAG CARS BLUE',
'20726': 'LUNCH BAG WOODLAND',
'21217': 'RED RETROSPOT ROUND CAKE TINS',
'85066': 'CREAM SWEETHEART MINI CHEST',
'20827': 'GOLD APERITIF GLASS',
'20825': 'GOLD WINE GLASS',
'20823': 'GOLD WINE GOBLET',
'20826': 'SILVER APERITIF GLASS',
'22087': 'PAPER BUNTING WHITE LACE',
'85206A': 'CREAM FELT EASTER EGG BASKET',
'21895': "POTTING SHED SOW 'N' GROW SET",
'21896': 'POTTING SHED TWINE',
'84946': 'ANTIQUE SILVER TEA GLASS ETCHED',
'22090': 'PAPER BUNTING RETROSPOT',
'90200A': 'PURPLE SWEETHEART BRACELET',
'90200D': 'PINK SWEETHEART BRACELET',
'90200B': 'BLACK SWEETHEART BRACELET',
'90200E': 'GREEN SWEETHEART BRACELET',
'90200C': 'BLUE SWEETHEART BRACELET',

'90185C': 'BLACK DIAMANTE EXPANDABLE RING',
'90185B': 'AMETHYST DIAMANTE EXPANDABLE RING',
'90003C': 'MIDNIGHT BLUE PAIR HEART HAIR SLIDE',
'90003D': 'CRYSTAL PAIR HEART HAIR SLIDES',
'90003B': 'ROSE COLOUR PAIR HEART HAIR SLIDES',
'90209C': 'PINK ENAMEL+GLASS HAIR COMB',
'90209A': 'PURPLE ENAMEL+GLASS HAIR COMB',
'90082D': 'DIAMANTE BOW BROOCH BLACK COLOUR',
'21977': 'PACK OF 60 PINK PAISLEY CAKE CASES',
'84991': '60 TEATIME FAIRY CAKE CASES',
'22273': 'FELTCRAFT DOLL MOLLY',
'22179': 'SET 10 LIGHTS NIGHT OWL',
'22320': 'BIRDS MOBILE VINTAGE DESIGN',
'22232': 'JIGSAW TOADSTOOLS 3 PIECE',
'84598': 'BOYS ALPHABET IRON ON PATCHES',
'79323S': 'SILVER CHERRY LIGHTS',
'84568': 'GIRLS ALPHABET IRON ON PATCHES ',
'21668': 'RED STRIPE CERAMIC DRAWER KNOB',
'21669': 'BLUE STRIPE CERAMIC DRAWER KNOB',
'21670': 'BLUE SPOT CERAMIC DRAWER KNOB',
'21672': 'WHITE SPOT RED CERAMIC DRAWER KNOB',
'21673': 'WHITE SPOT BLUE CERAMIC DRAWER KNOB',
'85071A': 'BLUE CHARLIE+LOLA PERSONAL DOORSIGN',
'85071B': 'RED CHARLIE+LOLA PERSONAL DOORSIGN',
'21161': 'KEEP OUT BOYS DOOR HANGER ',
'21162': 'TOXIC AREA DOOR HANGER ',
'21248': 'DOOR HANGER MUM + DADS ROOM',
'21159': 'MOODY BOY DOOR HANGER ',
'82494L': 'WOODEN FRAME ANTIQUE WHITE ',
'82486': 'WOOD S/3 CABINET ANT WHITE FINISH',
'82483': 'WOOD 2 DRAWER CABINET WHITE FINISH',
'21212': 'PACK OF 72 RETROSPOT CAKE CASES',
'72756': 'FAIRY CAKE CANDLES',
'22063': 'CERAMIC BOWL WITH STRAWBERRY DESIGN',
'37450': 'CERAMIC CAKE BOWL + HANGING CAKES',
'21707': 'FOLDING UMBRELLA BLACKBLUE POLKADOT',
'21710': 'FOLDING UMBRELLA PINKWHITE POLKADOT',
'21708': 'FOLDING UMBRELLA CREAM POLKADOT',
'21706': 'FOLDING UMBRELLA RED/WHITE POLKADOT',
'21711': 'FOLDING UMBRELLA WHITE/RED POLKADOT',
'20747': 'PICCADILLY TEA SET',
'20752': 'BLUE POLKADOT WASHING UP GLOVES',
'20753': 'RETRO PINK SPOT WASHING UP GLOVES',
'20751': 'FUNKY WASHING UP GLOVES ASSORTED',
'48185': 'DOORMAT FAIRY CAKE',
'21481': 'FAWN BLUE HOT WATER BOTTLE',
'85042': 'ANTIQUE LILY FAIRY LIGHTS',
'72760B': 'VINTAGE CREAM 3 BASKET CAKE STAND',
'22149': 'FELTCRAFT 6 FLOWER FRIENDS',
'21482': 'MUSHROOM BLUE HOT WATER BOTTLE',
'21484': 'CHICK GREY HOT WATER BOTTLE',
'20967': 'GREY FLORAL FELTCRAFT SHOULDER BAG',
'20970': 'PINK FLORAL FELTCRAFT SHOULDER BAG',
'84582': 'PINK CROCHET CAT WITH SCARF',
'20972': 'PINK CREAM FELT CRAFT TRINKET BOX ',
'84327A': 'PINK JUMPER LARRY THE LAMB',
'21490': "SET OF THREE 50'S GIFT WRAPS ",
'21063': 'PARTY INVITES JAZZ HEARTS',
'21058': 'PARTY INVITES WOODLAND',
'84997D': 'PINK 3 PIECE POLKADOT CUTLERY SET',
'84997C': 'BLUE 3 PIECE POLKADOT CUTLERY SET',
'22301': 'COFFEE MUG CAT + BIRD DESIGN',

'22300': 'COFFEE MUG DOG + BALL DESIGN',
'21231': 'SWEETHEART CERAMIC TRINKET BOX',
'20870': 'OPULENT VELVET CANDLE IN BOX',
'90121B': 'VINTAGE ROSE BEAD BRACELET BLUE',
'90175A': 'WHITE GLASS CHUNKY CHARM BRACELET',
'90195B': 'BLACK GEMSTONE BRACELET',
'90062': 'CARNIVAL BRACELET',
'21930': 'JUMBO STORAGE BAG SKULLS',
'21540': 'DAIRY MAID BUTTER DISH',
'21533': 'RETROSPOT LARGE MILK JUG',
'85083': 'KISS REINDEER SCANDINAVIAN STOCKING',
'21472': 'LADYBIRD + BEE RAFFIA FOOD COVER',
'47566B': 'TEA TIME PARTY BUNTING',
'22275': 'WEEKEND BAG VINTAGE ROSE PAISLEY',
'21479': 'WHITE SKULL HOT WATER BOTTLE ',
'21791': 'VINTAGE HEADS AND TAILS CARD GAME ',
'71477': 'COLOUR GLASS. STAR T-LIGHT HOLDER',
'37449': 'CERAMIC CAKE STAND + HANGING CAKES',
'37446': 'MINI CAKE STAND WITH HANGING CAKES',
'21258': 'VICTORIAN SEWING BOX LARGE',
'20892': 'SET/3 TALL GLASS CANDLE HOLDER PINK',
'47591D': 'PINK FAIRY CAKE CHILDRENS APRON',
'21867': 'PINK UNION JACK LUGGAGE TAG',
'82551': 'LAUNDRY 15C METAL SIGN',
'82567': 'AIRLINE LOUNGE,METAL SIGN',
'82613D': 'METAL SIGN CUPCAKE SINGLE HOOK',
'21179': 'NO JUNK MAIL METAL SIGN',
'21865': 'PINK UNION JACK PASSPORT COVER ',
'72783': "BLACK SIL'T SQU CANDLE PLATE ",
'20719': 'WOODLAND CHARLOTTE BAG',
'21869': 'POTTING SHED COFFEE MUG',
'21876': 'POTTERING MUG',
'84849B': 'FAIRY SOAP SOAP HOLDER',
'21864': 'UNION JACK FLAG PASSPORT COVER ',
'21470': 'FLOWER VINE RAFFIA FOOD COVER',
'20897': 'VINTAGE NOTEBOOK PARIS DAYS',
'44230': 'SQUARE MIRROR CURTAIN',
'37485': 'ENGLISH ROSE TEA SET IN GIFT BOX',
'84997A': 'GREEN 3 PIECE POLKADOT CUTLERY SET',
'21245': 'GREEN POLKADOT PLATE ',
'21241': 'GREEN POLKADOT CUP',
'21705': 'BAG 500g SWIRLY MARBLES',
'21913': 'VINTAGE SEASIDE JIGSAW PUZZLES',
'16207A': 'PINK STRAWBERRY HANDBAG ',
'21821': 'GLITTER STAR GARLAND WITH BELLS ',
'22319': 'HAIRCLIPS FORTIES FABRIC ASSORTED',
'21584': 'RETROSPOT SMALL TUBE MATCHES',
'21931': 'JUMBO STORAGE BAG SUKI',
'47591A': 'CAROUSEL CHILDRENS APRON',
'84997B': 'RED 3 PIECE RETROSPOT CUTLERY SET',
'22122': 'SET OF 2 FANCY FONT TEA TOWELS',
'21837': 'CROSS STITCH ALPHABET CUSHION COVER',
'84951A': 'SET OF 4 PISTACHIO LOVEBIRD COASTER',
'22335': 'HEART DECORATION PAINTED ZINC ',
'85129A': 'WHITE PEARL BEADED HEART, SMALL',
'84791': 'GINGHAM HEART DECORATION',
'21244': 'BLUE POLKADOT PLATE ',
'20675': 'BLUE POLKADOT BOWL',
'82616B': 'FRAPPUCINO SCARF KNITTING KIT',
'22276': 'WASH BAG VINTAGE ROSE PAISLEY',
'84912B': 'GREEN ROSE WASHBAG',
'85014C': 'BLUE/BROWN DOTS RUFFLED UMBRELLA',

'85014D': 'PINK/BROWN DOTS RUFFLED UMBRELLA',
'21326': 'AGED GLASS SILVER T-LIGHT HOLDER',
'84949': 'SILVER HANGING T-LIGHT HOLDER',
'21908': 'CHOCOLATE THIS WAY METAL SIGN',
'21371': 'MIRRORED WALL ART POPPIES',
'85152': 'HAND OVER THE CHOCOLATE SIGN ',
'21133': 'PAINTED BIRD ASSORTED CHRISTMAS',
'20984': '12 PENCILS TALL TUBE POSY',
'20982': '12 PENCILS TALL TUBE SKULLS',
'84971L': 'LARGE HEART FLOWERS HOOK ',
'90168': '2 DAISIES HAIR COMB',
'84884A': 'ANT WHITE WIRE HEART SPIRAL',
'90013C': 'GREEN VINTAGE EARRINGS ',
'90023': 'FILIGREE DIAMANTE EARRINGS',
'90067A': 'PINK VINTAGE VICTORIAN EARRINGS',
'90067B': 'BROWN VINTAGE VICTORIAN EARRINGS',
'90145': 'SILVER HOOP EARRINGS WITH FLOWER',
'90000A': 'RASPBERRY ANT COPPER FLOWER NECKLAC',
'90024B': "NEW BAROQUE B'FLY NECKLACE RED",
'90036D': 'FLOWER GLASS GARLAND NECKL.36"BLACK',
'90134': 'OLD ROSE COMBO BEAD NECKLACE',
'90131': 'PINK/AMETHYST/GOLD NECKLACE',
'90199C': '5 STRAND GLASS NECKLACE CRYSTAL',
'35751C': 'PURPLE CURRENT CANDLE RING',
'72739B': 'WHITE CHOCOLATE SCENT CANDLE',
'22212': 'FOUR HOOK WHITE LOVEBIRDS',
'48187': 'DOORMAT NEW ENGLAND',
'48195': 'DOOR MAT GREEN PAISLEY ',
'72741': 'GRAND CHOCOLATECANDLE',
'85232D': 'SET/3 DECOUPAGE STACKING TINS',
'20828': 'GLITTER BUTTERFLY CLIPS',
'22080': 'RIBBON REEL POLKADOTS ',
'22082': 'RIBBON REEL STRIPES DESIGN ',
'21784': 'SHOE SHINE BOX ',
'21172': 'PARTY METAL SIGN ',
'82600': 'NO SINGING METAL SIGN',
'22282': '12 EGG HOUSE PAINTED WOOD',
'21842': 'PINK HEART CANDY BUTTON CALCULATOR',
'22089': 'PAPER BUNTING VINTAGE PAISLEY',
'20734': 'SILVER MINI TAPE MEASURE ',
'35071': 'ASSORTED SANTA CHRISTMAS DECORATION',
'22121': 'NOEL WOODEN BLOCK LETTERS ',
'84836': 'ZINC METAL HEART DECORATION',
'37492A': 'YELLOW RED FLOWER PIGGY BANK',
'10133': 'COLOURING PENCILS BROWN TUBE',
'22113': 'GREY HEART HOT WATER BOTTLE',
'20782': 'CAMOUFLAGE EAR MUFF HEADPHONES',
'22190': 'LOCAL CAFE MUG',
'84029G': 'KNITTED UNION FLAG HOT WATER BOTTLE',
'21489': 'BLUE WHITE SCARF HOT WATER BOTTLE',
'20987': 'GEISHA GIRL CHOPSTICKS SET/5',
'20989': 'URBAN CHIC CHOPSTICKS SET/5',
'21868': 'POTTING SHED TEA MUG',
'21875': 'KINGS CHOICE MUG',
'21826': 'EIGHT PIECE DINOSAUR SET',
'21827': 'EIGHT PIECE CREEPY CRAWLIE SET',
'21828': 'EIGHT PIECE SNAKE SET',
'16225': 'RATTLE SNAKE EGGS',
'21829': 'DINOSAUR KEYRINGS ASSORTED',
'21638': 'ASSORTED TUTTI FRUTTI NOTEBOOK',
'22218': 'CAKE STAND LACE WHITE',
'22222': 'CAKE PLATE LOVEBIRD WHITE',

'22305': 'COFFEE MUG PINK PAISLEY DESIGN',
'22304': 'COFFEE MUG BLUE PAISLEY DESIGN',
'22302': 'COFFEE MUG PEARS DESIGN',
'22303': 'COFFEE MUG APPLES DESIGN',
'22127': 'PARTY CONES CARNIVAL ASSORTED',
'21143': 'ANTIQUE GLASS HEART DECORATION ',
'21639': 'ASSORTED TUTTI FRUTTI KEYRING',
'22154': 'ANGEL DECORATION 3 BUTTONS ',
'22155': 'STAR DECORATION RUSTIC',
'21889': 'WOODEN BOX OF DOMINOES',
'85045': 'GREEN CHRISTMAS TREE STRING 20LIGHT',
'85044': 'RED REINDEER STRING OF 20 LIGHTS',
'21587': 'COSY HOUR GIANT TUBE MATCHES',
'21588': 'RETROSPOT GIANT TUBE MATCHES',
'21818': 'GLITTER CHRISTMAS HEART ',
'21819': 'GLITTER CHRISTMAS STAR ',
'21817': 'GLITTER CHRISTMAS TREE',
'48111': 'DOORMAT 3 SMILEY CATS',
'20685': 'DOORMAT RED RETROSPOT',
'48194': 'DOORMAT HEARTS',
'21495': 'SKULLS AND CROSSBONES WRAP',
'47505': 'MINATURE COLOURED GARDENING SET',
'10135': 'COLOURING PENCILS BROWN TUBE',
'37492C': 'GREEN BLUE FLOWER PIGGY BANK',
'20677': 'PINK POLKADOT BOWL',
'20676': 'RED RETROSPOT BOWL',
'21086': 'SET/6 RED SPOTTY PAPER CUPS',
'21948': 'SET OF 6 CAKE CHOPSTICKS',
'21833': 'CAMOUFLAGE LED TORCH',
'21832': 'CHOCOLATE CALCULATOR',
'22315': '200 RED + WHITE BENDY STRAWS',
'21870': 'I CAN ONLY PLEASE ONE PERSON MUG',
'22115': 'METAL SIGN EMPIRE TEA',
'21166': 'COOK WITH WINE METAL SIGN ',
'21175': 'GIN + TONIC DIET METAL SIGN',
'21174': 'POTTERING IN THE SHED METAL SIGN',
'22197': 'SMALL POPCORN HOLDER',
'21094': 'SET/6 RED SPOTTY PAPER PLATES',
'72724': 'SET/6 CHRISTMAS ICICLE T-LIGHTS',
'22103': 'MIRROR MOSAIC T-LIGHT HOLDER ROUND',
'37424': 'WHITE WITH BLACK CATS MUG',
'21136': 'PAINTED METAL PEARS ASSORTED',
'21238': 'RED RETROSPOT CUP',
'21240': 'BLUE POLKADOT CUP',
'21239': 'PINK POLKADOT CUP',
'84007': 'MAGIC GARDEN MOUNT FUJI',
'84012': 'MAGIC SHEEP WOOL GROWING FROM PAPER',
'22077': '6 RIBBONS RUSTIC CHARM',
'85049A': 'TRADITIONAL CHRISTMAS RIBBONS',
'85049E': 'SCANDINAVIAN REDS RIBBONS',
'22075': '6 RIBBONS ELEGANT CHRISTMAS ',
'22076': '6 RIBBONS EMPIRE ',
'79323LP': 'LIGHT PINK CHERRY LIGHTS',
'79342B': 'BROWN FLOWER LIGHTS',
'79341': 'WILLOW BRANCH LIGHTS.',
'21976': 'PACK OF 60 MUSHROOM CAKE CASES',
'21498': 'RED RETROSPOT WRAP ',
'84948': 'SILVER HANGING T-LIGHT HOLDER DOME',
'21537': 'RED RETROSPOT PUDDING BOWL',
'20718': 'RED RETROSPOT SHOPPER BAG',
'21173': 'LOO ROLL METAL SIGN ',
'85124A': 'PINK JUICY FRUIT PHOTO FRAME',

'21539': 'RED RETROSPOT BUTTER DISH',
'16169F': 'WRAP BLIZZARD',
'21527': 'RED RETROSPOT TRADITIONAL TEAPOT ',
'21531': 'RED RETROSPOT SUGAR JAM BOWL',
'20914': 'SET/5 RED RETROSPOT LID GLASS BOWLS',
'21242': 'RED RETROSPOT PLATE ',
'21535': 'RED RETROSPOT SMALL MILK JUG',
'21844': 'RED RETROSPOT MUG',
'22073': 'RED RETROSPOT STORAGE JAR',
'22072': 'RED RETROSPOT TEA CUP AND SAUCER ',
'15056BL': 'EDWARDIAN PARASOL BLACK',
'15056N': 'EDWARDIAN PARASOL NATURAL',
'22254': 'FELT TOADSTOOL LARGE',
'22255': 'FELT TOADSTOOL SMALL',
'46000S': 'POLYESTER FILLER PAD 40x40cm',
'20964': 'POLYESTER FILLER PAD 60x40cm',
'21788': 'KIDS RAIN MAC BLUE',
'21789': 'KIDS RAIN MAC PINK',
'20682': 'RED RETROSPOT CHILDRENS UMBRELLA',
'84832': 'ZINC WILLIE WINKIE CANDLE STICK',
'85063': 'CREAM SWEETHEART MAGAZINE RACK',
'48129': 'DOORMAT TOPIARY',
'22030': 'SWALLOWS GREETING CARD',
'21081': 'SET/20 POSIES PAPER NAPKINS',
'20773': 'BLUE PAISLEY NOTEBOOK',
'20767': 'BLUE PAISLEY JOURNAL ',
'22028': 'PENNY FARTHING BIRTHDAY CARD',
'20622': 'VIPPASSPORT COVER ',
'21122': 'SET/10 PINK POLKADOT PARTY CANDLES',
'21124': 'SET/10 BLUE POLKADOT PARTY CANDLES',
'20657': 'TROPICAL LUGGAGE TAG',
'16156N': 'WRAP, FRIENDSHIP ON GREEN',
'20673': 'STRAWBERRIES PRINT BOWL',
'84520B': 'PACK 20 ENGLISH ROSE PAPER NAPKINS',
'20619': 'TROPICAL PASSPORT COVER ',
'21866': 'UNION JACK FLAG LUGGAGE TAG',
'20660': 'VIP LUGGAGE TAG ',
'20777': 'CHRYSANTHEMUM NOTEBOOK',
'20774': 'GREEN FERN NOTEBOOK',
'22037': 'ROBOT BIRTHDAY CARD',
'22036': 'DINOSAUR BIRTHDAY CARD',
'21726': 'MULTI HEARTS STICKERS',
'21544': 'SKULLS WATER TRANSFER TATTOOS ',
'21004': 'ROSE DU SUD CHILDRENS APRON',
'84402D': 'PINK DRESS JEWELLERY STAND ',
'84660B': 'BLACK STITCHED WALL CLOCK',
'21429': 'RED GINGHAM ROSE JEWELLERY BOX',
'72754B': 'F FAIRY POTPOURRI CUSHIONS LAVENDER',
'21500': 'PINK POLKADOT WRAP ',
'21431': 'GREEN GINGHAM FLOWER JEWELLERY BOX',
'20748': 'KENSINGTON COFFEE SET',
'72802B': 'OCEAN SCENT CANDLE IN JEWELLED BOX',
'72803A': 'ROSE SCENT CANDLE JEWELLED DRAWER',
'79029': 'GLASS HONEYPOT WASP CATCHER',
'21591': 'COSY HOUR CIGAR BOX MATCHES ',
'72801E': '4 FESTIVE GREEN DINNER CANDLES',
'21034': 'REX CASH+CARRY JUMBO SHOPPER',
'85049D': 'BRIGHT BLUES RIBBONS ',
'21985': 'PACK OF 12 HEARTS DESIGN TISSUES ',
'21980': 'PACK OF 12 RED RETROSPOT TISSUES ',
'16235': 'RECYCLED PENCIL WITH RABBIT ERASER',
'84347': 'ROTATING SILVER ANGELS T-LIGHT HLDR'

'84990': '60 GOLD AND SILVER FAIRY CAKE CASES',
'85231E': 'STRAWBERRY SCENTED SET/9 T-LIGHTS',
'85231B': 'CINAMMON SET OF 9 T-LIGHTS',
'72008': 'FROSTED BOX 9 WHITE T-LIGHT CANDLES',
'22071': 'SMALL WHITE RETROSPOT MUG IN BOX ',
'22070': 'SMALL RED RETROSPOT MUG IN BOX ',
'84270': 'ASSTD DESIGN BUBBLE GUM RING',
'85049C': 'ROMANTIC PINKS RIBBONS ',
'84352': 'SILVER CHRISTMAS TREE BAUBLE STAND ',
'79067': 'CORONA MEXICAN TRAY',
'22203': 'MILK PAN RED RETROSPOT',
'21353': 'SILVER SLEIGHBELL WREATH',
'21538': 'DAIRY MAID PUDDING BOWL',
'21534': 'DAIRY MAID LARGE MILK JUG',
'85053': 'FRENCH ENAMEL CANDLEHOLDER',
'22118': 'JOY WOODEN BLOCK LETTERS',
'35832': 'WOOLLY HAT SOCK GLOVE ADVENT STRING',
'D': 'Discount',
'21375': 'LARGE CAMPHOR WOOD FIELD MUSHROOM ',
'22186': 'RED STAR CARD HOLDER',
'47504K': 'ENGLISH ROSE GARDEN SECATEURS',
'47504F': 'ENGLISH ROSE TORCH',
'21809': 'CHRISTMAS HANGING TREE WITH BELL',
'21186': 'WHITE DOVE HONEYCOMB PAPER GARLAND',
'21611': 'SET OF 12 LILY BOTANICAL T-LIGHTS',
'21480': 'PINK STRIPE HOT WATER BOTTLE',
'85170A': 'SET/6 IVORY BIRD T-LIGHT CANDLES',
'85033S': 'SET/6 SILVER REINDEER T-LIGHTS',
'35991': 'SET/6 BLACK SNOWFLAKE T-LIGHTS ',
'22185': 'SLATE TILE NATURAL HANGING',
'21786': 'POLKADOT RAIN HAT ',
'84388': 'WISE MAN STAR SHAPE EGG PAN',
'21918': 'SET 12 KIDS COLOUR CHALK STICKS',
'51012C': 'FEATHER DUSTER, COAL BLACK',
'84671B': 'CROCHET BEAR WITH BLUE STRIPES',
'21747': 'SMALL SKULL WINDMILL',
'22330': '72 ROUND PINK DOILIES',
'84375': 'SET OF 20 KIDS COOKIE CUTTERS',
'84952B': 'BLACK LOVE BIRD T-LIGHT HOLDER',
'20961': 'STRAWBERRY BATH SPONGE ',
'20960': 'WATERMELON BATH SPONGE',
'20966': 'SANDWICH BATH SPONGE',
'17164D': "ASS COL SMALL SAND FROG P'WEIGHT",
'21890': 'S/6 WOODEN SKITTLES IN COTTON BAG',
'85170D': 'SET/6 PINK BIRD T-LIGHT CANDLES',
'17091J': 'VANILLA INCENSE IN TIN',
'20975': '12 PENCILS SMALL TUBE RED RETROSPOT',
'21440': '12 MINI TOADSTOOL PEGS',
'22174': 'PHOTO CUBE',
'22150': '3 STRIPEY MICE FELTCRAFT',
'22147': 'FELTCRAFT BUTTERFLY HEARTS',
'21356': 'TOAST ITS - FAIRY FLOWER',
'21357': 'TOAST ITS - DINOSAUR',
'21677': 'HEARTS STICKERS',
'21680': 'WOODLAND STICKERS',
'21730': 'GLASS STAR FROSTED T-LIGHT HOLDER',
'70007': 'HI TEC ALPINE HAND WARMER',
'70006': 'LOVE HEART POCKET WARMER',
'21650': 'ASSORTED TUTTI FRUTTI BRACELET',
'21892': 'TRADITIONAL WOODEN CATCH CUP GAME ',
'85227': 'SET OF 6 3D KIT CARDS FOR KIDS',
'20655': 'QUEEN OF SKIES LUGGAGE TAG',

'20655': 'QUEEN OF SKIES LUGGAGE TAG',
'21990': 'MODERN FLORAL STATIONERY SET',
'21934': 'SKULL SHOULDER BAG',
'21578': 'WOODLAND DESIGN COTTON TOTE BAG',
'21579': 'LOLITA DESIGN COTTON TOTE BAG',
'21577': 'SAVE THE PLANET COTTON TOTE BAG',
'20713': 'JUMBO BAG OWLS',
'20712': 'JUMBO BAG WOODLAND ANIMALS',
'16047': 'POP ART PEN CASE & PENS',
'21561': 'DINOSAUR LUNCH BOX WITH CUTLERY',
'20814': 'SILVER FINCH DECORATION',
'22068': 'BLACK PIRATE TREASURE CHEST',
'21377': 'SMALL CAMPHOR WOOD FIELD MUSHROOM',
'21379': 'CAMPHOR WOOD PORTOBELLO MUSHROOM',
'20956': 'PORCELAIN T-LIGHT HOLDERS ASSORTED',
'21018': 'BIRD BOX CHRISTMAS TREE DECORATION',
'35995': 'SET/6 BLACK GLITTER REINDEER CANDLE',
'22355': 'CHARLOTTE BAG SUKI DESIGN',
'21581': 'SKULLS DESIGN COTTON TOTE BAG',
'20727': 'LUNCH BAG BLACK SKULL.',
'21071': 'VINTAGE BILLBOARD DRINK ME MUG',
'21068': 'VINTAGE BILLBOARD LOVE/HATE MUG',
'22189': 'CREAM HEART CARD HOLDER',
'48179': 'DOOR MAT SWEET HOME',
'48188': 'DOORMAT WELCOME PUPPIES',
'48184': 'DOORMAT ENGLISH ROSE ',
'21777': 'RECIPE BOX WITH METAL HEART',
'20969': 'RED FLORAL FELTCRAFT SHOULDER BAG',
'20968': 'GREEN FLORAL FELTCRAFT SHOULDER BAG',
'21737': 'SET 3 WICKER LOG BASKETS',
'84899F': 'YELLOW PURPLE DAISY FELT PURSE KIT',
'84898F': 'YELLOW FLOWERS FELT HANDBAG KIT',
'22207': 'FRYING PAN UNION FLAG',
'21016': 'LIGHT BIRD HOUSE TREE DECORATION',
'20829': 'GLITTER HANGING BUTTERFLY STRING',
'85231G': 'ORANGE SCENTED SET/9 T-LIGHTS',
'21407': 'BROWN CHECK CAT DOORSTOP ',
'72800F': '4 IVORY DINNER CANDLES GOLD FLOCK',
'21993': 'FLORAL FOLK STATIONERY SET',
'21486': 'PINK HEART DOTS HOT WATER BOTTLE',
'21813': 'GARLAND WITH STARS AND BELLS',
'21812': 'GARLAND WITH HEARTS AND BELLS',
'21488': 'RED WHITE SCARF HOT WATER BOTTLE',
'75013B': 'STRING OF 8 BUTTERFLIES,PINK',
'72800E': '4 IVORY DINNER CANDLES SILVER FLOCK',
'72801G': '4 BURGUNDY WINE DINNER CANDLES',
'72801F': 'S/4 FESTIVE RED DINNER CANDLES',
'22183': 'CAKE STAND VICTORIAN FILIGREE MED',
'21732': 'RED HANGING METAL STAR TLIGHT HLDR',
'21254': 'SET OF KITCHEN WALL STICKERS',
'85041': 'SET/4 PINK ORCHID CANDLES IN BOWL',
'85040B': 'SET/4 BLUE FLOWER CANDLES IN BOWL',
'37461': 'FUNKY MONKEY MUG',
'21700': 'BIG DOUGHNUT FRIDGE MAGNETS',
'20692': 'FLORAL ELEPHANT SOFT TOY',
'16237': 'SLEEPING CAT ERASERS',
'37410': 'BLACK AND WHITE PAISLEY FLOWER MUG',
'84251F': 'GREETING CARD, RITA',
'21080': 'SET/20 RED RETROSPOT PAPER NAPKINS ',
'21967': 'PACK OF 12 SKULL TISSUES',
'22110': 'BIRD HOUSE HOT WATER BOTTLE',
'84977': 'WIRE FLOWER T-LIGHT HOLDER',
'18010': 'ASSORTED FRAGRANCE BATH CONFETTI',

'18010': 'ASSORTED FRAGRANCE BATH CONFETTI',
'22152': 'PLACE SETTING WHITE STAR',
'72267': 'PREMIUM CHURCH CANDLE',
'22081': 'RIBBON REEL FLORA + FAUNA ',
'22079': 'RIBBON REEL HEARTS DESIGN ',
'20758': 'ABSTRACT CIRCLES POCKET BOOK',
'20756': 'GREEN FERN POCKET BOOK',
'21126': 'SET OF 6 GIRLS CELEBRATION CANDLES',
'21125': 'SET 6 FOOTBALL CELEBRATION CANDLES',
'22102': 'MIRROR MOSAIC T-LIGHT HOLDER ',
'35955': 'SMALL FOLKART CHRISTMAS TREE DEC',
'35954': 'SMALL FOLKART STAR CHRISTMAS DEC',
'35957': 'SMALLFOLKART BAUBLE CHRISTMAS DEC',
'84050': 'PINK HEART SHAPE EGG FRYING PAN',
'47568': 'ENGLISH ROSE DESIGN PEG BAG',
'84782A': 'PINK 3 TIER GLASS PLATE STAND',
'84782C': 'GREEN 3 TIER GLASS PLATE STAND',
'22169': 'FAMILY ALBUM WHITE PICTURE FRAME',
'84933A': 'LARGE GREEN PEONY FOLDING STOOL',
'84933B': 'LARGE PINK PEONY FOLDING STOOL',
'22201': 'FRYING PAN BLUE POLKADOT',
'37503': 'TEA TIME CAKE STAND IN GIFT BOX',
'21593': 'SWALLOW CIGAR BOX MATCHES ',
'22085': 'PAPER CHAIN KIT SKULLS ',
'85099C': 'JUMBO BAG BAROQUE BLACK WHITE',
'22348': 'TEA BAG PLATE RED RETROSPOT',
'85061W': 'WHITE JEWELLED HEART DECORATION',
'21781': 'MA CAMPAGNE CUTLERY BOX',
'20768': 'GREEN FERN JOURNAL ',
'21082': 'SET/20 FRUIT SALAD PAPER NAPKINS ',
'47591B': 'SCOTTIES CHILDRENS APRON',
'22196': 'SMALL HEART MEASURING SPOONS',
'84968C': 'SET OF 16 VINTAGE PISTACHIO CUTLERY',
'84968F': 'SET OF 16 VINTAGE SKY BLUE CUTLERY',
'84968B': 'SET OF 16 VINTAGE IVORY CUTLERY',
'85231L': 'LAVENDER SCENTED SET/9 T-LIGHTS',
'21618': '4 WILDFLOWER BOTANICAL CANDLES',
'21615': '4 LAVENDER BOTANICAL DINNER CANDLES',
'21619': '4 VANILLA BOTANICAL CANDLES',
'21612': 'SET 12 WILDFLOWER BOTANICAL T-LIGHT',
'85049B': 'LUSH GREENS RIBBONS',
'85049G': 'CHOCOLATE BOX RIBBONS ',
'21770': 'OPEN CLOSED METAL SIGN',
'37500': 'TEA TIME TEAPOT IN GIFT BOX',
'21637': 'ASSORTED SANSKRIT MINI NOTEBOOK',
'21901': 'KEY FOB , BACK DOOR ',
'21900': 'KEY FOB , SHED',
'21902': 'KEY FOB , FRONT DOOR ',
'21223': 'SET/4 BADGES WOODLAND ANIMAL',
'85049H': 'URBAN BLACK RIBBONS ',
'21576': 'LETS GO SHOPPING COTTON TOTE BAG',
'21580': 'RABBIT DESIGN COTTON TOTE BAG',
'35004S': 'SET OF 3 SILVER FLYING DUCKS',
'35001W': 'HAND OPEN SHAPE DECO.WHITE',
'35001B': 'HAND OPEN SHAPE BLACK',
'21260': 'FIRST AID TIN',
'20866': 'BLUE ROSE FABRIC MIRROR',
'20867': 'PINK ROSE FABRIC MIRROR',
'20979': '36 PENCILS TUBE RED RETROSPOT',
'20983': '12 PENCILS TALL TUBE RED RETROSPOT',
'20981': '12 PENCILS TALL TUBE WOODLAND',
'20977': '36 PENCILS TUBE WOODLAND',

'20974': '12 PENCILS SMALL TUBE SKULL',
'21114': 'LAVENDER SCENTED FABRIC HEART',
'85232A': 'SET/3 POLKADOT STACKING TINS',
'35971': 'ROSE FOLKART HEART DECORATIONS',
'21740': 'COSY SLIPPER SHOES LARGE RED',
'21624': 'VINTAGE UNION JACK DOORSTOP',
'21738': 'COSY SLIPPER SHOES SMALL RED ',
'22280': 'POCKET BAG PINK PAISELY BROWN SPOT',
'21841': 'BABY MOUSE RED GINGHAM DRESS',
'21621': 'VINTAGE UNION JACK BUNTING',
'21622': 'VINTAGE UNION JACK CUSHION COVER',
'20750': 'RED RETROSPOT MINI CASES',
'82482': 'WOODEN PICTURE FRAME WHITE FINISH',
'20665': 'RED RETROSPOT PURSE ',
'84837': 'WHITE SWEETHEARTS CABINET 4 DRAWERS',
'21439': 'BASKET OF TOADSTOOLS',
'16169D': 'WRAP WINTER FOREST',
'16169C': 'WRAP BLUE REINDEER',
'21017': 'RED BIRD HOUSE TREE DECORATION',
'21210': 'SET OF 72 RETROSPOT PAPER DOILIES',
'21039': 'RED RETROSPOT SHOPPING BAG',
'21432': 'SET OF 3 CASES WOODLAND DESIGN',
'21424': 'WOODLAND STORAGE BOX LARGE ',
'21426': 'WOODLAND STORAGE BOX SMALL',
'21891': 'TRADITIONAL WOODEN SKIPPING ROPE',
'21098': 'CHRISTMAS TOILET ROLL',
'22415': 'WHITE TISSUE REAM',
'17165D': "ASS COL LARGE SAND FROG P'WEIGHT",
'17165B': "ASS COL LARGE SAND GECKO P'WEIGHT",
'85208': 'SET/12 FUNKY FELT FLOWER PEG IN BAG',
'90090': 'PINK CRYSTAL HEART PHONE CHARM',
'90092': 'BLUE CRYSTAL BOOT PHONE CHARM',
'16161C': 'WRAP SPEED DEMONS ',
'35160': 'BUDDHA INCENSE HOLDER ASSORTED',
'17108D': 'FLOWER FAIRY SUMMER BOUQUET SACHET',
'21805': 'WHITE CHRISTMAS TREE DECORATION ',
'90125A': 'PINK BERTIE GLASS BEAD BAG CHARM',
'84578': 'ELEPHANT TOY WITH BLUE T-SHIRT',
'84901': 'BLACK FACE LAMB WITH PANTS',
'35931': 'PINK STAR CHRISTMAS DECORATION',
'35933': 'PINK STOCKING CHRISTMAS DECORATION',
'90087': 'CRYSTAL SEA HORSE PHONE CHARM',
'90085': 'CRYSTAL STILETTO PHONE CHARM',
'90093': 'CLEAR CRYSTAL STAR PHONE CHARM',
'90084': 'PINK CRYSTAL GUITAR PHONE CHARM',
'90088': 'CRYSTAL KEY+LOCK PHONE CHARM',
'21807': 'WHITE CHRISTMAS STAR DECORATION',
'21806': 'WHITE CHRISTMAS HEART DECORATION',
'21815': 'STAR T-LIGHT HOLDER ',
'21816': 'CHRISTMAS TREE T-LIGHT HOLDER',
'35051B': 'PINK BEADED CONE CHRISTMAS',
'22153': 'ANGEL DECORATION STARS ON DRESS',
'20943': 'SET/6 GINGERBREAD STAR T-LIGHTS',
'72816': 'SET/3 CHRISTMAS DECOUPAGE CANDLES',
'85038': '6 CHOCOLATE LOVE HEART T-LIGHTS',
'20978': '36 PENCILS TUBE SKULLS',
'20942': 'SET/6 GINGERBREAD TREE T-LIGHTS',
'84559A': '3D SHEET OF DOG STICKERS',
'84466': 'TOP SECRET PEN SET',
'20994': 'JAZZ HEARTS SPIRAL NOTEPAD',
'20992': 'JAZZ HEARTS PURSE NOTEBOOK',
'84670': 'JAPANESE CROCHETED ANIMAL',

'84077': 'WORLD WAR 2 GLIDERS ASSTD DESIGNS',
'84029C': 'BROWN FURRY HOT WATER BOTTLE',
'22336': 'DOVE DECORATION PAINTED ZINC ',
'22156': 'HEART DECORATION WITH PEARLS ',
'22338': 'STAR DECORATION PAINTED ZINC ',
'21825': 'PAINTED METAL TREE WITH HOLLY BELLS',
'21822': 'GLITTER CHRISTMAS TREE WITH BELLS',
'22277': 'COSMETIC BAG VINTAGE ROSE PAISLEY',
'35979': 'SIX STRING SCANDINAVIAN HEART DECS',
'47566': 'PARTY BUNTING',
'21344': 'MOROCCAN BEATEN METAL DISH',
'21477': 'STRING OF 20 ROSE LIGHTS',
'85047': 'WHITE BEADED GARLAND STRING 20LIGHT',
'20681': 'PINK POLKADOT CHILDRENS UMBRELLA',
'84648': 'NEW BAROQUE WALL MIRROR',
'84472': 'PINK AND LILAC QUILTED THROW ',
'84910A': 'PINK PAISLEY CUSHION COVER ',
'84656': 'WHITE ROSE C/COVER',
'84580': 'MOUSE TOY WITH PINK T-SHIRT',
'84622': 'PINK GINGHAM ROSE CUSHION COVER',
'22151': 'PLACE SETTING WHITE HEART',
'35962': 'SET/12 ZINC CHRISTMAS DECORATIONS',
'22337': 'ANGEL DECORATION PAINTED ZINC ',
'90125C': 'TURQUOISE BERTIE GLASS BEAD CHARM',
'90125D': 'PURPLE BERTIE GLASS BEAD BAG CHARM',
'90214H': 'LETTER "H" BLING KEY RING',
'20901': 'VINTAGE KEEPSAKE BOX PINK FLOWER',
'22172': 'METAL SHELF WITH RAIL ',
'85067': 'CREAM SWEETHEART WALL CABINET',
'21156': 'RETROSPOT CHILDRENS APRON',
'72127': 'COLUMBIAN CANDLE ROUND ',
'84992': '72 SWEETHEART FAIRY CAKE CASES',
'84929': 'ASSTD FRUIT+FLOWERS FRIDGE MAGNETS',
'21649': 'ASSORTED TUTTI FRUTTI PHOTO FRAME',
'21463': 'MIRRORED DISCO BALL ',
'85162A': 'SILVER RETRODISC LAMPSHADE',
'21761': 'WOOD AND GLASS MEDICINE CABINET',
'84839': 'SWEETHEART KEY CABINET',
'85064': 'CREAM SWEETHEART LETTER RACK',
'84706C': 'PINK FLOCK TABLE CLOCK ',
'21804': 'WHITE CHRISTMAS GARLAND STARS TREES',
'21218': 'RED SPOTTY BISCUIT TIN',
'85103': 'SILVER T-LIGHT SETTING',
'20894': 'HANGING BAUBLE T-LIGHT HOLDER LARGE',
'85036C': 'ROSE 1 WICK MORRIS BOXED CANDLE',
'22162': 'HEART GARLAND RUSTIC PADDED',
'22297': 'HEART IVORY TRELLIS SMALL',
'85104': 'SILVER GLASS T-LIGHT SET',
'84813': 'SET OF 4 DIAMOND NAPKIN RINGS',
'72799D': 'BLUE PILLAR CANDLE SILVER FLOCK',
'21741': 'COSY SLIPPER SHOES LARGE GREEN',
'85071D': 'CHARLIE+LOLA MY ROOM DOOR SIGN',
'21724': 'PANDA AND BUNNIES STICKER SHEET',
'21524': 'DOORMAT SPOTTY HOME SWEET HOME',
'48197': 'DOORMAT BIRD ON THE WIRE',
'84684': 'BEACH HUT PHOTO FRAME',
'20812': 'EMBOSSED FLOWER TRELLIS PHOTO FRAME',
'22035': 'VINTAGE CARAVAN GREETING CARD ',
'21508': 'VINTAGE KID DOLLY CARD ',
'22027': 'TEA PARTY BIRTHDAY CARD',
'85059': 'FRENCH ENAMEL WATER BASIN',
'21465': 'PINK FLOWER CROCHET FOOD COVER',

'21466': 'RED FLOWER CROCHET FOOD COVER',
'21467': 'CHERRY CROCHET FOOD COVER',
'21471': 'STRAWBERRY RAFFIA FOOD COVER',
'21314': 'SMALL GLASS HEART TRINKET POT',
'84975': 'HEART SHAPED MIRROR',
'21949': 'SET OF 6 STRAWBERRY CHOPSTICKS',
'21947': 'SET OF 6 HEART CHOPSTICKS',
'21528': 'DAIRY MAID TRADITIONAL TEAPOT ',
'21548': 'CERAMIC BIRDHOUSE FINCH BLUE ROOF',
'21684': 'SMALL MEDINA STAMPED METAL BOWL ',
'21549': 'CERAMIC BIRDHOUSE RED ROOF SMALL',
'21545': 'CERAMIC BIRDHOUSE BUTTERFLY SMALL',
'21547': 'CERAMIC BIRDHOUSE CRESTED TIT SMALL',
'85054': 'FRENCH ENAMEL POT W LID',
'21532': 'DAIRY MAID SUGAR JAM BOWL',
'21530': 'DAIRY MAID TOASTRACK',
'37444A': 'YELLOW BREAKFAST CUP AND SAUCER',
'37444B': 'BLUE BREAKFAST CUP AND SAUCER ',
'22042': 'CHRISTMAS CARD SINGING ANGEL',
'84247C': 'PACK/12 BLUE REINDEER CARD',
'22034': 'ROBIN CHRISTMAS CARD',
'84754': 'S/15 SILVER GLASS BAUBLES IN BAG',
'21313': 'GLASS HEART T-LIGHT HOLDER ',
'84509A': 'SET OF 4 ENGLISH ROSE PLACEMATS',
'84510A': 'SET OF 4 ENGLISH ROSE COASTERS',
'84340': 'LARGE FIBRE OPTIC CHRISTMAS TREE',
'21625': 'VINTAGE UNION JACK APRON',
'18096C': 'WHITE ROUND PORCELAIN TLIGHT HOLDER',
'85147': 'JARDIN ETCHED GLASS CANDLEHLDR LONG',
'37444C': 'PINK BREAKFAST CUP AND SAUCER ',
'85055': 'FRENCH ENAMEL UTENSIL HOLDER',
'85028L': 'FRENCH CHATEAU LARGE FRUIT BOWL ',
'85028S': 'FRENCH CHATEAU SMALL FRUITBOWL',
'21214': 'SET OF 72 IVORY PAPER DOILIES',
'84985B': 'SET OF 72 BLACK PAPER DOILIES ',
'84985A': 'SET OF 72 GREEN PAPER DOILIES',
'47021G': 'SET/6 BEAD COASTERS GAUZE BAG GOLD',
'21911': 'GARDEN METAL SIGN ',
'21070': 'VINTAGE BILLBOARD MUG ',
'21066': 'VINTAGE RED MUG',
'21067': 'VINTAGE RED TEATIME MUG',
'21069': 'VINTAGE BILLBOARD TEA MUG',
'21986': 'PACK OF 12 PINK POLKADOT TISSUES',
'22325': 'MOBILE VINTAGE HEARTS ',
'20940': 'SET/6 GARDEN GNOME T-LIGHT CANDLES',
'72798C': 'SET/4 GARDEN ROSE DINNER CANDLE',
'40018F': 'CHERRY DESIGN PAPERLANTERNS',
'40018E': 'TEA TIME PAPER LANTERNS',
'85176': 'SEWING SUSAN 21 NEEDLE SET',
'21893': 'POTTING SHED SEED TIN',
'21745': 'GAOLERS KEYS DECORATIVE GARDEN ',
'17084J': 'LOVE POTION MASALA INCENSE',
'71279': 'PINK GLASS CANDLEHOLDER',
'21417': 'COCKLE SHELL DISH',
'22088': 'PAPER BUNTING COLOURED LACE',
'22093': 'MOTORING TISSUE BOX',
'84288B': 'SET/2 BLUE NOODLE BASKETS',
'35924': 'HANGING FAIRY CAKE DECORATION',
'20941': 'FROG CANDLE',
'21553': 'CERAMIC BIRDHOUSE FINCH BLUE LARGE',
'21554': 'CERAMIC BIRDHOUSE RED ROOF LARGE',
'22215': 'CAKE STAND WHITE TWO TIER LACE',

'20804': 'GLASS CHICKEN BUTTER DISH',
'21015': 'DARK BIRD HOUSE TREE DECORATION',
'21014': 'SWISS CHALET TREE DECORATION ',
'21917': 'SET 12 KIDS WHITE CHALK STICKS',
'22069': 'BROWN PIRATE TREASURE CHEST ',
'47591C': 'BLUE FAIRY CAKE CHILDRENS APRON',
'21743': 'STAR PORTABLE TABLE LIGHT ',
'21744': 'SNOWFLAKE PORTABLE TABLE LIGHT ',
'37501': 'TEA TIME TEA SET IN GIFT BOX',
'21385': 'IVORY HANGING DECORATION HEART',
'85099F': 'JUMBO BAG STRAWBERRY',
'35001G': 'HAND OPEN SHAPE GOLD',
'84855': 'BLOKEY SILVER TOOL SET',
'84854': 'GIRLY PINK TOOL SET',
'21216': 'SET 3 RETROSPOT TEA,COFFEE,SUGAR',
'21243': 'PINK POLKADOT PLATE ',
'21155': 'RED RETROSPOT PEG BAG',
'21412': 'VINTAGE GOLD TINSEL REEL',
'72799F': 'IVORY PILLAR CANDLE GOLD FLOCK',
'72799E': 'IVORY PILLAR CANDLE SILVER FLOCK',
'72799C': 'PINK PILLAR CANDLE SILVER FLOCK',
'21499': 'BLUE POLKADOT WRAP',
'21246': 'RED RETROSPOT BIG BOWL',
'21247': 'SET OF RED SALAD SERVERS ',
'85136C': 'RED SHARK HELICOPTER',
'21975': 'PACK OF 60 DINOSAUR CAKE CASES',
'21213': 'PACK OF 72 SKULL CAKE CASES',
'72807B': 'SET/3 OCEAN SCENT CANDLE JEWEL BOX',
'72802C': 'VANILLA SCENT CANDLE JEWELLED BOX',
'72799B': 'PURPLE PILLAR CANDLE BLACK FLOCK',
'84970L': 'SINGLE HEART ZINC T-LIGHT HOLDER',
'21768': 'WALL MOUNTED VINTAGE ORGANISER',
'84984D': 'ORIENTAL RED CUSHION COVER ',
'84773': 'RED ROSE AND LACE C/COVER',
'84601A': 'NEW BAROQUE IVORY CUSHION COVER ',
'85174': 'S/4 CACTI CANDLES',
'71028B': 'SILVER BLUE EGG T-LIGHT HOLDER',
'84755': 'COLOUR GLASS T-LIGHT HOLDER HANGING',
'72140F': 'HAPPY ANNIVERSARY CANDLE LETTERS',
'84950': 'ASSORTED COLOUR T-LIGHT HOLDER',
'85175': 'CACTI T-LIGHT CANDLES',
'72487': 'ROSES WHITE ROUND CANDLE',
'84360': 'TABLE LAMP WHITE SHADE WOOD BASE',
'21335': 'LIBRARY LAMP',
'84396': '6 HOOK JEWELLERY STAND BLACK DRESS',
'72753A': 'S/3 FLWR FAIRY LILLY CANDLES',
'72750C': 'F.FAIRY S/3 CANDLE/GLASS, ROSE',
'79323GR': 'GREEN CHERRY LIGHTS',
'22120': 'WELCOME WOODEN BLOCK LETTERS',
'85211': 'S/4 GROOVY CAT MAGNETS',
'84341B': 'SMALL PINK MAGIC CHRISTMAS TREE',
'22211': 'WOOD STAMP SET FLOWERS',
'22283': '6 EGG HOUSE PAINTED WOOD',
'22187': 'GREEN CHRISTMAS TREE CARD HOLDER',
'21735': 'TWO DOOR CURIO CABINET',
'48189': 'DOORMAT FRIENDSHIP ',
'48116': 'DOORMAT MULTICOLOUR STRIPE',
'79323B': 'BLACK CHERRY LIGHTS',
'21257': 'VICTORIAN SEWING BOX MEDIUM',
'21767': 'FRENCH STYLE WIRE DOOR CABINET ',
'85068': 'CREAM SWEETHEART SHELF + HOOKS',
'22171': '3 HOOK PHOTO SHELF ANTIQUE WHITE',

'72349B': 'SET/6 PURPLE BUTTERFLY T-LIGHTS',
'84596K': 'JELLY BABIES YELLOW SMALL BOWL',
'84596E': 'SMALL LICORICE DES PINK BOWL',
'84596G': 'SMALL CHOCOLATES PINK BOWL',
'37448': 'CERAMIC CAKE DESIGN SPOTTED MUG',
'22131': 'FOOD CONTAINER SET 3 LOVE HEART ',
'37447': 'CERAMIC CAKE DESIGN SPOTTED PLATE',
'22055': 'MINI CAKE STAND HANGING STRAWBERY',
'22062': 'CERAMIC BOWL WITH LOVE HEART DESIGN',
'84596B': 'SMALL DOLLY MIX DESIGN ORANGE BOWL',
'22347': 'PIZZA SLICE DISH',
'84507C': 'BLUE CIRCLES DESIGN MONKEY DOLL',
'84508A': 'CAMOUFLAGE DESIGN TEDDY',
'79302M': 'ART LIGHTS,FUNK MONKEY',
'37502': 'TEA TIME MUG IN GIFT BOX',
'84534B': 'FAIRY CAKE NOTEBOOK A5 SIZE',
'84535B': 'FAIRY CAKES NOTEBOOK A6 SIZE',
'84536B': 'FAIRY CAKES NOTEBOOK A7 SIZE',
'21752': 'GLASS CAKE STAND MIRRORED BASE',
'85090': 'CANDY SPOT APRON',
'21121': 'SET/10 RED POLKADOT PARTY CANDLES',
'90167': 'BEADED LOVE HEART JEWELLERY SET',
'21543': 'BUTTERFLY WATER TRANSFER TATTOOS ',
'21642': 'ASSORTED TUTTI FRUTTI PEN',
'72023F': 'RED ORGANIC CANDLE IN GLASS',
'20886': 'BOX OF 9 PEBBLE CANDLES',
'85172': 'HYACINTH BULB T-LIGHT CANDLES',
'21469': 'POLKA DOT RAFFIA FOOD COVER',
'21897': 'POTTING SHED CANDLE CITRONELLA',
'21898': 'POTTING SHED ROSE CANDLE',
'20771': 'CHRYSANTHEMUM JOURNAL',
'72815': '3 WICK CHRISTMAS BRIAR CANDLE ',
'47504G': 'ENGLISH ROSE TAPE MEASURE',
'84986B': 'SET OF 36 SILVER PAPER DOILIES',
'20684': 'STRAWBERRY DREAM CHILDS UMBRELLA',
'85006': 'SET 4 NURSERY DES ROUND BOXES',
'85199S': 'SMALL HANGING IVORY/RED WOOD BIRD',
'85199L': 'LARGE HANGING IVORY & RED WOOD BIRD',
'21556': 'CERAMIC STRAWBERRY MONEY BOX',
'82081': 'PINK HEARTS PHOTO ALBUM',
'21164': 'HOME SWEET HOME METAL SIGN ',
'21843': 'RED RETROSPOT CAKE STAND',
'47504E': 'ENGLISH ROSE SCISSORS',
'84934': 'ASSTD DESIGN POP ART SLIM PEN',
'21165': 'BEWARE OF THE CAT METAL SIGN ',
'21877': 'HOME SWEET HOME MUG',
'21398': 'RED POLKADOT COFFEE MUG',
'16161U': 'WRAP SUKI AND FRIENDS',
'21497': 'FANCY FONTS BIRTHDAY WRAP',
'22051': 'BLUE SCANDINAVIAN PAISLEY WRAP',
'22117': 'METAL SIGN HER DINNER IS SERVED ',
'22116': 'METAL SIGN HIS DINNER IS SERVED',
'21042': 'RED RETROSPOT APRON ',
'22164': 'STRING OF STARS CARD HOLDER',
'22165': 'DIAMANTE HEART SHAPED WALL MIRROR, ',
'46000P': 'POLYESTER FILLER PAD 65CMx65CM',
'46000M': 'POLYESTER FILLER PAD 45x45cm',
'21685': 'MEDINA STAMPED METAL TRAY ',
'84658': 'WHITE STITCHED CUSHION COVER',
'20887': 'BOX OF 6 PEBBLE CANDLES',
'21692': 'SILVER LATTICE VANILLA CANDLE POT',
'21691': 'SILVER VANILLA FLOWER CANDLE POT'

'21814': 'HEART T-LIGHT HOLDER ',
'47570B': 'TEA TIME TABLE CLOTH',
'21541': 'DINOSAURS WATER TRANSFER TATTOOS ',
'51014C': 'FEATHER PEN,COAL BLACK',
'20659': 'ECONOMY LUGGAGE TAG',
'20658': 'RED RETROSPOT LUGGAGE TAG',
'21355': 'TOAST ITS - I LOVE YOU ',
'84378': 'SET OF 3 HEART COOKIE CUTTERS',
'84380': 'SET OF 3 BUTTERFLY COOKIE CUTTERS',
'22144': 'CHRISTMAS CRAFT LITTLE FRIENDS',
'21330': 'WOODLAND ANIMAL WRITING SET ',
'72814': 'SET OF 3 CHRISTMAS BRIAR CANDLES',
'85125': 'SMALL ROUND CUT GLASS CANDLESTICK',
'85126': 'LARGE ROUND CUTGLASS CANDLESTICK',
'72529W': 'WHITE ADVENT CANDLE',
'21795': 'SET OF 3 BLUE STRIPE BASKET',
'21796': 'SET OF 3 RED STRIPE BASKETS',
'15036': 'ASSORTED COLOURS SILK FAN',
'21742': 'LARGE ROUND WICKER PLATTER ',
'21221': 'SET/4 BADGES CUTE CREATURES',
'84559B': '3D SHEET OF CAT STICKERS',
'84559D': '3D SHEET OF SEA WORLD STICKERS',
'15056P': 'EDWARDIAN PARASOL PINK',
'84199': 'GLOW IN DARK DOLPHINS',
'17003': 'BROCADE RING PURSE ',
'15034': 'PAPER POCKET TRAVELING FAN ',
'16238': 'PARTY TIME PENCIL ERASERS',
'35085E': 'IVORY ROSE LAMPSHADE ',
'10120': 'DOGGY RUBBER',
'17164B': "ASS COL SMALL SAND GECKO P'WEIGHT",
...}

为了演示方便，做一个非常粗糙但是可能很不合理的假设。 我们假设这个数据提供的，基于同一个customerid的购买记录是有顺序 关系的。那么，我们就可以把itermid看做是一个类似于文本的 序列数据，可以尝试使用word2vec模型。 因此，需要首先将所有的购买记录根据customerid分组

In [4]:

```
train_list = []    # 建立空列表
groups = records.groupby('customerid')    # 分组
for group in groups:
    temp_df = group[1]      # 提取元素
    my_records = list(temp_df['itemid'])    # 构建列表
    train_list.append(my_records)          # 添加到空列表
train_list[0]                          # 打印列表
```

Out[4]:

['TEST001',
 'TEST001',
 'TEST001',
 'TEST001',
 'TEST002',
 'TEST001',
 'TEST001',
 'TEST001',
 'TEST001',
 'TEST001',
 'ADJUST',
 '20682',
 '20679',
 '15056N',

'15056BL',
'15056P',
'21524',
'22692',
'22660',
'22687',
'48173C',
'22691',
'48111',
'22690',
'21523',
'48138',
'22689',
'22365',
'48185',
'22688',
'48188',
'48187',
'22366',
'20685',
'21955',
'M',
'M',
'M',
'M',
'M',
'M',
'M',
'M',
'M',
'M',
'D',
'TEST001']

尝试建立一个100维的word2vec模型

In [5]:

```
DIM=100   # 设立维度
model = Word2Vec(train_list, size=DIM, min_count=1)   # 模型建立
```

首先获得所有不通的word（通过index2word），计算这些WORD的个数，获得他们的空间坐标Z

In [6]:

```
products=model.wv.index2word   # 模型调研
M=len(products)                # 计算长度
Z=model[products]              # 设置模型元素
Z[0:3]                  # 打印列表
```

Out[6]:

```
array([[ 0.64142764, -0.12151847,  0.283735  ,  0.8760767 ,  0.2111246 ,
        -0.21458533, -0.387277  , -0.9541201 ,  0.04498236,  0.85430354,
         0.38846955,  0.55404586,  0.50015104, -0.45160267, -0.43055737,
        -0.6586711 , -0.1223806 ,  0.6634704 , -0.07301805, -0.24491575,
        -0.23016156, -1.0702176 , -0.63186556, -0.52576923, -0.5732349 ,
         0.6393819 , -0.4793257 ,  0.35970706,  0.01755434, -0.27413034,
        -0.31692272, -0.74574864,  0.5125053 , -0.5739702 ,  0.033603  ,
         0.21602231, -0.21184759, -1.0250087 ,  0.9766222 , -0.28887644,
        -0.20086497,  0.320183  ,  0.520856  , -1.0304912 ,  0.11483935,
```

```
 -0.40349317, -0.33167094, -0.07318869, -0.7638836 ,  0.40819347,
  0.42095935, -0.9233155 ,  0.18878882,  0.56585807, -0.2154034 ,
 -0.46314165,  0.5949456 , -0.6764075 , -0.7534002 , -0.60770464,
 -0.61932063,  0.20419332,  0.9132327 , -0.35003224, -0.03646546,
 -0.55970645,  0.6423672 ,  0.48929498, -0.6376039 , -0.0136434 ,
  0.3684285 ,  1.0887609 , -0.4662261 , -0.1798205 ,  1.3000172 ,
  0.77955127, -0.25434124, -0.07277333, -0.34983635,  0.54522663,
 -0.31995448, -0.17762907, -0.5350758 ,  0.9119324 , -0.19576621,
 -0.5811731 , -0.08834836, -0.2901597 ,  0.09320474, -0.69854647,
 -0.05125995, -0.24651171, -0.65065235,  0.37061188, -0.02153244,
  0.20943074,  0.7770775 ,  0.19270326,  0.40915456,  0.5970408 ],
[ 0.15053566, -0.21257354,  0.65138066, -0.19117694, -0.01742698,
 -0.2885174 , -1.1115094 ,  0.5209125 , -0.16963798,  0.03611134,
  0.28297538, -0.0860139 ,  0.72693163,  1.2254909 ,  0.35257447,
 -0.52785987, -0.290453  ,  1.2326783 ,  0.15126427,  0.01306852,
 -0.55170685, -0.7845006 , -0.6762014 ,  0.23969337, -0.10283015,
  0.6000467 ,  0.676376  ,  0.41804084, -0.3659454 ,  0.5590178 ,
  0.46632886,  0.03923266, -0.32891995, -1.0407909 , -0.70257914,
 -0.37628928,  0.5297289 ,  0.2988036 ,  0.9193447 ,  0.01771698,
  0.86651057,  1.0971856 ,  0.61075693,  1.1373799 , -0.4016995 ,
 -0.3849545 , -0.5101094 , -1.1200743 ,  0.17623757,  0.4552217 ,
  0.15112261, -0.35682046, -0.558497  ,  0.79808533,  0.46880296,
  0.46232572,  0.06726908, -0.16148761, -0.269681  , -0.19070825,
  0.50251395,  0.12917382, -0.3425356 ,  0.28307027, -0.20157662,
  0.1809374 ,  0.07195481, -0.314426  ,  0.31865337, -0.33729374,
  1.3875242 , -0.2336829 ,  0.06267579, -0.00681723, -0.37446186,
 -0.19540255, -0.00291107,  0.20069468, -0.91074723,  0.5034267 ,
 -0.23375966,  0.4647358 ,  0.34217182,  1.1772932 ,  0.32226366,
 -0.01291609,  0.63492453, -0.47898763,  0.7357939 ,  0.7868993 ,
  0.6283331 ,  0.6843309 , -1.4759805 ,  0.6892896 ,  0.32231247,
  0.06706129,  0.09548617, -0.04222035, -0.43366963, -0.39379227],
[-0.08316184, -0.13172904, -0.2920211 , -0.28057033, -0.01828005,
 -0.33913293, -0.41193554,  0.5343503 ,  0.30171528, -0.3051102 ,
 -0.19253685,  1.4422504 ,  0.19178522, -0.09749951, -0.51092976,
 -0.20519789, -0.28437296,  1.0875324 ,  0.6902937 ,  0.4522468 ,
 -2.0382354 , -1.1438586 , -0.21006314,  0.15698464, -0.25686097,
  1.1026198 , -0.19689548, -0.03597373,  0.8922151 , -0.06989169,
 -0.6086415 , -0.8022141 , -0.17667441, -0.97670585, -0.7736524 ,
  1.0259031 ,  0.6652292 ,  0.04642349,  1.4914975 , -0.23495273,
 -0.36007658, -0.9068238 ,  0.7664474 , -0.17706579, -0.35502848,
 -0.2016522 , -0.0847242 , -0.4243803 , -0.50189096, -0.94151706,
 -0.6013569 , -0.41148043, -0.15668046,  0.2179042 ,  0.4969789 ,
  1.1140894 ,  0.47291762, -0.28876492, -0.6416374 , -0.85512775,
 -0.34108773, -0.12327233,  1.3413712 , -0.09035023,  0.24574454,
 -0.74027807,  0.21687742,  0.09576017,  0.20232292,  0.47601795,
  1.659293  , -0.04908349,  0.47489786, -0.4215974 ,  0.9258439 ,
  0.15568256,  0.9826967 ,  0.30875674, -0.9845636 ,  0.49440718,
  0.59192437,  0.30929118,  1.4432107 ,  1.535837  ,  0.32804307,
  0.7299716 , -0.23616993, -0.63426226, -0.8597054 , -0.05368056,
  0.689651  , -0.64465517,  0.33423626,  0.45136988,  0.9672736 ,
  0.18627456,  0.3037388 ,  0.18864463, -0.71529293,  0.2298125 ]],
dtype=float32)
```

每一个消费者可能购买多个商品。每个商品有一个空间坐标 然后，将多个空间坐标求均值，用于描述该消费者的空间位置

In [7]:

```
N=len(train_list)   # 计算训练集长度
X=np.zeros([N,DIM])  # 构建以0填充的2维矩阵
for i in range(N):
```

```
    my_records=train_list[i]  # 训练集元素索引
    my_state=model[my_records]  # 模型调用
    X[i]=np.mean(my_state,0)  # 均值计算
```

任意指定MyID号顾客，打印他的消费记录

In [8]:

```
MyID=20  # 指定顾客编号
my_records=train_list[MyID]  # 训练集中索引元素
for i in range(len(my_records)): print(product_dictionary[my_records[i]])  # 列表输出
```

```
POSTAGE
ROUND SNACK BOXES SET OF4 WOODLAND
RED RETROSPOT PICNIC BAG
SET/20 WOODLAND PAPER NAPKINS
SET/20 RED RETROSPOT PAPER NAPKINS
RETROSPOT LARGE MILK JUG
LUNCH BAG RED RETROSPOT
RED RETROSPOT CUP
GREEN POLKADOT CUP
JUMBO BAG RED RETROSPOT
WOODLAND CHARLOTTE BAG
RED RETROSPOT CHARLOTTE BAG
LUNCH BAG WOODLAND
WOODLAND DESIGN  COTTON TOTE BAG
MUSHROOM BLUE HOT WATER BOTTLE
RED RETROSPOT PURSE
RED SPOTTY TABLECLOTH
RED RETROSPOT TRADITIONAL TEAPOT
JUMBO BAG WOODLAND ANIMALS
LUNCH BAG CARS BLUE
RED RETROSPOT SHOPPER BAG
RETROSPOT CHILDRENS APRON
RED RETROSPOT CUP
RED POLKADOT PUDDING BOWL
RED POLKADOT BEAKER
RED POLKADOT COFFEE  MUG
RED RETROSPOT TRADITIONAL TEAPOT
RETROSPOT LARGE MILK JUG
RED RETROSPOT MUG
ROUND SNACK BOXES SET OF4 WOODLAND
ROUND SNACK BOXES SET OF 4 SKULLS
SKULLS PARTY BAG + STICKER SET
RED RETROSPOT SHOPPER BAG
CHARLOTTE BAG SUKI DESIGN
SUKI  SHOULDER BAG
ROUND SNACK BOXES SET OF 4 FRUITS
POSTAGE
```

计算该消费者的空间坐标位置与其他各个产品之间的距离位置，并寻找最近的产品

In [9]:

```
mystate=X[MyID]  # 定义基于顾客编号的索引
similarity=Z-np.dot(np.ones([M,1]),mystate.reshape([1,DIM]))  # 计算相似度
similarity=np.mean(similarity**2,1)  # 计算相似度均值
TopProd=sorted(zip(similarity,products))  # 压缩并排序
TopProd=TopProd[0][1]  # 定义索引
TopProd=product_dictionary[TopProd]  # 打包成字典
TopProd
```

Out[9]:

'RED RETROSPOT SHOPPING BAG'

NLP - SEM Keyword Prediction

In [10]:

```
import jieba      # 导入库
import pandas as pd
import collections
from gensim.models import Word2Vec
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
from operator import itemgetter
import numpy as np
import warnings
warnings.filterwarnings('ignore')
```

读入SEM数据。第一列是关键词文本，第二列是对数变化后的平均展现量

In [11]:

```
SEM=pd.read_csv("F:/大三（上）/深度学习/TASK6.1：Wod-2-Vector - SEM关键词展现量预测/SEM.csv")
SEM   # 导入数据
```

Out[11]:

	kw	logImp
0	1.5折机票	2.302585
1	10.1机票	1.609438
2	10.1机票查询	1.791759
3	10.1特价机票	1.568616
4	10月份特价机票	2.036882
...	...	...
4836	最低折扣机票	3.786373
4837	最低折扣机票查询	2.826656
4838	最好的机票网站	1.845827
4839	最好的机票预订网站	1.386294
4840	最优惠机票	1.945910

4841 rows × 2 columns

利用jieba分词：

（1）wdtrain是一个长度为样本量的list，每一个元素又是一个list，其中每一个元素是分词后的最基本的词。

（2）wdlist是一个长度不确定的list，它是把所有的样本的分词结果全部都串在一起了

In [12]:

In [12]:

```
wdtrain=[];wdlist=[]    # 建立空列表
for kw in SEM.kw:
    wd=jieba.lcut(kw)    # 分词
    wdtrain.append(wd)    # 添加入列表
    wdlist=wdlist+wd      # 合并所有产生的列表
```

```
Building prefix dict from the default dictionary ...
Loading model from cache C:\Users\hp\AppData\Local\Temp\jieba.cache
Loading model cost 0.823 seconds.
Prefix dict has been built successfully.
```

通过对wdlist的各种操作，找到每个【词根】出现的频率，并降序排列

In [13]:

```
tab=collections.Counter(wdlist)
df = pd.DataFrame.from_dict(tab,orient="index").reset_index()  # 转化成数据框
df = df.rename(columns={'index':'kw', 0:'count'})  # 重命名
df = df.sort_values(by='count',ascending=False)    # 排序
```

生成一个字典，将词根变为整数，从0开始到ncov结束

In [14]:

```
ncov=len(df.kw)    # 计算长度
topkw=list(df.kw)  # 转化成列表
mydict=dict.fromkeys(topkw)  #会造一个value是空的字典
for i in range(ncov):
    mydict[topkw[i]]=i  # 循环打印
mydict
```

Out[14]:

```
{'机票': 0,
 '飞机票': 1,
 '特价机票': 2,
 '查询': 3,
 '预订': 4,
 '便宜': 5,
 '-': 6,
 '打折': 7,
 '特价': 8,
 '到': 9,
 '深圳': 10,
 '北京': 11,
 '航班': 12,
 '广州': 13,
 '的': 14,
 '上海': 15,
 '网站': 16,
 '预定': 17,
 '网': 18,
 '订': 19,
 '飞机': 20,
 '网上': 21,
 '订机票': 22,
 '机票价格': 23,
 '三亚': 24,
 '价格': 25,
 '订票': 26,
```

'买': 27,
'十一': 28,
'五一': 29,
'折扣': 30,
'南京': 31,
'厦门': 32,
'订购': 33,
'重庆': 34,
'武汉': 35,
'天津': 36,
'海口': 37,
'长沙': 38,
'昆明': 39,
'去': 40,
'最': 41,
'哈尔滨': 42,
'多少': 43,
'定': 44,
'大连': 45,
'乌鲁木齐': 46,
'钱': 47,
'哪个': 48,
'杭州': 49,
'哪里': 50,
'旅游': 51,
'郑州': 52,
'长春': 53,
'官网': 54,
'南宁': 55,
'济南': 56,
'电话': 57,
'呼和浩特': 58,
'机票网': 59,
'成都': 60,
'西安': 61,
'南昌': 62,
'票价': 63,
'时刻表': 64,
'珠海': 65,
'烟台': 66,
'春节': 67,
'低价': 68,
'太原': 69,
'端午节': 70,
'北京机票': 71,
'自助游': 72,
'什么': 73,
'西双版纳': 74,
'沈阳': 75,
'廉价': 76,
'往返机票': 77,
'哪儿': 78,
'青岛': 79,
'攻略': 80,
'好': 81,
'怎么': 82,
'成都机票': 83,
'放假': 84,
'携程网': 85,
'优惠机票': 86,
'2013': 87,
'西安机票': 88,

'如何': 89,
'黄山': 90,
'吗': 91,
'元旦': 92,
'往返': 93,
'怎样': 94,
'哪': 95,
'携程': 96,
'头等舱': 97,
'贵阳': 98,
'优惠': 99,
'丽江': 100,
'九寨沟': 101,
'张家界': 102,
'期间': 103,
'购买': 104,
'机票价': 105,
'酷讯': 106,
'最低': 107,
'提前': 108,
'航空': 109,
'电子机票': 110,
'温州': 111,
'机场': 112,
'在': 113,
'特惠': 114,
'年': 115,
'南航': 116,
'团购': 117,
'国内': 118,
'桂林': 119,
'时候': 120,
'艺龙': 121,
'折': 122,
'哪网': 123,
'途牛': 124,
'同程': 125,
'热线': 126,
'同程网': 127,
'航空公司': 128,
'5.1': 129,
'石家庄': 130,
'那个': 131,
'定票': 132,
'途牛网': 133,
'多久': 134,
'1': 135,
'国内机票': 136,
'北海': 137,
'网订': 138,
'一折': 139,
'合肥': 140,
'官方网站': 141,
'最好': 142,
'信用卡': 143,
'安排': 144,
'牡丹江': 145,
'南方': 146,
'银川': 147,
'宁波': 148,
'徐州': 149,
'兰州': 150,

'月份': 151,
'拉萨': 152,
'九江': 153,
'绵阳': 154,
'查': 155,
'义乌': 156,
'泸州': 157,
'旅游网': 158,
'儿童': 159,
'南阳': 160,
'福州': 161,
'常州': 162,
'全国': 163,
'西宁': 164,
'包头': 165,
'哪家': 166,
'临沂': 167,
'艺龙网': 168,
'酷讯网': 169,
'推荐': 170,
'庆阳': 171,
'阜阳': 172,
'5': 173,
'宜昌': 174,
'民航': 175,
'海南': 176,
'榆林': 177,
'南苑': 178,
'票': 179,
'衢州': 180,
'天气预报': 181,
'海拉尔': 182,
'旅游景点': 183,
'旅行网': 184,
'柳州': 185,
'儿童票': 186,
'可靠': 187,
'威海': 188,
'无锡': 189,
'广州白云机场': 190,
'通辽': 191,
'敦煌': 192,
'安徽': 193,
'可以': 194,
'天气': 195,
'延安': 196,
'有': 197,
'乌兰浩特': 198,
'喀什': 199,
'库尔勒': 200,
'大理': 201,
'湛江': 202,
'查寻': 203,
'武夷山': 204,
'中国': 205,
'适合': 206,
'丹东': 207,
'门票': 208,
'出发': 209,
'汕头': 210,
'团购网': 211,
'延吉': 212,

'景德镇': 213,
'揭阳': 214,
'云南': 215,
'会': 216,
'晋江': 217,
'眞的': 218,
'南通': 219,
'阿克苏': 220,
'佳木斯': 221,
'宜宾': 222,
'赤峰': 223,
'盐城': 224,
'梅州': 225,
'小孩': 226,
'广元': 227,
'洛阳': 228,
'南充': 229,
'促销': 230,
'潍坊': 231,
'最低价': 232,
'降价': 233,
'查找': 234,
'安全': 235,
'海航': 236,
'安庆': 237,
'抢': 238,
'路线': 239,
'哪买': 240,
'和田': 241,
'购票': 242,
'梧州': 243,
'格尔木': 244,
'购': 245,
'乌海': 246,
'好去处': 247,
'比较': 248,
'湖北': 249,
'一周': 250,
'黑河': 251,
'过年': 252,
'贵州': 253,
'吉林': 254,
'驴友': 255,
'齐齐哈尔': 256,
'机票信息': 257,
'林芝': 258,
'贵': 259,
'襄樊': 260,
'黄花': 261,
'常德': 262,
'中心': 263,
'系统': 264,
'今年': 265,
'连云港': 266,
'芒市': 267,
'悠哉': 268,
'网定': 269,
'网买': 270,
'达州': 271,
'穷游网': 272,
'保山': 273,
'锡林浩特': 274,

'秦皇岛': 275,
'6': 276,
'锦州': 277,
'黄金周': 278,
'正规': 279,
'坐飞机': 280,
'恩施': 281,
'涨价': 282,
'2': 283,
'思茅': 284,
'五一节': 285,
'井冈山': 286,
'10.1': 287,
'东营': 288,
'穷游': 289,
'那儿': 290,
'游': 291,
'一日': 292,
'抢购': 293,
'大全': 294,
'要': 295,
'自由': 296,
'塔城': 297,
'行': 298,
'地方': 299,
'青海': 300,
'四川': 301,
'越': 302,
'网是': 303,
'飞机场': 304,
'人少': 305,
'旅行': 306,
'秒杀': 307,
'假期': 308,
'最佳': 309,
'是': 310,
'国庆节': 311,
'前': 312,
'小': 313,
'长假': 314,
'时间': 315,
'攀枝花': 316,
'几折': 317,
'澳门': 318,
'哪定': 319,
'伊宁': 320,
'婴儿': 321,
'几天': 322,
'特价网': 323,
'那里': 324,
'加格达奇': 325,
'算': 326,
'白云机场': 327,
'标准': 328,
'达县': 329,
'打折票': 330,
'半': 331,
'价钱': 332,
'州': 333,
'到达': 334,
'东方航空': 335,
'全攻略': 336,

'朝阳': 337,
'哪些': 338,
'流程': 339,
'安康': 340,
'福建': 341,
'西昌': 342,
'甘肃': 343,
'怎么样': 344,
'中秋': 345,
'虹桥机场': 346,
'黄岩': 347,
'河南': 348,
'河北': 349,
'浙江': 350,
'新疆': 351,
'天天': 352,
'迪庆': 353,
'咸阳': 354,
'月': 355,
'网购': 356,
'池州': 357,
'永州': 358,
'广州白云国际机场': 359,
'赣州': 360,
'阿勒泰': 361,
'剩余': 362,
'短途': 363,
'淘': 364,
'万县': 365,
'淮安': 366,
'艺龙定': 367,
'大同': 368,
'深航': 369,
'周边': 370,
'介绍': 371,
'兴义': 372,
'国际': 373,
'折价': 374,
'宜春': 375,
'国航': 376,
'一元': 377,
'五月': 378,
'学生': 379,
'专题': 380,
'格': 381,
'起': 382,
'多长时间': 383,
'搜': 384,
'自助': 385,
'云南旅游': 386,
'十一月份': 387,
'同城网': 388,
'网络': 389,
'平台': 390,
'去处': 391,
'网易': 392,
'飞': 393,
'旅游胜地': 394,
'干什么': 395,
'报价': 396,
'花费': 397,

'打九折': 398,
'活动': 399,
'出行': 400,
'前后': 401,
'淘宝': 402,
'海边': 403,
'西塘': 404,
'国庆': 405,
'贤贤网': 406,
'玩': 407,
'查询电话': 408,
'1.5': 409,
'规定': 410,
'山西': 411,
'淡季': 412,
'春运': 413,
'改签': 414,
'退': 415,
'认证': 416,
'代': 417,
'网址': 418,
'东北': 419,
'东莞': 420,
'二折': 421,
'预订电话': 422,
'超低价': 423,
'余票': 424,
'关于': 425,
'广交会': 426,
'广西': 427,
'嘉峪关': 428,
'扬州': 429,
'遵义': 430,
'邯郸': 431,
'济宁': 432,
'超': 433,
'超低': 434,
'国外': 435,
'jipiao': 436,
'10': 437,
'17u': 438,
'.': 439,
'cn': 440,
'元': 441,
'51': 442,
'e': 443,
'龙': 444,
'feijipiao': 445,
'tejiajipiao': 446,
'已': 447,
'遨游': 448,
'白菜价': 449,
'阿尔山': 450,
'佛山': 451,
'鸡西': 452,
'长治': 453,
'两日': 454,
'北京首都机场': 455,
'伊春': 456,
'在线': 457,
'黑龙江': 458,
'厦门航空': 459,

'宁夏': 460,
'预约': 461,
'历史': 462,
'两折': 463,
'辽宁': 464,
'临沧': 465,
'售票': 466,
'哪有': 467,
'那拉提': 468,
'南方航空公司': 469,
'平价': 470,
'七天': 471,
'青春': 472,
'机': 473,
'那网': 474,
'能': 475,
'三峡': 476,
'凤凰': 477,
'三折': 478,
'厦航': 479,
'厦门航空公司': 480,
'库车': 481,
'京东': 482,
'湖南': 483,
'卡': 484,
'风景区': 485,
'注意事项': 486,
'亲子': 487,
'惠州': 488,
'比价': 489,
'采购': 490,
'东航': 491,
'呢': 492,
'及': 493,
'定购': 494,
'江西': 495,
'附加费': 496,
'号': 497,
'货到付款': 498,
'酒店': 499,
'票号': 500,
'搜索': 501,
'可信': 502,
早订': 503,
'江苏': 504,
'舟山': 505}

生成一个X的0-1矩阵。如果一个关键词包含相应词根，那么对应的X元素取值为1，否则为0

In [15]:

```
ss=len(SEM.kw)  # 计算列表长度
ncov=len(mydict)  # 计算列表长度
X=np.zeros([ss,ncov+1])  # 构建以0填充的数组
for i in range(ss):
    columnID=itemgetter(*wdtrain[i])(mydict)
    X[i,columnID]=1            # 赋值
    X[i,ncov]=len(SEM.kw[i])   # 计算长度
```

In [16]:

```
X # 打印出数组
```

Out[16]:

```
array([[1., 0., 0., ..., 0., 0., 6.],
       [1., 0., 0., ..., 0., 0., 6.],
       [1., 0., 0., ..., 0., 0., 8.],
       ...,
       [1., 0., 0., ..., 0., 0., 7.],
       [1., 0., 0., ..., 0., 0., 9.],
       [0., 0., 0., ..., 0., 0., 5.]])
```

做一个最普通的线性回归，看看外样本的预测精度。外样本的预测精度其实是非常不稳定的，充满了异常值。因此，没有用均值，而是采用的中位数。

In [17]:

```
Y=np.array(SEM.logImp)      # 构建数组
Y=(Y-np.mean(Y))/np.std(Y)  # 标准化数据
niter=1000                  # 参数设置
R21=np.zeros(niter)         # 构建1000个以0填充的1维数组
regmodel = LinearRegression()   # 设置线性回归模型
for i in range(niter):
    X0,X1,Y0,Y1= train_test_split(X,Y,test_size=0.5,random_state=i) # 拆分数据集
    regmodel.fit(X0,Y0)    # 模型调用
    Yhat=regmodel.predict(X1)  # 模型预测
    R21[i]=100.0*(1-np.median((Yhat-Y1)**2))  # 计算中位数
np.round(np.median(R21),2)  # 输出中位数
```

Out[17]:

```
72.77
```

接下来做一个word2vec模型，假设laten空间的维数为100

In [18]:

```
DIM=100
wvmodel = Word2Vec(wdtrain, size=DIM, min_count=1)  # 函数设置
Z=np.zeros([ss,DIM])   # 构建以0填充的2维数组
for i in range(ss):
    vec=wvmodel[wdtrain[i]]  # 训练
    Z[i,:]=np.mean(vec,0)  # 计算均值
```

把word2vect学习出来的空间坐标的平均值作为一个新的解释性变量Z与前面的X整合，形成新的矩阵XZ，然后做Y对XZ的回归分析，并检测外样本预测精度

In [19]:

```
XZ=np.hstack((X,Z))  # 将矩阵进行行连接
niter=1000      # 参数设置
R22=np.zeros(niter)  # 构建以0填充的1维数组
regmodel = LinearRegression()   # 构建1000个以0填充的1维数组
for i in range(niter):
    XZ0,XZ1,Y0,Y1= train_test_split(XZ,Y,test_size=0.5,random_state=i) # 训练集拆分
    regmodel.fit(XZ0,Y0) # 模型调用
    Yhat=regmodel.predict(XZ1)  # 模型预测
    R22[i]=100.0*(1-np.median((Yhat-Y1)**2))  # 计算中位数
np.round(np.median(R22),2)    # 输出中位数
```

Out[19]:

75.92

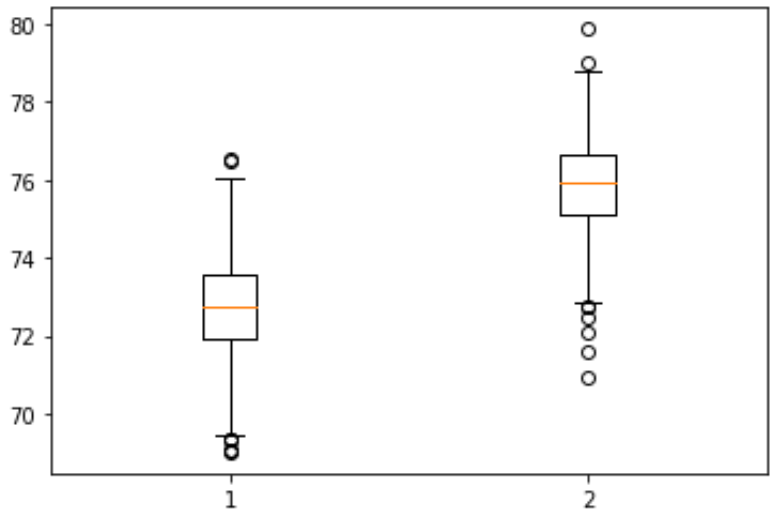
对两个不同的模型，多次随机模型的重复结果，做箱线图对比如下。结果令人振奋！

In [20]:

```
plt.boxplot((R21,R22)) # 绘制箱线图
```

Out[20]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x2c0a7d72208>,
  <matplotlib.lines.Line2D at 0x2c0a7d77688>,
  <matplotlib.lines.Line2D at 0x2c0a7d94bc8>,
  <matplotlib.lines.Line2D at 0x2c0a7d94348>],
 'caps': [<matplotlib.lines.Line2D at 0x2c0a07671c8>,
  <matplotlib.lines.Line2D at 0x2c0a7d77c48>,
  <matplotlib.lines.Line2D at 0x2c0a7d94408>,
  <matplotlib.lines.Line2D at 0x2c0a9062248>],
 'boxes': [<matplotlib.lines.Line2D at 0x2c0a7d8cb88>,
  <matplotlib.lines.Line2D at 0x2c0a7d99608>],
 'medians': [<matplotlib.lines.Line2D at 0x2c0a7d79388>,
  <matplotlib.lines.Line2D at 0x2c0a9062888>],
 'fliers': [<matplotlib.lines.Line2D at 0x2c0a7d79548>,
  <matplotlib.lines.Line2D at 0x2c0a7d8a6c8>],
 'means': []}
```

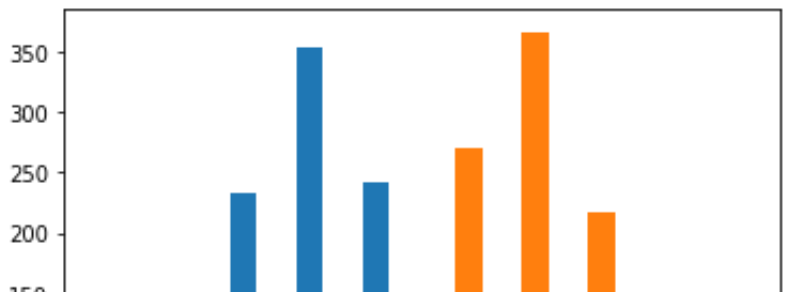


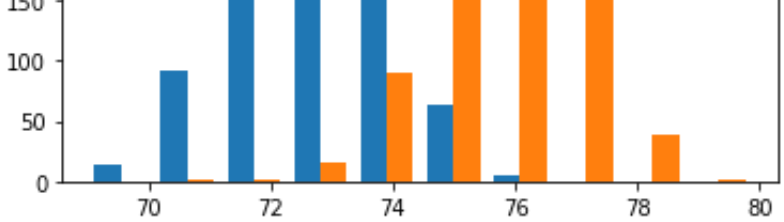
In [21]:

```
plt.hist((R21,R22)) # 绘制直方图
```

Out[21]:

```
([array([ 14.,  91., 233., 353., 241.,  63.,   5.,   0.,   0.,   0.]),
  array([  0.,   1.,   2.,  15.,  89., 270., 366., 217.,  38.,   2.])],
 array([69.01098546, 70.09760916, 71.18423287, 72.27085658, 73.35748028,
        74.44410399, 75.5307277 , 76.6173514 , 77.70397511, 78.79059881,
        79.87722252]),
 <a list of 2 Lists of Patches objects>)
```




150
100
50
0
70
72
74
76
78
80