**Application of Association rule to build Market Basket Analysis Case**

**Business Objective**: Steve is store manager of a Multi brand retail store. He has access to the transactions data stored every time a customer buys something from his stores. He knew by experience that certain products are often bought together but with huge variety of products available, he wanted to perform some statistical test to capture the insightful association available in the transaction records available in his database.

He had plans to re-design the store layout so that customers find the associated products next to each other. He wanted to rethink over the offers and promotions to ensure that they make sense to the customers. Finally, this basket analysis will help him proactively recommend “What else” and “What next” to his loyal customers.

Steve will perform Market basket analysis and based on the output results, plan his store layout, promotions and recommendations.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Market basket analysis also known as Affinity of products is based on the application of the Association rule concept of the data mining. The algorithm works on the conditional probability and generates the output based on the analysis of certain parameters.

Customers purchase and stuff their baskets with some subset of the products, and we analyze what products people buy together, even if we don't know who they are. We can sue this information to position products in future sales and enhance the shopping experience by controlling the way a typical customer browses the store.

Selection of affinity rules from the set of all possible rules needs some constraints on various measures of significance like Support, Confidence and Lift.

**Important definitions:**

**Itemset** or **transaction dataset**: The data grid comprising of the transactions to be analyzed. The transaction id variable should be unique and should have the list of items bought in the basket. The rules for the transaction are based on the Antecedents and Consequents in the sequence of the products bought.

**Rules** are statements represented in the following form

{i 1 ,i 2 ,...}⇒ {i k}

It is interpreted as the items in item set (on the left hand side of the rule i.e. {i 1 ,i 2 ,...}  ), then it can be safely assumed that the customer will be interested in the item on the right hand side (RHS i.e. {i k }  .

The output of a market basket analysis is generally a set of rules, that we can use to make business decisions (related to marketing or product placement, promotions etc).

**Support**: The support, Supp() of an itemset is defined as the proportion of transactions in the data set which contain the itemset.

Support Score Supp(A) = (Count of product A in N transaction)/ Total Transactions (N)

**Confidence**: The confidence of a rule conf(A,B) is given as a ratio of the support(AUB) andsupport(A). Here, support (AUB) means "support for occurrences of transactions where A and B both appear"

Confidence can be interpreted as an estimate of the probability P(B|A), the probability of finding the RHS of the rule in transactions under the condition that these transactions also contain the LHS.

**Lift** is defined as Lift (A,B) and is given as the ratio of the observed support to that expected if A and B were [independent](http://en.wikipedia.org/wiki/Independence_(probability_theory))

Lift (A,B) = Supp(AUB)/Supp(A) X Supp(B)

The Lift score of greater than 1 is desired and indicates possible correlation between two itemsets.

***Conviction*** is defined as Conv(A,B) and is given as ratio of the (1- supp(A,B)) to (1-conf(A,B))

Conv (A,B) = (1-Supp(AUB) ) / (1-conf(A,B)

It can be interpreted as the ratio of the expected frequency that A occurs without B i.e the frequency that the rule makes an incorrect prediction) if A and B were independent divided by the observed frequency of incorrect predictions. The conviction value of 1.x shows that the rule would be incorrect x% more often (1.x times as often) if the association between A and B was purely random chance.

The above concepts can be applied other way as negations rules as well i.e the low Lift(A,B) can be an indicator of association of (A with NOT B).

**Analysis steps and Code**

### Set the work directory, change as per your work folder

setwd("C:/Users/babycorn/Documents/market basket analysis")

### Import/read the data

txn\_data<-read.csv("Retail\_Data.csv")

### See the data summary (verify Data)

head(txn\_data)

transaction\_id Prod1 Prod2 Prod3

1 100001 D E I

2 100002 B G I

3 100003 B F I

4 100004 D F H

5 100005 C F I

6 100006 D G H

tail(txn\_data)

transaction\_id Prod1 Prod2 Prod3

11995 111995 B G H

11996 111996 D G H

11997 111997 C E H

11998 111998 B E I

11999 111999 A F H

12000 112000 B G H

summary(txn\_data)

transaction\_id Prod1 Prod2 Prod3

Min. :100001 A:2983 E:3962 H:5907

1st Qu.:103001 B:3024 F:4053 I:6093

Median :106001 C:3047 G:3985

Mean :106001 D:2946

3rd Qu.:109000

Max. :112000

### Install and run the following libraries

library(arules)

library(arulesSequences)

### Take a subset for analysis from the complete Dataset

smpl\_dat<-txn\_data[1:5000,]

### Factorization of variables to get all the combinations

for ( i in 1:ncol(smpl\_dat))

{

smpl\_dat[,i]=as.factor(smpl\_dat[,i])

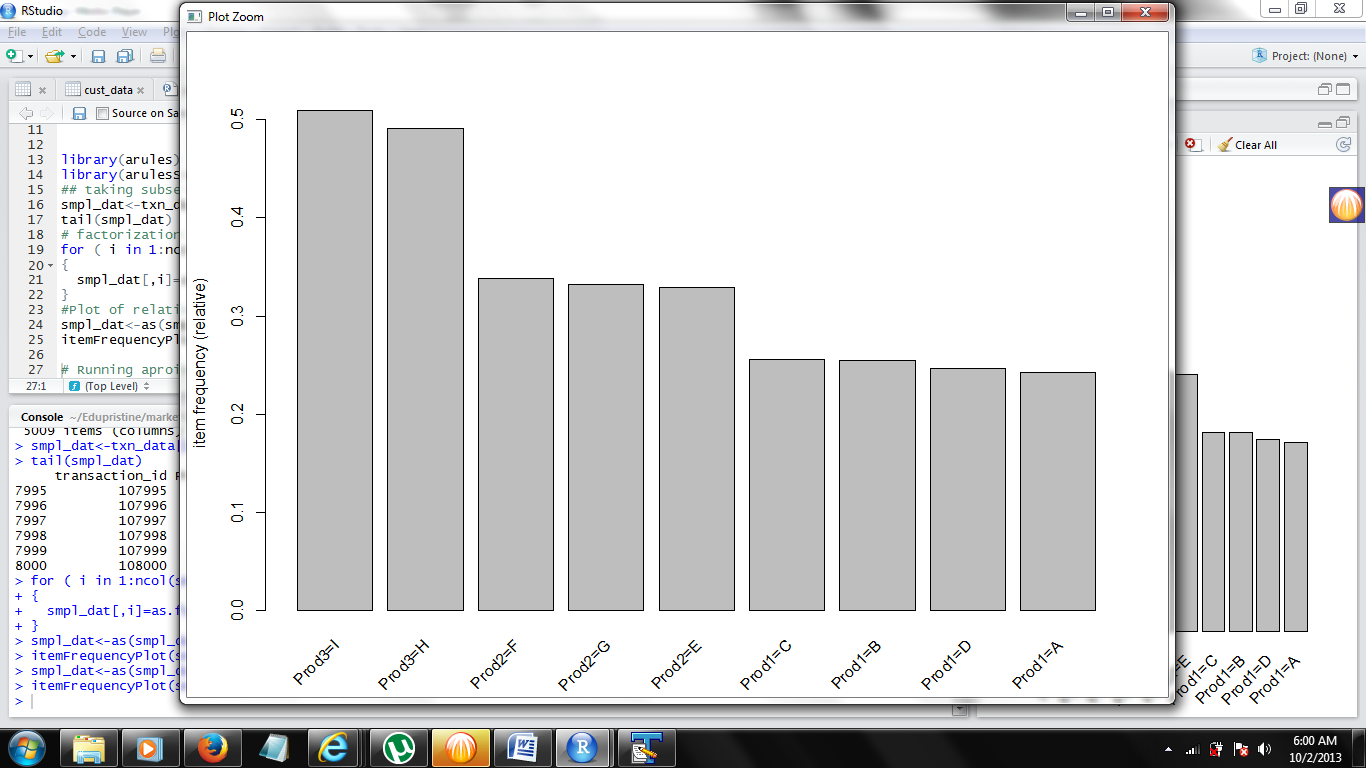
}

###Plot of relative frequency; update topN variable to the number of distinct products

smpl\_dat<-as(smpl\_dat,"transactions")

itemFrequencyPlot(smpl\_dat, topN = 9)

This gives a plot highlighting the relative frequency of all the products bought by the customers.



### Run the apriori algorithm command

### Observe the constraints set on the lift and confidence as covered at the top of the document. The minlen and maxlen sets the analysis level. Maxlen 3 implies the rule will analyze the association of 3rd product as combinations of 1st and 2nd product bought together.

basket\_rules <- apriori(smpl\_dat, parameter = list(sup = 0.005, conf = 0.01, target="rules", minlen=2,maxlen=3))

### See the summary of the outputs from apriori rule

summary( basket\_rules)

set of 124 rules

rule length distribution (lhs + rhs):sizes

2 3

52 72

Min. 1st Qu. Median Mean 3rd Qu. Max.

2.000 2.000 3.000 2.581 3.000 3.000

summary of quality measures:

support confidence lift

Min. :0.03562 Min. :0.2279 Min. :0.8956

1st Qu.:0.04113 1st Qu.:0.2572 1st Qu.:0.9757

Median :0.04437 Median :0.3298 Median :1.0003

Mean :0.07258 Mean :0.3548 Mean :1.0000

3rd Qu.:0.08700 3rd Qu.:0.4838 3rd Qu.:1.0247

Max. :0.17400 Max. :0.5455 Max. :1.1056

mining info:

data ntransactions support confidence

smpl\_dat 8000 0.005 0.01

### See all the rules generated by the apriori rule

inspect( basket\_rules) # see all the rules

lhs rhs support confidence lift

1 {Prod1=A} => {Prod2=E} 0.083375 0.3436373 1.0429053

2 {Prod2=E} => {Prod1=A} 0.083375 0.2530349 1.0429053

3 {Prod1=A} => {Prod2=G} 0.078875 0.3250902 0.9780825

4 {Prod2=G} => {Prod1=A} 0.078875 0.2373073 0.9780825

5 {Prod1=A} => {Prod2=F} 0.080375 0.3312725 0.9797339

6 {Prod2=F} => {Prod1=A} 0.080375 0.2377079 0.9797339

7 {Prod1=A} => {Prod3=H} 0.120125 0.4951056 1.0088754

8 {Prod3=H} => {Prod1=A} 0.120125 0.2447784 1.0088754

9 {Prod1=A} => {Prod3=I} 0.122500 0.5048944 0.9914470

10 {Prod3=I} => {Prod1=A} 0.122500 0.2405498 0.9914470

11 {Prod1=D} => {Prod2=E} 0.078375 0.3182741 0.9659305

12 {Prod2=E} => {Prod1=D} 0.078375 0.2378604 0.9659305

13 {Prod1=D} => {Prod2=G} 0.080875 0.3284264 0.9881200

14 {Prod2=G} => {Prod1=D} 0.080875 0.2433246 0.9881200

15 {Prod1=D} => {Prod2=F} 0.087000 0.3532995 1.0448784

16 {Prod2=F} => {Prod1=D} 0.087000 0.2573013 1.0448784

17 {Prod1=D} => {Prod3=H} 0.119625 0.4857868 0.9898865

18 {Prod3=H} => {Prod1=D} 0.119625 0.2437596 0.9898865

19 {Prod1=D} => {Prod3=I} 0.126625 0.5142132 1.0097461

20 {Prod3=I} => {Prod1=D} 0.126625 0.2486500 1.0097461

21 {Prod1=B} => {Prod2=E} 0.086500 0.3392157 1.0294861

22 {Prod2=E} => {Prod1=B} 0.086500 0.2625190 1.0294861

23 {Prod1=B} => {Prod2=G} 0.085875 0.3367647 1.0132071

24 {Prod2=G} => {Prod1=B} 0.085875 0.2583678 1.0132071

25 {Prod1=B} => {Prod2=F} 0.082625 0.3240196 0.9582835

26 {Prod2=F} => {Prod1=B} 0.082625 0.2443623 0.9582835

27 {Prod1=B} => {Prod3=H} 0.126625 0.4965686 1.0118566

28 {Prod3=H} => {Prod1=B} 0.126625 0.2580234 1.0118566

29 {Prod1=B} => {Prod3=I} 0.128375 0.5034314 0.9885741

30 {Prod3=I} => {Prod1=B} 0.128375 0.2520864 0.9885741

31 {Prod1=C} => {Prod2=E} 0.081250 0.3172279 0.9627554

32 {Prod2=E} => {Prod1=C} 0.081250 0.2465857 0.9627554

33 {Prod1=C} => {Prod2=G} 0.086750 0.3387018 1.0190351

34 {Prod2=G} => {Prod1=C} 0.086750 0.2610004 1.0190351

35 {Prod1=C} => {Prod2=F} 0.088125 0.3440703 1.0175831

36 {Prod2=F} => {Prod1=C} 0.088125 0.2606285 1.0175831

37 {Prod1=C} => {Prod3=H} 0.124375 0.4856027 0.9895114

38 {Prod3=H} => {Prod1=C} 0.124375 0.2534386 0.9895114

39 {Prod1=C} => {Prod3=I} 0.131750 0.5143973 1.0101075

40 {Prod3=I} => {Prod1=C} 0.131750 0.2587138 1.0101075

41 {Prod2=E} => {Prod3=H} 0.155500 0.4719272 0.9616448

42 {Prod3=H} => {Prod2=E} 0.155500 0.3168619 0.9616448

43 {Prod2=E} => {Prod3=I} 0.174000 0.5280728 1.0369619

44 {Prod3=I} => {Prod2=E} 0.174000 0.3416789 1.0369619

45 {Prod2=G} => {Prod3=H} 0.162875 0.4900338 0.9985407

46 {Prod3=H} => {Prod2=G} 0.162875 0.3318900 0.9985407

47 {Prod2=G} => {Prod3=I} 0.169500 0.5099662 1.0014063

48 {Prod3=I} => {Prod2=G} 0.169500 0.3328424 1.0014063

49 {Prod2=F} => {Prod3=H} 0.172375 0.5097967 1.0388114

50 {Prod3=H} => {Prod2=F} 0.172375 0.3512481 1.0388114

51 {Prod2=F} => {Prod3=I} 0.165750 0.4902033 0.9625986

52 {Prod3=I} => {Prod2=F} 0.165750 0.3254786 0.9625986

53 {Prod1=A,

Prod2=E} => {Prod3=H} 0.038750 0.4647676 0.9470558

54 {Prod1=A,

Prod3=H} => {Prod2=E} 0.038750 0.3225806 0.9790004

55 {Prod2=E,

Prod3=H} => {Prod1=A} 0.038750 0.2491961 1.0270835

56 {Prod1=A,

Prod2=E} => {Prod3=I} 0.044625 0.5352324 1.0510209

57 {Prod1=A,

Prod3=I} => {Prod2=E} 0.044625 0.3642857 1.1055712

58 {Prod2=E,

Prod3=I} => {Prod1=A} 0.044625 0.2564655 1.0570449

59 {Prod1=A,

Prod2=G} => {Prod3=H} 0.040250 0.5103011 1.0398392

60 {Prod1=A,

Prod3=H} => {Prod2=G} 0.040250 0.3350676 1.0081012

61 {Prod2=G,

Prod3=H} => {Prod1=A} 0.040250 0.2471220 1.0185349

62 {Prod1=A,

Prod2=G} => {Prod3=I} 0.038625 0.4896989 0.9616080

63 {Prod1=A,

Prod3=I} => {Prod2=G} 0.038625 0.3153061 0.9486457

64 {Prod2=G,

Prod3=I} => {Prod1=A} 0.038625 0.2278761 0.9392112

65 {Prod1=A,

Prod2=F} => {Prod3=H} 0.041125 0.5116641 1.0426166

66 {Prod1=A,

Prod3=H} => {Prod2=F} 0.041125 0.3423517 1.0125005

67 {Prod2=F,

Prod3=H} => {Prod1=A} 0.041125 0.2385787 0.9833227

68 {Prod1=A,

Prod2=F} => {Prod3=I} 0.039250 0.4883359 0.9589316

69 {Prod1=A,

Prod3=I} => {Prod2=F} 0.039250 0.3204082 0.9476027

70 {Prod2=F,

Prod3=I} => {Prod1=A} 0.039250 0.2368024 0.9760017

71 {Prod1=D,

Prod2=E} => {Prod3=H} 0.035625 0.4545455 0.9262261

72 {Prod1=D,

Prod3=H} => {Prod2=E} 0.035625 0.2978056 0.9038108

73 {Prod2=E,

Prod3=H} => {Prod1=D} 0.035625 0.2290997 0.9303540

74 {Prod1=D,

Prod2=E} => {Prod3=I} 0.042750 0.5454545 1.0710939

75 {Prod1=D,

Prod3=I} => {Prod2=E} 0.042750 0.3376111 1.0246163

76 {Prod2=E,

Prod3=I} => {Prod1=D} 0.042750 0.2456897 0.9977245

77 {Prod1=D,

Prod2=G} => {Prod3=H} 0.040750 0.5038640 1.0267223

78 {Prod1=D,

Prod3=H} => {Prod2=G} 0.040750 0.3406479 1.0248901

79 {Prod2=G,

Prod3=H} => {Prod1=D} 0.040750 0.2501919 1.0160076

80 {Prod1=D,

Prod2=G} => {Prod3=I} 0.040125 0.4961360 0.9742484

81 {Prod1=D,

Prod3=I} => {Prod2=G} 0.040125 0.3168806 0.9533826

82 {Prod2=G,

Prod3=I} => {Prod1=D} 0.040125 0.2367257 0.9613225

83 {Prod1=D,

Prod2=F} => {Prod3=H} 0.043250 0.4971264 1.0129932

84 {Prod1=D,

Prod3=H} => {Prod2=F} 0.043250 0.3615465 1.0692688

85 {Prod2=F,

Prod3=H} => {Prod1=D} 0.043250 0.2509065 1.0189095

86 {Prod1=D,

Prod2=F} => {Prod3=I} 0.043750 0.5028736 0.9874788

87 {Prod1=D,

Prod3=I} => {Prod2=F} 0.043750 0.3455084 1.0218363

88 {Prod2=F,

Prod3=I} => {Prod1=D} 0.043750 0.2639517 1.0718852

89 {Prod1=B,

Prod2=E} => {Prod3=H} 0.041375 0.4783237 0.9746790

90 {Prod1=B,

Prod3=H} => {Prod2=E} 0.041375 0.3267522 0.9916608

91 {Prod2=E,

Prod3=H} => {Prod1=B} 0.041375 0.2660772 1.0434399

92 {Prod1=B,

Prod2=E} => {Prod3=I} 0.045125 0.5216763 1.0244012

93 {Prod1=B,

Prod3=I} => {Prod2=E} 0.045125 0.3515093 1.0667959

94 {Prod2=E,

Prod3=I} => {Prod1=B} 0.045125 0.2593391 1.0170160

95 {Prod1=B,

Prod2=G} => {Prod3=H} 0.041500 0.4832606 0.9847388

96 {Prod1=B,

Prod3=H} => {Prod2=G} 0.041500 0.3277394 0.9860531

97 {Prod2=G,

Prod3=H} => {Prod1=B} 0.041500 0.2547966 0.9992024

98 {Prod1=B,

Prod2=G} => {Prod3=I} 0.044375 0.5167394 1.0147068

99 {Prod1=B,

Prod3=I} => {Prod2=G} 0.044375 0.3456670 1.0399909

100 {Prod2=G,

Prod3=I} => {Prod1=B} 0.044375 0.2617994 1.0266644

101 {Prod1=B,

Prod2=F} => {Prod3=H} 0.043750 0.5295008 1.0789623

102 {Prod1=B,

Prod3=H} => {Prod2=F} 0.043750 0.3455084 1.0218363

103 {Prod2=F,

Prod3=H} => {Prod1=B} 0.043750 0.2538071 0.9953220

104 {Prod1=B,

Prod2=F} => {Prod3=I} 0.038875 0.4704992 0.9239062

105 {Prod1=B,

Prod3=I} => {Prod2=F} 0.038875 0.3028238 0.8955971

106 {Prod2=F,

Prod3=I} => {Prod1=B} 0.038875 0.2345400 0.9197646

107 {Prod1=C,

Prod2=E} => {Prod3=H} 0.039750 0.4892308 0.9969043

108 {Prod1=C,

Prod3=H} => {Prod2=E} 0.039750 0.3195980 0.9699484

109 {Prod2=E,

Prod3=H} => {Prod1=C} 0.039750 0.2556270 0.9980557

110 {Prod1=C,

Prod2=E} => {Prod3=I} 0.041500 0.5107692 1.0029833

111 {Prod1=C,

Prod3=I} => {Prod2=E} 0.041500 0.3149905 0.9559651

112 {Prod2=E,

Prod3=I} => {Prod1=C} 0.041500 0.2385057 0.9312084

113 {Prod1=C,

Prod2=G} => {Prod3=H} 0.040375 0.4654179 0.9483808

114 {Prod1=C,

Prod3=H} => {Prod2=G} 0.040375 0.3246231 0.9766773

115 {Prod2=G,

Prod3=H} => {Prod1=C} 0.040375 0.2478895 0.9678457

116 {Prod1=C,

Prod2=G} => {Prod3=I} 0.046375 0.5345821 1.0497440

117 {Prod1=C,

Prod3=I} => {Prod2=G} 0.046375 0.3519924 1.0590219

118 {Prod2=G,

Prod3=I} => {Prod1=C} 0.046375 0.2735988 1.0682238

119 {Prod1=C,

Prod2=F} => {Prod3=H} 0.044250 0.5021277 1.0231842

120 {Prod1=C,

Prod3=H} => {Prod2=F} 0.044250 0.3557789 1.0522111

121 {Prod2=F,

Prod3=H} => {Prod1=C} 0.044250 0.2567078 1.0022753

122 {Prod1=C,

Prod2=F} => {Prod3=I} 0.043875 0.4978723 0.9776580

123 {Prod1=C,

Prod3=I} => {Prod2=F} 0.043875 0.3330171 0.9848934

124 {Prod2=F,

Prod3=I} => {Prod1=C} 0.043875 0.2647059 1.0335027

### Sort the rules by Lift and get top n ( 20 in the following case) rules generated by the apriori rule. The following 20 rules can be analyzed for association and affinity.

inspect(head(sort( basket\_rules,by="lift"),20))

lhs rhs support confidence lift

1 {Prod1=A,

Prod3=I} => {Prod2=E} 0.044625 0.3642857 1.105571

2 {Prod1=B,

Prod2=F} => {Prod3=H} 0.043750 0.5295008 1.078962

3 {Prod2=F,

Prod3=I} => {Prod1=D} 0.043750 0.2639517 1.071885

4 {Prod1=D,

Prod2=E} => {Prod3=I} 0.042750 0.5454545 1.071094

5 {Prod1=D,

Prod3=H} => {Prod2=F} 0.043250 0.3615465 1.069269

6 {Prod2=G,

Prod3=I} => {Prod1=C} 0.046375 0.2735988 1.068224

7 {Prod1=B,

Prod3=I} => {Prod2=E} 0.045125 0.3515093 1.066796

8 {Prod1=C,

Prod3=I} => {Prod2=G} 0.046375 0.3519924 1.059022

9 {Prod2=E,

Prod3=I} => {Prod1=A} 0.044625 0.2564655 1.057045

10 {Prod1=C,

Prod3=H} => {Prod2=F} 0.044250 0.3557789 1.052211

11 {Prod1=A,

Prod2=E} => {Prod3=I} 0.044625 0.5352324 1.051021

12 {Prod1=C,

Prod2=G} => {Prod3=I} 0.046375 0.5345821 1.049744

13 {Prod1=D} => {Prod2=F} 0.087000 0.3532995 1.044878

14 {Prod2=F} => {Prod1=D} 0.087000 0.2573013 1.044878

15 {Prod2=E,

Prod3=H} => {Prod1=B} 0.041375 0.2660772 1.043440

16 {Prod1=A} => {Prod2=E} 0.083375 0.3436373 1.042905

17 {Prod2=E} => {Prod1=A} 0.083375 0.2530349 1.042905

18 {Prod1=A,

Prod2=F} => {Prod3=H} 0.041125 0.5116641 1.042617

19 {Prod1=B,

Prod3=I} => {Prod2=G} 0.044375 0.3456670 1.039991

20 {Prod1=A,

Prod2=G} => {Prod3=H} 0.040250 0.5103011 1.039839

**Conclusion**:

Based on the above results we can say the whenever Product A and Product I will be bought together, there are high chances that product E will also be bought. Similarly we can interpret the other rows and plan the promotions or store layouts to catch customer’s attention.