

# Proposal of Item-sets in View of Simultaneous Purchase in Convenience Store Using Association Analysis

**Wei Zang**

Department of Industrial and Systems Engineering  
Aoyama Gakuin University, Kanagawa, Japan  
Tel: (+81) 42-759-6311, Email: [zang@ise.aoyama.ac.jp](mailto:zang@ise.aoyama.ac.jp)

**Mio Ishiwata**

MARUI GROUP CO., LTD. Japan  
Tel: (+81) 90-7175-2576, Email: [l\\_mimimitaro@yahoo.co.jp](mailto:l_mimimitaro@yahoo.co.jp)

**Satoshi Kumagai**

Department of Industrial and Systems Engineering  
Aoyama Gakuin University, Kanagawa, Japan  
Tel: (+81) 42-759-6312, Email: [kumagai@ise.aoyama.ac.jp](mailto:kumagai@ise.aoyama.ac.jp)

**Abstract.** In a limited area, it is important for convenience store to display and sell items that contribute to income. In this study we propose a new approach to design new item-sets and the prices of them in order to increase profit of convenience store, and validate it by using POS-data (the data includes three different type stores' 1 year purchasing data). Firstly, classify the items in view of simultaneous purchase with ABC analysis and association analysis into three time zones (breakfast time zone, lunch time zone and dinner time zone). Secondly, calculate the benefit of each item-set in view of simultaneous purchase and raking them for each time zone. Then, pick up the top 5 goods of each class and propose new set for each time zone. Then design a questionnaire survey to gather the expected price information of each item-set. Decision of item-set' price is based on the survey data. By comparison with POS data, to verify the contribution potential to profit increase due to item-sets.

**Keywords:** Association Analysis, item-set, ABC Analysis, POS-data, R

## 1. INTRODUCTION

In this study, we focus on the Heibai structure between multiple items for net profit improvement of at a convenience store. The purchase history of the customer from multiple Heibai relationship of items and analysis from the ID with POS data, to set the item-sets. From the setting of item-sets, the acquisition of new customers, increase the number of items purchased, it is an object of the net income improvement in the convenience store. In the association analysis, using support, confidence, lift, the three indicators to analyze the relevance of the items.

## 2. METHODS OF ANALYSIS

### 2.1 ASSOCIATION ANALYSIS

Associations analysis is a function representing the strength of association between each item. To take advantage

of the transaction data, which is a one-time information, such as that which item and which item in the shopping has been purchased in the retail cash register system, between the items in receipt relevance analysis in the early 1990s in order to do about, is a technique that has been developed by IBM Research. The three indicators are defined as the following equation 2.1~2.3

Support :

$$\text{support}(x \rightarrow y) = P(x \cap y) \quad (2.1)$$

$P(x)$  = x commodity purchasers number / total number of purchasers

$P(y)$  = y commodity purchasers number / total purchasers number,

$P(x \cap y)$  = commodity x, and purchasers' number / total purchasers' number of y.

It means the percentage of conclusion (y) occurs when the condition (x) has occurred. Condition (x) and the

transaction satisfying conclusion (y) at the same time is a percentage of all transactions, it is an index representing the appearance frequency of the rule.

In the whole of the population, if this rule is the extent to which occurs, the higher the percentage of the rule appears in the whole, whether small and rarely appeared not clogging the rule to the contrary is supported much means.

Confidence :

$$confidence(x \rightarrow y) = \frac{P(x \cap y)}{P(x)} \quad (2.2)$$

It shows that there is a strong relationship between the two instruments to this number is concluded that the conditions the greater means that likely to occur at the same time.

Confidence be thought as an index meaningful near correlation coefficient can be determined by conditional probabilities on the calculation.

Lift :

$$lift(x \rightarrow y) = \frac{P(x \cap y)}{P(x) \cdot P(y)} \\ = \frac{confidence(x \rightarrow y)}{support(y)} = \frac{confidence(y \rightarrow x)}{support(x)} \quad (2.3)$$

The lift value is an index indicating whether the probability that the conclusion of the association rule (y) is purchased is turned up by the purchase condition (x).

The lower the lift value, the conclusion (y) is generated by a single reason, conditions than association with items (x) is considered to have occurred at the conclusion (y) specific reasons. In other words, condition (x) and also as a confidence was high in the conclusion (y), its relevance is to the interpretation that much sense not. If the lift value is greater than 1, as a general guideline is valid rules.

$lift(x \rightarrow y) > 1$  if a strong relationship between the two items, when you purchase one of the items the other item there is a positive association that tends to be purchasing at the same time. On the contrary,  $lift(x \rightarrow y) < 1$  is weak 1 If relevance, The other items when purchasing one item represents that there is a negative association that hardly purchasing

## 2.2 ABC ANALYSIS

ABC analysis (emphasis analysis) is one of the typical method for determining the degree of importance of the commodity in the retail industry. The ABC analysis, to determine what's in handling all items are selling, is a method to divide in order of importance to the three ranks of A · B · C in each item. When arranged in sales of high order of goods,

it can be regarded as are important commodity earns a lot of sales in a few items of higher order. On the contrary, not as large a percentage of net sales also attracted a large number of items which is located in the lower. Therefore, any items rather than managing similarly, classified into three groups of ABC in accordance with the importance of an analysis to determine the importance of the time of administration.

Procedure of ABC analysis is as follows.

① aggregate sales for each commodity, sorted in descending order of the amount of money in the commodity group.

② sales to total rearranged the order to in ①, to calculate the cumulative component ratio of total net sales.

Based on the ③ sales cumulative composition ratio to divide the items into three classes of A( $\leq 70\%$ ) · B( $\leq 95\%$ ) · C( $\leq 100\%$ ).

## 3. METHOD OF DETERMINING ITEM-SETS AND SELLING PRICE

In this study, the method of determining the item-sets and selling price is proceeded in the flow as shown in **Figure 1**.

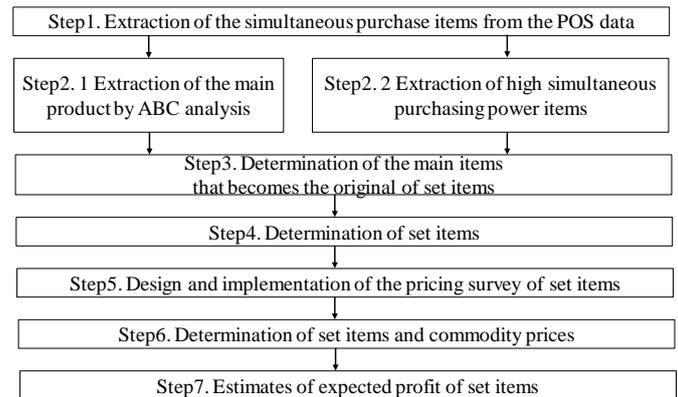


Figure 1: The procedure of this study

### 3.1 METHOD OF DETERMINING THE MAIN ITEMS IN THE ITEM-SETS

Using association analysis, we rearrange items in descending order of association analysis. In addition, to extract the items of Category A by using the ABC analysis. From among them in each time zone to determine the 1- and 2-position of the item. At that time, all of association analysis in the time zone is high, other than the general milk 1L is a mainstay item, the general milk 1L, cigarettes, pastries, rice

balls, rice balls to be removed from the main item. Flagship item go out the candidate two by two in each time zone.

### 3.2 METHOD OF DETERMINING THE SUPPLIED ITEMS IN THE ITEM-SETS

As set commodity, a commodity that comes with the main item, a decision from the ranking of goods that have been bought along with the main item.

Among the items that are sold together, taking into account the results of the ABC analysis, in addition to the combination of the items of Category A, B division, make the goods from one set of items candidates of Category C.

Comes with the item for each core item, go out the candidates one by three.

### 3.3 PRICE SETTING METHOD

With respect to the proposed each item-set candidate, when the purpose items to go to the convenience store had been sold as (○○ circle argument to purchase at the same time) item-sets, After that it was less than what yen, rather than the purpose items separately, I had you answer to 100 people about what to buy as a item-set at the same time.

For the amount of money, plus the amount of the selling price of the items separately, from the amount obtained by rounding off as easy-to-understand, 10 yen discount, 20 yen discount, to be carried out a discount of 30 yen discount. It goes down from there to the cost, capped with a ○ when there in the combination and the price that would want to buy as an item-set among them.

### 3.4 CALCULATION OF ITEM-SETS' EXPECTED PROFIT

#### Setting the cost of items

In this study, the cost of items in the convenience store taking into account the selling, general and administrative expenses and the purchase cost, is defined as it is 70% of the selling price.

#### Method of calculating the expected profit

From questionnaire of aggregate data, by performing a discount to what extent, about what can be the induction of simultaneous purchase, it is possible to further make a profit due to item-sets sales, reveal by calculating the expected profit.

$$\text{Profit} = \text{selling price} - \text{Cost} \quad (3.1)$$

## 4. ADAPTATION IN CONVENIENCE STORE

### 4.1 CALCULATION OF ITEM-SETS'

### EXPECTED PROFIT

In this study, we use the POS data of convenience store to calculate ABC analysis and association analysis. The information about the POS data is as follows:

Period:2010.5.1~2011.4.30

Store location: 4 stores in front of subway station

Receipt number:1~83407

Item number:6372

Maker number:485

It also includes the information of price, sales number, age, day of the week, sex etc.

### 4.2 CALCULATION OF ITEM-SETS' EXPECTED PROFIT

Before we start the association analysis, it is necessary to change the POS data into transactional data (0-1 type). The transaction data, customer is data collected set of items purchased by a single transaction, also called basket data. Each row of the transaction data corresponds to a single receipt, each column corresponds classification name. It has been described 1 if the purchase column of the item, if not purchased a item of the column 0 is described. If each line to the receipt number can be represented what number of people had bought which item. Also how to manage each column order to facilitate the management of each item is divided into the same classification name. Receipt of the POS data in each column using a pivot table feature of Excel number, one to the goods that have been bought to use VBA to create a table of classification name on each line, as put in Table 3-2 and 0 in the blank part a 0-1 type was converted to the transaction data of.

### 4.3 ANALYSIS RESULT

#### 4.3.1 RESULT OF ITEM-SETS

In this section, we show the result of ABC analysis and association analysis

Table1.The result of breakfast time(6~11a.m.)

Rank	Item Name	Association Analysis	Sales Number	ABC Analysis
1	紙パック入り紅茶飲料	8.3450	131	A
2	紙パック入り果汁100%飲料	5.9310	117	A

3	チョコバー・粒チョコ レート	5.5546	158	A
4	ロールパン・コッペ パン	5.0762	148	A
5	ペットボトル入り健 康茶飲料	4.8999	170	A
6	サラダ	4.1924	279	A
7	紙パック入り野菜ジ ュース	4.0907	220	A
8	ペットボトル入りミネ ラルウォーター	3.7987	307	A
9	ハードヨーグルト	3.7521	206	A
10	即席カップ中華そば	3.3978	203	A

Table2.The result of lunch time(12a.m.~5p.m.)

Rank	Item Name	Association Analysis	Sales Number	ABC Analysis
1	ペットボトル入り無 果汁炭酸飲料	10.2444	88	A
2	栄養補給ドリンク	8.8574	74	A
3	せんべい	8.7846	91	A
4	鶏卵	8.7343	108	A
5	ハードヨーグルト	8.7106	235	A
6	ドリンクヨーグルト	8.2725	121	A
7	マドレーヌ・ケーキ マフィン	7.7439	69	A
8	サラダ	7.3256	210	A
9	スティック包装のど あめ	6.9142	97	A
10	ジャガ芋系スナック	6.9016	91	A

Table3.The result of breakfast time(6~11p.m.)

Rank	Item Name	Association Analysis	Sales Number	ABC Analysis
1	スティック包装のど あめ	9.3731	199	A
2	イカ珍味	6.9468	112	A
3	チョコレート菓子	6.8167	134	A
4	マドレーヌ・ケーキマ フィン	6.4179	117	A
5	小麦系スナック	6.1492	103	A
6	アイスモナカ	5.9131	88	A
7	コーン系スナック	5.7567	112	A
8	揚げせんべい・あら れ・おかき	5.7467	89	A
9	チューハイ	5.6182	186	A
10	チョコバー・粒チョコ レート	5.5822	247	A

Rank	Item Name	Association Analysis	Sales Number	ABC Analysis
1	紙パック入り紅茶飲料	131		
1	菓子パン	24	A	
2	サンドイッチ	17	A	
3	おにぎり・おむすび	10	A	
4	缶入りコーヒー飲料	6	A	
5	弁当	6	A	
6	ジャガイモ系スナック	4	A	
7	デニッシュ・ペストリー	4	A	
8	ロールパン・コッペパン	4	A	
9	サラダ	3	A	
10	その他レギュラーアイス	3	B	

Table4.The result of supported items at breakfast time(紙パ  
ック入り紅茶飲料)

Rank	Item Name	Sales Number	ABC Analysis
	紙パック入り紅茶飲料	131	
1	菓子パン	24	A
2	サンドイッチ	17	A
3	おにぎり・おむすび	10	A
4	缶入りコーヒー飲料	6	A
5	弁当	6	A
6	ジャガイモ系スナック	4	A
7	デニッシュ・ペストリー	4	A
8	ロールパン・コッペパン	4	A
9	サラダ	3	A
10	その他レギュラーアイス	3	B

Table5.The result of supported items at breakfast time(紙パ  
ック入り果汁100%飲料)

Rank	Item Name	Sales Number	ABC Analysis
	紙パック入り紅茶飲料	117	
1	菓子パン	33	A
2	サラダ	16	A
3	おにぎり・おむすび	14	A
4	弁当	14	A
5	その他の調理パン	10	A
6	たばこ	8	A

7	チルドカップ入りコーヒー乳飲料	8	A
8	ハードヨーグルト	7	A
9	チルドシュークリーム	6	B
10	大福もち	6	C

Table6.The result of supported items at lunch time  
(ペットボトル入り無果汁炭酸飲料)

Rank	Item Name	Sales Number	ABC Analysis
	ペットボトル入り無果汁炭酸飲料	88	
1	菓子パン	22	A
2	おにぎり・おむすび	11	A
3	缶入りコーヒー飲料	9	A
4	普通食パン	9	A
5	コーンアイス	7	B
6	アルドショートケーキ・パイ	7	A
7	ペットボトル入り緑茶飲料	7	A
8	コーヒー乳飲料	6	A
9	せんべい	6	A
10	フルーツ入りヨーグルト	6	A

Table7.The result of supported items at lunch time  
(栄養補給ドリンク)

Rank	Item Name	Sales Number	ABC Analysis
	栄養補給ドリンク	74	
1	その他栄養補給食品	14	C
2	菓子パン	14	A
3	板チョコレート	10	B
4	おにぎり・おむすび	8	A
5	ペットボトル入りスポーツ飲料	8	A
6	ペットボトル入り健康茶飲料	5	B
7	チョコパー・粒チョコレート	4	A
8	アルドショートケーキ・パイ	4	A
9	ドーナツ	4	A
10	ペットボトル入りミネラルウォーター	4	A

Table8.The result of supported items at dinner time  
(スティック包装のどあめ)

Rank	Item Name	Sales Number	ABC Analysis
	スティック包装のどあめ	199	
1	まんじゅう	36	B
2	ペットボトル入りミネラルウォーター	18	A

3	パウンドケーキ	16	B
4	弁当	16	A
5	菓子パン	15	A
6	即席ワナン	15	B
7	栄養補給錠剤・錠菓	12	C
8	レトルトカレー	10	B
9	クラッカー	9	B
10	ぬれディッシュ	9	C

Table9.The result of supported items at dinner time  
(イカ珍味)

Rank	Item Name	Sales Number	ABC Analysis
	イカ珍味	112	
1	菓子パン	20	A
2	チューハイ	14	A
3	おにぎり・おむすび	13	A
4	チルドプリン	12	A
5	即席カップ中華そば	8	A
6	缶入りコーヒー	6	A
7	紙パック入り果汁100%飲料	6	A
8	即席カップうどん	6	A
9	カクテルドリンク	5	A
10	カップアイス	5	A
11	ジャガイモ系スナック	5	A
12	チョコパー・粒チョコレート	5	A
13	パーティタイプアイス	5	A
14	ペットボトル入りコーヒー飲料	5	B

### 4.3.2 QUESTIONNAIRE RESULT

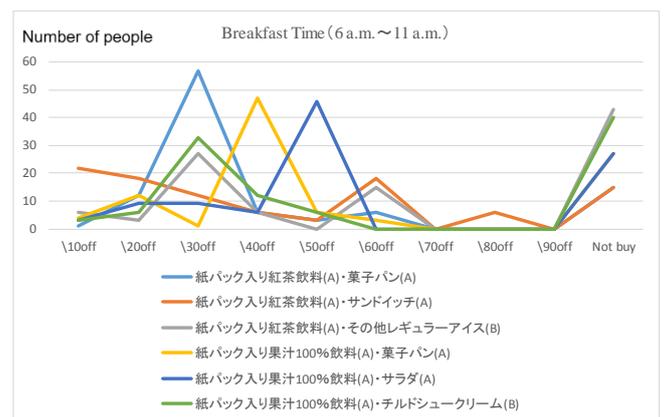


Figure2. The result of item-sets at breakfast time

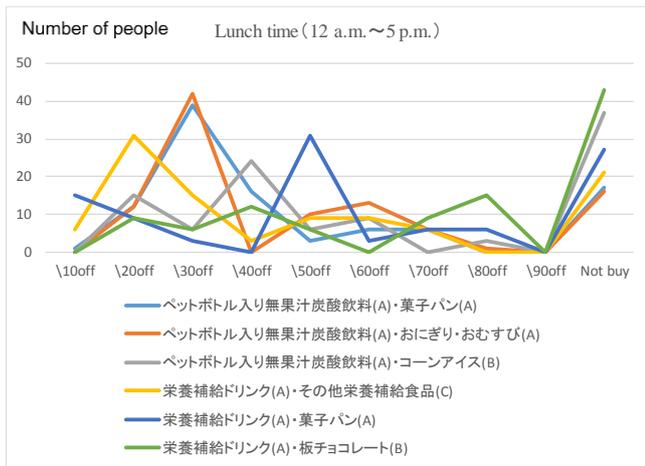


Figure3. The result of item-sets at lunch time

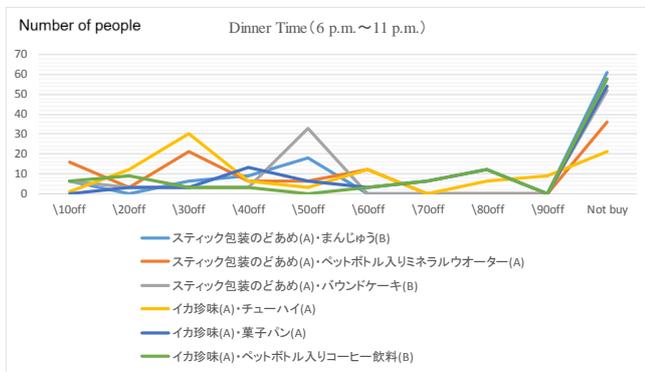


Figure4. The result of item-sets at dinner time

### 4.3.3 CALCULATION OF ITEM-SETS' EXPECTED PROFIT



Figure5. The result of profit at breakfast time

Here, the name of No.1~6 in Figure5 are showed below.

- 1: 紙パック入り紅茶飲料(A)・菓子パン(A)
- 2: 紙パック入り紅茶飲料(A)・サンドイッチ(A)

- 3: 紙パック入り紅茶飲料(A)・その他レギュラーアイス(B)
- 4: 紙パック入り果汁100%飲料(A)・菓子パン(A)
- 5: 紙パック入り果汁100%飲料(A)・サラダ(A)
- 6: 紙パック入り果汁100%飲料(A)・チルドシュークリーム(B).

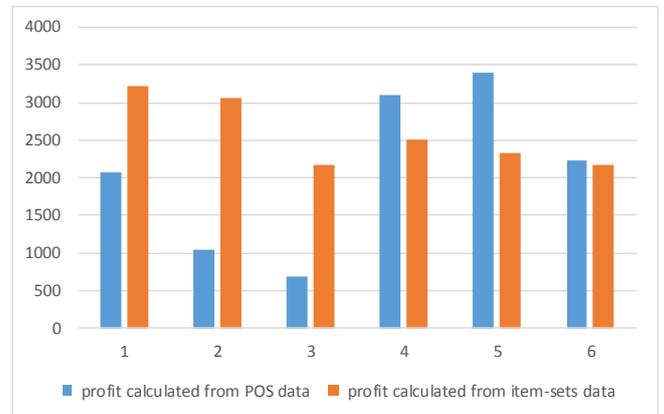


Figure6. The result of profit at lunch time

Here, the name of No.1~6 in Figure6 are showed below.

- 1: ペットボトル入り無果汁炭酸飲料・菓子パン
- 2: ペットボトル入り無果汁炭酸飲料・おにぎり・おむすび
- 3: ペットボトル入り無果汁炭酸飲料・コーンアイス
- 4: 栄養補給ドリンク・その他栄養補給食品
- 5: 栄養補給ドリンク・菓子パン
- 6: 栄養補給ドリンク・板チョコレート

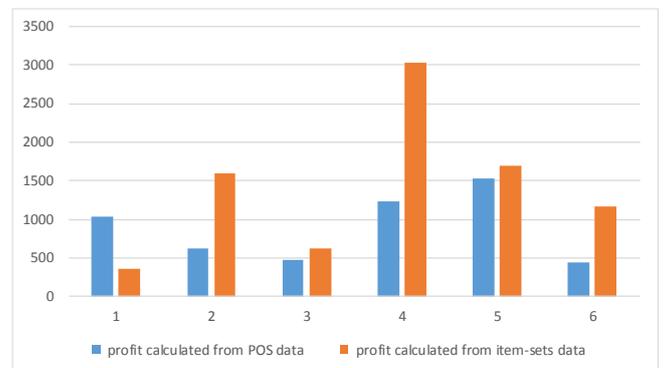


Figure7. The result of profit at dinner time

Here, the name of No.1~6 in Figure6 are showed below.

- 1: スティック包装のどあめ・まんじゅう
- 2: スティック包装のどあめ・ペットボトル入りミネラルウォーター
- 3: スティック包装のどあめ・パウンドケーキ
- 4: イカ珍味・チューハイ
- 5: イカ珍味・菓子パン
- 6: イカ珍味・ペットボトル入りコーヒー飲料

Using association analysis result of the set-items designed, it has become possible to improve the benefits in each time period. We continue to discuss the results. Firstly, the purpose products in each time period were found by performing the ABC analysis and association analysis. It is considered to be able to induce more simultaneous purchase if the prices are set having exact change. The person who answered like below was less. We do not buy things that are a combination of the trade with each other, which is classified as the Category A in the ABC analysis. On the other hand, the percentage of people who answered “Even the result of association analysis was high, we do not buy a combination with products other than Category A in the ABC analysis” was greater

## 5. CONCLUSION

In this study, on the basis of the POS data, we analyze the Simultaneous purchase influence of the time zone of the item-sets by association analysis. By analyzing whether the customer has purchased any products at the same time, we perform item-sets design. In addition, from the implementation of the questionnaire, it was pricing of item-sets. By performing the discount due to item-sets sales, it is possible to perform the induction of simultaneous purchase more than ever, contributing to the improvement of the profit is expected.

Further, by performing their own set merchandise that has not been done to date, considered also appears the item-sets customers are looking. Thereby, it believed it is possible to perform differentiation for each company.

## REFERENCES

- Agrawal R., Imielinski T. and Swami A., (1993) Mining association rules between sets of items in large databases, *Proceedings of 1993ACM-SIGMOD Int. Conf. Management of Data*, Washington,D.C. , 207-216.
- Agrawal R. and Srikant R. (1994) Fast algorithms for mining association rules. *Proceedings of the 20th Int'l Conference on Very Large Data Bases*, Santiago, Chile, pp. 487–499.
- ALBERT Inc.(2016) Method of commodity analysis: ABC analysis and association analysis.  
[http://www.albert2005.co.jp/technology/marketing/item\\_analysis.html](http://www.albert2005.co.jp/technology/marketing/item_analysis.html)
- Garbage NEWS.com(2015)The top four companies of sales and number of stores in the convenience store industry.
- G.Fang et al. (2015) An efficient algorithm for mining frequent closed itemsets, *Informatica* (39)pp.87-98.
- Hidetaro(2013) Association analyses by R.  
<http://qiita.com/hidetarou2013@github/items/dc3b448542c5fce7a6ce>
- Inter-Knowledge Partners(2011) Trends of the past growth in

the convenience store industry.

<http://www.ikpi.co.jp/bizblog/blog20111121.html>

Kim M.(2007)*Data science by R*, Morikita Publishing Co., Ltd.

Nakamura N.(2009) *Data science by R- Multidimensional data analysis method-*, Kyoritsu Shuppan Co., Ltd.

NTT DATA Mathematical Systems Inc. (2015) Association analyses

[http://www.msi.co.jp/vmstudio/materials/tech\\_web/association.html](http://www.msi.co.jp/vmstudio/materials/tech_web/association.html)