International Journal of scientific research and management (IJSRM)
\||Volume||3||Issue||7||Pages|| 3324-3330||2015||
Website: www.ijsrm.in ISSN (e): 2321-3418

Association Rule Mining using Apriori algorithm for work-related beliefs of Generation X and Generation Y.

Dr. Bhagirathi Nayak [1] CA. Vijaya Batth [2]

Asst. Professor (FMS),
Head (IT & Systems),
Sri Sri University, Cuttack, Odisha.
Phone: +91 8456023038
Email: bhagirathi.n@srisriuniversity.edu.in
Asst. Professor (FMS),
Finance and Accounts,
Sri Sri University, Cuttack, Odisha.
Phone: +91 9437006081
Email: vijayabatth@srisriuniversity.edu.in

Abstract
Market-Basket Analysis is a process to analyze the habits of buyers to find the relationship between different items in their market basket. The discovery of these relationships can help the merchant to develop a sales strategy by considering the items frequently purchased together by customers. In this article, the data mining with market basket analysis method is implemented, for work-related beliefs of Generation X and Generation Y. The data testing is conducted from collection of questionnaire. There are 20 questions in our questionnaire. The data for frequent questions performed by Apriori algorithm to get the relation that often appear in the database. The questions are generated association rules after decoding. One frequent question can generate association rules and find the confidence. The test results show, the application can generate the information what kind of work-related beliefs are frequently bought in the same Generation according to the Association Rules criteria. Results from the mining process show a correlation between the data (association rules) including the support and confidence that can be analyzed. This information will give additional consideration for work-related beliefs of Generation X and Generation Y to make further decision.

Key words: Generation X and Y, Apriori algorithm, Association Rules, Market-Basket Analysis.

Introduction

A new generation of employees with new expectations is entering the workplace. They are known by many names, but most will recognize them as the Millennial or Generation Y (Gen Y). The 1999-2000 Occupational Outlook Quarterly has indicated that in the 1998-2008 ten year period, the number of 35 to 44 year old Generation X (Gen X) workforce members is expected to decline by 7%, additionally those in Gen X that are in the 25 to 34 year old age category are expected to fall by 1%, whereas those from Gen Y in the age group between 16 to 24 years old will increase by 15% (Cole, Lucas, & Smith, 2002). To get some perspective of what these changes in headcount actually represent, it is approximated that while there are 75 million Traditionalists, 76 million Baby Boomers, and 44 million Gen X members, Gen Y is closer to 80 million (Sujansky, 2004). Furthermore, although Gen Y members are already in the workforce, they have not begun to reach critical mass. As a result, managers are likely to be required to deal with the generational differences that appear to exist among employees and understand the unique needs of Gen Y (Smola & Sutton, 2002; Sujansky, 2004). Only in this manner will the inevitable clash amongst the generations become more controllable, lessening the possible misunderstandings, miscommunications, and mixed

signals among employees of different generations (Smola & Sutton, 2002). The entrance of Gen Y to the workplace and this generation's unique work attributes have created quite the commotion in the workplace, as employers appear to be scrambling to find out everything they can about them (Raines, 2002). Therefore, the purpose of the present study is to explore possible differences between Gen X and Gen Y on their work-related perceptions. This article begins by describing the two generations of focus in this article. Then, previous industry and research findings concerning potential differences between Gen X and Gen Y on three work-related beliefs (i.e., work engagement, career development, and teamwork) are reviewed.

Generational Categories

A generational group, often referred to as a cohort, includes those individuals who share historical and/or social life experiences (Kupperschmidt, 2000; Smola & Sutton, 2002; Weston, 2006). Such experiences unite people of the same generation, lead them to come to share common values, and a large number of them may then come to experience the world in similar ways (Patterson, 2007; Smola & Sutton, 2002). Although each individual within a generation is unique, each generation tends to develop a collective personality that influences the way members live their lives, including feelings toward authority and organizations, their participation in and desires from work, and even how they plan to attain those desires (Kupperschmidt, 2000; Smola & Sutton, 2002; Weston, 2006). Although it is inappropriate to make generalizations about the individuals in each generational cohort, it is inevitable to recognize that those who are born in the same era have had common influential experiences that predispose them to similar expectations, and as a result such life experiences are what tend to distinguish one generation from another (Patterson, 2007; Smola & Sutton, 2002). Although there appears to be agreement as to what to label a generation that is defined by the shared birth years and significant life events, there are inconsistencies as to what exact years to use to classify the generations (Smola & Sutton, 2002). Of the two generational roups in this study, there is less agreement as to what the birth years of Gen X are. Some report that they begin in the mid-1960s and extend until 1975, whereas others end the Gen X years in 1982 (Barton & Skiba, 2006; Greene, 2003; Patterson, 2007; Smola & Sutton, 2002; Sujansky, 2004; Vejar, 2008). This study defines Gen X as those who were born between 1965 and 1977. As for Gen Y, considering they follow Gen X, their generation beginning year also tends to vary ranging from 1978 to 2000 (Greene, 2003; Howe, 2004; Leo, 2003; Patterson, 2007; Smola & Sutton, 2002; Sujansky, 2004; Vejar, 2008). Since this study is about the experience of real-world working Gen Y-ers, the year range that will define Gen Y will be that of 1978 to 1990. With the retirement of some of the earlier generations, Gen X-ers are and will be the experienced employees and managers as Gen Y continues to enter the workforce (Smola & Sutton, 2002). These two generations are the ones I choose to focus on due to the fact that these two generations will be working closely together for at least the next 20 years.

Generation X

In a study about the civic engagement of Generation X, the U.S. Census Bureau defined this segment of the population as consisting of individuals born between 1968 and 1979. However, the upper limit of Generation X in some cases has been as high as 1982, while the lower limit has been as low as 1963 (Karp et al., 2002). This generation was also called the baby bust generation, because of its small size relative to the generation that preceded it, the Baby Boom generation. The term Generation X spread into popular parlance following the publication of Douglas Coupland's book about a generation of individuals who would come of age at the end of the 20th century. Members of Generation X [Hereinafter Xers] are the children of older boomers, who grew up in a period of financial, familial and societal insecurity. They witnessed their parents get laid off and the decline of the American global power. They grew up with a stagnant job market, corporate downsizing, and limited wage mobility, and are the first individuals predicted to earn less than their parents did. They have grown up in homes where both parents worked, or in single parent household because of high divorce rates, and as such, became latchkey kids forced to fend for themselves (Karp et al., 2002). They were influenced by MTV, AIDS and worldwide competition and are accustomed to receiving instant feedback from playing computer and video games (O'Bannon, 2001). Among the characteristics attributed to Xers, the following appear most often. They aspire more than previous generations to achieve a balance between work and life (Jenkins, 2007; Karp et al, 2002; www.valueoptions.com) they are more independent, autonomous and self-reliant than previous generations

(Jenkins, 2007; Zemke et al., 2000) having grown up as latchkey kids. They are not overly loyal to their employers (Bova & Kroth, 2001; Karp et al, 2002; The National Oceanographic and Atmospheric Association Office of Diversity, 2006) although they have strong feelings of loyalty towards their family and friends (Karp et al., 2002). They value continuous learning and skill development (Bova & Kroth, 2001). They have strong technical skills (Zemke et al., 2000), are results focused (Crampton & Hodge, 2006), and are "ruled by a sense of accomplishment and not the clock" (Joyner, 2000). Xers naturally question authority figures and are not intimidated by them (The National Oceanographic and Atmospheric Association Office of Diversity, 2006; Zemke et al., 2000). Money does not necessarily motivate members of this generation, but the absence of money might lead them to lose motivation (Karp et al., 2002). They like to receive feedback (The National Oceanographic and Atmospheric Association Office of Diversity, 2006), are adaptable to change (Zemke et al., 2000) and prefer flexible schedules (Joyner, 2000). They can tolerate work as long as it is fun (Karp et al., 2002). They are entrepreneurial (The National Oceanographic and Atmospheric Association Office of Diversity, 2006), pragmatic (Niemiec, 2002), and creative (The National Oceanographic and Atmospheric Association Office of Diversity, 2006). Although they are individualistic, they may also like teamwork, more so than boomers (Karp et al., 2002).

Generation Y

The lower limit for Generation Y may be as low as 1978, while the upper limit may be as high as 2002, depending on the source. Members of Generation Y may include individuals born between 1980 and 1999 (Campton & Hodge, 2006); 1978 and 1995 (The National Oceanographic and Atmospheric Association Office of Diversity, 2006); 1980 and 2002 (Kersten, 2002); and 1978 and 1988 (Martin, 2005). The label associated with this generation is not yet finalized. Current labels include Millenials, Nexters, Generation www, the Digital generation, Generation E, Echo Boomers, N-Gens and the Net Generation. Members of the generation have labeled themselves as the Non-Nuclear Family generation, the Nothing-Is-Sacred Generation, the Wannabees, the Feel-Good Generation, Cyberkids, the Do-or-Die Generation, and the Searching-for-an-Identity Generation. This generation has been shaped by parental excesses, computers (Niemiec, 2000), and dramatic technological advances. One of the most frequently reported characteristics of this generation is their comfort with technology (Kersten, 2002). In general, Generation Y shares many of the characteristics of Xers. They are purported to value team work and collective action (Zemke et al., 2000), embrace diversity (The National Oceanographic and Atmospheric Office of Diversity, 2006), be optimistic (Kersten, 2002), and be adaptable to change (Jenkins, 2007). Furthermore, they seek flexibility (Martin, 2005), are independent, desire a more balanced life (Crampton & Hodge, 2006), are multi-taskers (The National Oceanographic and Atmospheric Office of Diversity, 2006), and are the most highly educated generation. They also value training (www.valueoptions.com). They have been characterized as demanding (Martin, 2005), and as the most confident generation (Glass, 2007). Like Xers, they are also purported to be entrepreneurial, and as being less process focused (Crampton & Hodge, 2006).

Problem Statement

The aim of this article is to find the work-related beliefs relation between the Generation X and Y from IT Sectors, Banking and Manufacturing sectors. These sectors employees those who are agree or disagree determine the questions that are frequently used together by the employee. Here the four major parameters through twenty numbers of questions namely i) stability ii) innovation iii) challenging and iv) security to measure greatest work-place benefit among the two generations.

Market Basket Analysis

This is a process which analyses the habits of the buyer to find a relationship between different items on their shopping cart (market basket). The discovery of these relationships can help the seller to develop a sales strategy to consider items frequently purchased together by customers. For example, if a buyer buys bread, how likely they will buy butter on the same transaction.

Association Rules

Association rule is a procedure which is looking for a relationship among an item with other items. Association rule is usually used "if" and "then" such as "if A then B and C", this shows if A then B and C. To determine the Association's rules, it needs to be specified the support and confidence to restrict whether the rule is interesting or not.

- Support: A measure that indicates how much the level of dominance of an item or item set of the overall transaction
- Confidence: A measure that shows the relationship between items in a conditional (e.g. how frequently purchased item B if the person buying the item A).

Apriori Algorithm

Apriori is an iteration approach known as level-wise search, where k-item set is used to explore (k +1)-item set. First, a collection of 1-itemset is found by checking the database to accumulate counts for each item, and records of the items. The result is denoted by L1. Furthermore, L1 is used to find L2, a collection of 2-itemset that is used to search for L3, and so on until there are no k-item set to be found. L_k invention requires inspector's entire database [1]. To increase the efficiency of frequent item set search, it is used Apriori rule which says "All parts not empty from the frequent item set are also frequent". When written in the form of pseudo code, Apriori algorithm is as follows.

C_k: Candidate itemset of size k
L_k : frequent itemset of size k
L_1 = {frequent items};
for (k = 1; L_k !=∅; k++) **do begin**
  C_{k+1} = candidates generated from L_k;
  for each transaction t in database do
    increment the count of all candidates in C_{k+1}
  L_{k+1} = candidates in C_{k+1} with min_support
  end
return $\cup_k L_k$;

While the pseudo code of the formation of joint candidate item set is given below after the results obtained, just made a strong association rule from these results. This can be obtained by following rule strength measures.

a) Support

The rule $X \Rightarrow Y$ holds with support s if s% of transactions in D contains $X \cup Y$. Rules that have as greater than a user-specified support is said to have minimum support.

$$\text{Support} = \frac{\text{number of transactions that contain antecedents}}{\text{total number of transactions}}$$

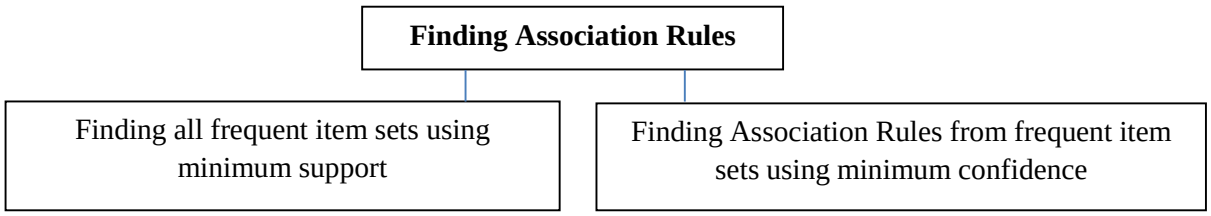
b) Confidence

The rule $X \Rightarrow Y$ holds with confidence c if c% of the transactions in D that contain X also contain Y. Rules that have a c greater than a user-specified confidence is said to have minimum confidence.

$$\text{Confidence} = \frac{\text{support}\ (X \cup Y)}{\text{support}\ (X)}$$

Basic architecture

1) Input Data: Giving the existing data set
2) Training the data: The algorithm will learn about the data
3) Building the Model: based on that knowledge the model will be build.
4) Knowledge: Obtain rules form the model
5)



6) Generating Association Rules
7) Decision making: Take decisions based on the rules

Association Rules generated using Data mining tools.

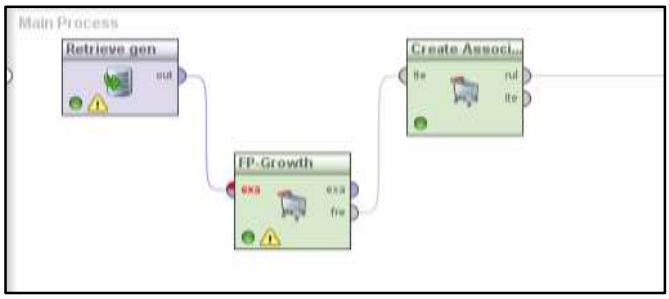


Figure 1: Main Process to create Association Rules

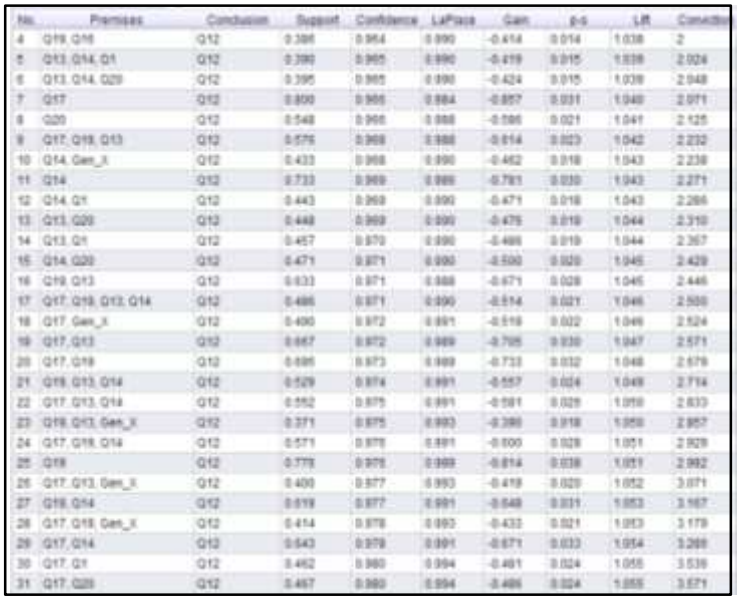


Figure 2: Supports and confidence of Generation X and Y

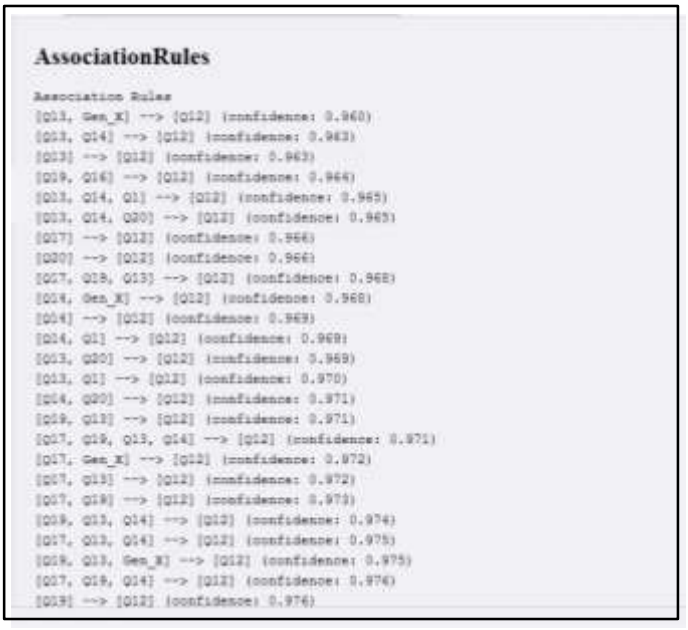


Figure 3: Create association rules of Generation X and Y

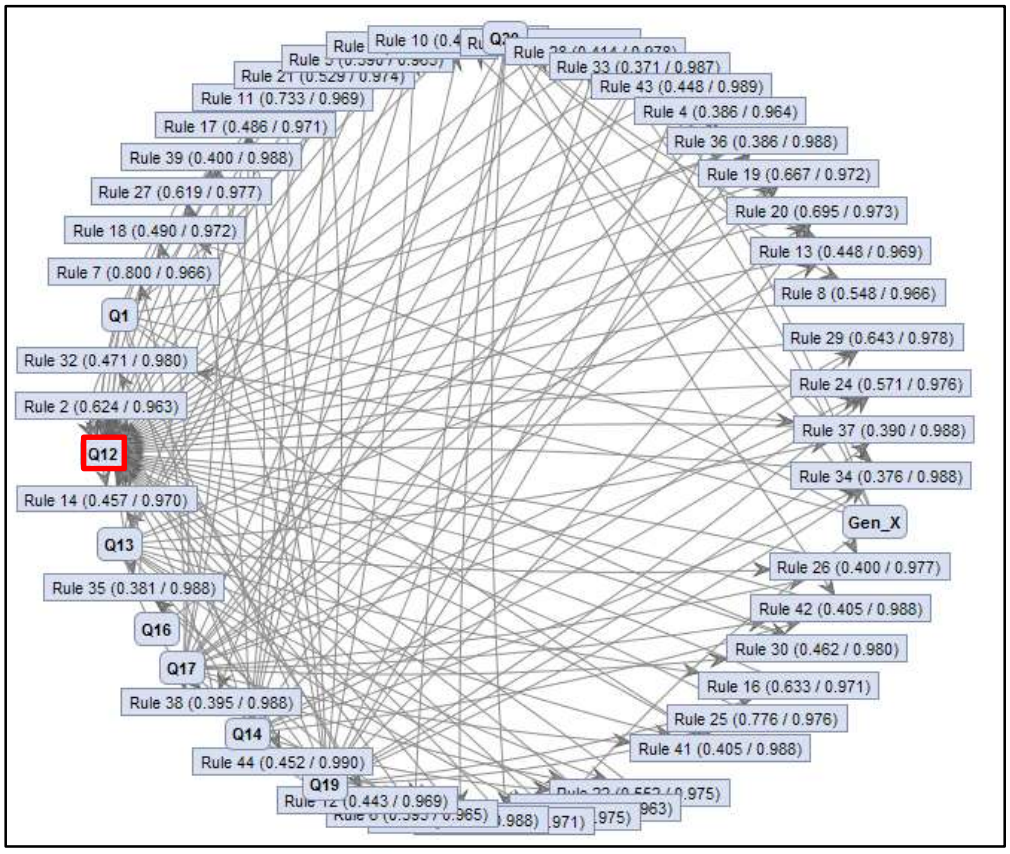


Figure 4: Node graph of Association rules for Generation X and Y

Conclusion

This article is basically for individual work-related beliefs of Generation X and Generation Y employees as respondents in IT, Banking and manufacturing sectors. We have noted that Generation X and Generation Y differences exist but we found through our analysis that both are associate in work-related beliefs in challenging parameters out of our four parameters. We also found that the parameter seeking readiness for accepting challenging project is strongly associated in all types of environment. Having such a sample gives

credibility to the results knowing that the findings came from individual question and environment where these results could be similarly applied. In the survey, data were collected from more than 500 employees from said sectors. The above analysis result shows that our question (Q12. Are you ready for accepting challenging project?) is strongly associated in Generation X and Generation Y. Shedding light on generational differences gives organizations the knowledge to better understand their current workforce and make the workplace a more enjoyable place for all generations and productive for the organization.

References

[1] A Publication of the Defense Acquisition University http://www.dau.mil

[2] Armour, S. (2005), "*Generation Y: They've arrived at work with a new attitude*", *USATODAY.com.* Retrieved April 25, 2008, from http://www.usatoday.com/money/workplace/2005-11-06-gen-y_x.htm

[3] D. Olson and S. Yong (2007), "*Introduction to Business Data Mining*". New York: McGraw-Hill.

[4] Egeler, D. (2003). "*Mentoring Millennials: Shaping the next generation",* NavPress Publishing Group, ISBN: 1576833828

[5] H. Jiawei and K. Micheline, (2001), "*Data Mining: Concepts and Techniques*". Morgan Kaufmann.

[6] Howe, N., Strauss, W., & Matson, R. J. (2000), "*Millennials rising: The next great generation",* Vintage, ISBN: 0375707190

[7] http://scholarworks.sjsu.edu/etd_theses

[8] Huntley, R. (2001), "*The world according to Y: Inside the new adult generation",* HRD Press, ISBN: 1741148456

[9] Marston, C. (2007), "*Motivating the "What's in it for me" workforce: Manage across the generational divide and increase profits"* Wiley, ISBN: 0470124148

[10] Mohit K. Gupta and Geeta Sikka (2003), "*Association Rules Extraction using Multi-objective Feature of Genetic Algorithm*" , International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6, June 2013