Apply clustering to analyze categorical data in longitudinal studies

Mohammad Mahdi Hassan[1†]
Computer Science Department
Qassim University
Al Qassim, Saudi Arabia

Martin Blom[2††]
Computer Science Department
Karlstad University Karlstad, Sweden

Gufran Ahmad Ansari[3†††]
Department Computer Applications
BSARC Institute of
Science & Technology
Vandalur, Chennai, India

Summary
It is common to collect data from practitioners in the software engineering field using surveys and questionnaires. This data is usually analyzed using descriptive statistics where the entire population is considered as an undivided group, sometimes complemented by sampling methods to obtain variations within the sample. In many cases, the survey population is partitioned into smaller groups by using available background knowledge of the participants. These techniques are valid, but can only reveal opinion diversity if that correlates with the background variables, and fail to identify sub-groups across multiple background variables. The existing approaches can thus capture the general trends but might miss opinions of different minority sub-groups. This problem becomes more complex in longitudinal studies where minority opinions might fade or resolute over time. Data from longitudinal studies may contain patterns which can be extracted using a clustering process. These patterns may unveil supplementary information and draw attention to alternative viewpoints than those exhibited by the sample population as a whole. This approach may reveal the range of opinion variations between diverse groups over time and makes it possible to identify the minorities. In our research, we have investigated the suitability of clustering techniques for analyzing categorical data from longitudinal studies.

Key words:
Empirical Survey, Longitudinal Study, Clustering, Partitioning, Grouping, Data Mining, Expert Opinion, Diversity.

1. Introduction

Various forms of data are generated during the software development process. Some typical forms of data are as follows [1]- 1) Program Code, 2) Trace logs, 3) Design and Code revision history, 4) Defect databases etc.

Recently large investments in automation of software processes have been made to reduce the cost and to improve the quality of the product. Various automation processes not only generate traditional forms of data as listed above but also make it possible to both preserve and obtain new forms of software engineering data. New forms of data like, 1) Test cases, 2) System build traces, 3) Team and Personal data, and 4) Development and process data etc., are readily available in many software development organizations [2].
Besides the above mentioned data that can be collected both in industry and academia, it is also common, especially in academia, to conduct surveys to analyze personal opinions of various software engineering artefacts, processes and techniques.
According to Pfleeger [3], one of the major research domain of software engineering is gathering information from software development practitioners through surveys and analyze those data to get new insights. As online facilities and tools have become more available in recent years, collecting survey opinions frequently has been made easier, resulting in large amounts of survey data now existing in many software organizations.
This data is usually analyzed using descriptive statistics (like mean, median, variance, and various analytic tests) where the entire population is considered as an undivided group [4], sometimes complemented by sampling techniques to obtain variations within the sample [5]. In many cases, the population is segmented into smaller groups by using available background information. In most cases, this approach rarely exposes opinion variations precisely as alike opinions might exist in separate sections of the population, whereas in the same segments people can have alternate opinions. The problem becomes more complex in case of a longitudinal study[1] where minority

[1] A Longitudinal Study (LS) is an empirical research approach in which data is collected for the same subjects repeatedly over a time period. LS projects can continue over years or even decades. LS allows researchers to study changes over time through the same individuals who are observed over the study period. LS generates valuable empirical data. Moreover, LS allows changes over a long time to be traced. This approach suggests that the life of a process, practice or an entire system can be understood in a deeper way by observing the temporal aspects of various changes. The scope and strength of data gathered over an extended period of time is a piece of valuable empirical evidence which can be used to understand the study subject in a better way [6] issues related to our approach.

Manuscript received April 5, 2019
Manuscript revised April 20, 2019

opinions might fade or resolute over time. In our study, we applied clustering techniques on data gathered in a longitudinal survey in which the data was collected in categorical or numerical forms. The clustering approach divides the population without any perceived bias into different segments of subpopulations which to some extent have a similar view. Using this approach, we have observed the following benefits and opportunities:

- Reduced grouping manipulation, as groups are generated based on opinions. If background information is interesting for certain opinions, it can be incorporated with opinions during the clustering process.
- Exhibits opinion differences among the participants more accurately. Statistical variance [4] merely suggest general consensus or divergence, on the other hand, partitioning by clustering can reveal variation within each group and show intra-group disagreements and agreements.
- Identifying minorities which would not be recognized otherwise. If results are presented in an aggregated fashion, minority groups often lose their voices.
- Groups with distinct characteristics may be revealed by combining opinion differences with background information. These findings can lead to new hypotheses that in turn can inspire new research for investigating the groups and the hypotheses.
- In certain cases, some forms of relationship between different features of opinion only appear inside a cluster.
- In a longitudinal study statistically, similar groups can be identified which makes it possible to observe characteristics of similar opinions over time.

To investigate the application of clustering approach on LS (Longitudinal Studies) we used a longitudinal opinion survey conducted by a Swedish firm over four years. To analyze they used standard statistical methods and reached some general conclusions on the population as a whole. As per their analysis testing process was in good shape, but the requirement process lacks competence. By applying data mining techniques, we were able to find some significant groups within the participants who have different viewpoints than the claimed general conclusion both in terms of strength as well as in direction. During the study period, some of the opinions persisted over time whereas some others disappeared.

The remaining paper is structured as follows: Section II contains related work, in Section III we present the sample LS survey used for the study and provide relevant background information, Section IV is dedicated to analyzing the longitudinal study using clustering, in Section V we discuss some issues related to our approach. Finally, in Section VI we conclude with some future goals.

2. Related Works

After searching the literature on “data analysis of expert opinion survey” using various data mining techniques, there seems to be an inadequacy of research within the software engineering field in this regard. Where data mining research on survey data is being done is in marketing [7] and business oriented fields [8] mostly based on data collected from ordinary customers. The internet inflation has made it convenient for the business establishments to collect customers’ opinions through web forms [9] but other more traditional methods are also used (phone, paper-based etc.) but preserved in digital form. The aim of marketing and business research in this area is to understand consumption patterns and use this to improve services and increase revenues [10].

Conducting opinion surveys on software engineering professionals is one of the key research exercises. There exist conventional guidelines on how to analyze survey data, which in principle consists of rational examination approach using some simple statistical methods. Barbara A. Kitchenham [4] & [11] mentions such methods with a recommendation for using superior statistical techniques like Bayesian analysis. They found that Bayesian analysis approaches are not commonly used in software engineering studies and suggest to get some assistance from statisticians. John Moses [12], [13] & [14] has introduced a quality prediction model of software build on the experts’ opinion using Markov Chain Monte Carlo (MCMC) simulation and Bayesian inference. In general, descriptive statistical procedures, with some hypothesis tests, are used to examine opinion survey [15].

In the survey research domain, we can observe a kind of uneasiness towards using superior statistical methods as well as techniques based on machine learning, which are built on complex statistical and mathematical models [16]. Researchers are also dissuaded to use DM approaches for analysis due to the inadequate amount of data from those opinion surveys. We have observed a similar weakness regarding analyzing longitudinal studies too where only some simple statistics are used for analysis and comparison purposes [17]. On the other hand, Hassan and Blom [2] showed that by applying DM on opinion survey data one can observe different viewpoints which may be neglected and uninvestigated using conventional analytical and simple statistical methods. In our current study, we have extended their work to analyze longitudinal studies using clustering techniques.

Before presenting our planned study, we provide a short overview of DM tasks, according to Shaw et. al. [18] following tasks are significant during the DM process (also shown in Figure 1): 1) Dependency Analysis, 2) Class Identification, 3) Concept Description, 4) Deviation Detection, 5) Data Visualization. Hassan and Blom [2]

discussed some datamining related questioners in details in their study.
In our study we performed Class Identification for group identifications and Concept Description for characterization on sample survey data. Later sections provide the details of the process.

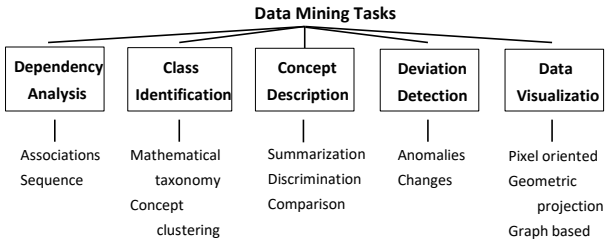

Fig. 1 General data mining tasks[18]

3. Sample Survey Overview

The survey was conducted by an IT education and consulting company, QTEMA [19], with the purpose of assessing the state-of-practice in the Swedish IT industry. The questionnaire form was comprised of 21 questions on various aspects related to working in the IT industry. It included background questions as well as questions on technical and non-technical aspects related to the development processes used by the participants.
The study was repeated annually and in this study, the questionnaires from 2010 to 2013 were used. On average the questionnaire was answered by 150 respondents each year. We have chosen to include both numerical and categorical data to evaluate the methodology on both kinds of data. The questions used in our study are listed below1, the answers and resulting grouping are presented in the next section.[2]:

1. "Which of the following sentences best describes what development methodology you use most often?"3
2. "Does your company/unit have a functioning organization and process for working with requirements?"
3. "Does your company/unit have a functioning organization and process for working with testing/verification/ validation?"
4. "How much time (in %) do you spend on requirements?"[5]
5. "How much time (in %) do you spend on test, verification, and validation?"
6. "How much experience do you have working with your current tasks?"
7. "Do you have the required competence to work professionally?"
8. "Which of the following best describes when in a typical development project, you meet the customer?'

The motivation to why these questions were selected out of the 21 possible questions was the focus of QTEMA and the possible correlation between those attributes and the development method used. The first three questions (1-3) were used for clustering categorical data and the following questions were used to understand the characteristics of the significant clusters.
The overall results from the study suggests that requirement related activities are in bad condition whereas the testing activities were in a good condition 4 . The overall conclusions are presented in Figures 2 and 3 respectively.

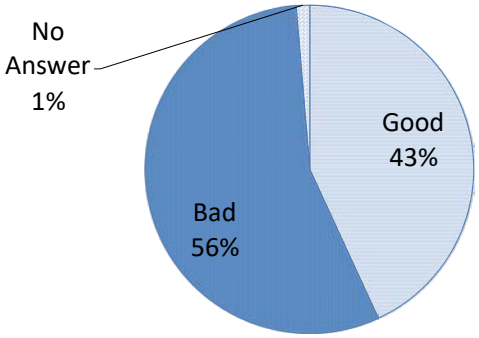

Fig. 2 Overall requirement status for year 2010

[1] We have focused on questions related to requirements and testing as these topics are the main focus of Qtema. We also tried to cluster using other questions initially, but it did not produce any interesting patterns. In larger datasets, the attributes would perhaps better be reduced by a systematic reduction process, but in this rather small dataset it was possible to do this ad hoc.
[2] Since the original study was in Swedish, the questions and answers have been translated by the authors.
[3] For readability we use the terms Traditional, Agile and Blend for options a), b) and c) respectively to show the answer to this question (see the distribution table 1 in next section). 5 This question, as well as the other questions related to time consumption, have been separated from a compound question to increase readability. The original question was - "How much time (in %) do you spend on the following activities? (For each activity you can state a number between 0-100, but the sum for all activities should be 100)". Since the respondents could enter numbers freely in these questionnaires, we made categories (based on cut points). To provide a bit more information we present not only the percentage of respondents but also the mean value (see the distribution table 1 in next section).
[4] Bad is an aggregation of the answers "To a very low degree" and "To a low degree", Good is an aggregation of the answers "To a very high degree" and "To a high degree" (see the distribution table 1 in next section).

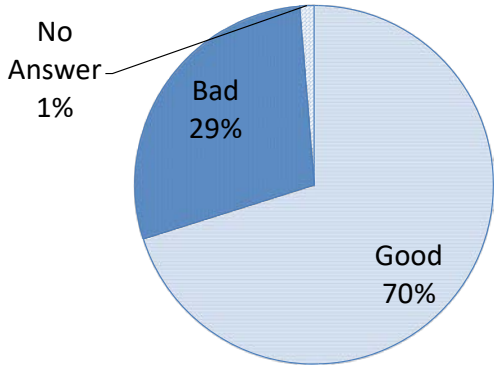

Fig. 3 Overall test status for year 2010

4. Analyzing Longitudinal Study

In our previous study [2], we have described an approach to identify and analyze interesting groups with a diverse opinion from a single year survey (i.e. data from the year 2010). Initially, the clustering process starts with a low expected number of clusters and then gradually increase the number. In each step, we identify cohesive and significant clusters and we labeled them. The process continues even as the overall clustering results deteriorate (see the log likelihood graph for different numbers of expected clusters in Figure 4). We stop the process when no new significant groups emerge. In each step, the size of identified groups may change but we recognize them based on their statistical closeness. Figure 5 shows the change of size of the different groups as we increase the expected number of clusters. For further analysis of each group, we extract each group's data when its size is largest.

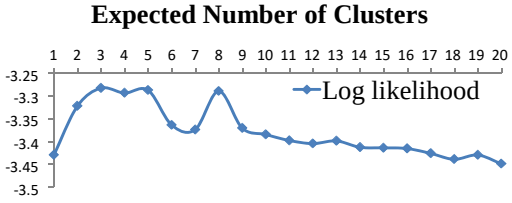

Fig. 4 Log likelihood for different settings of expected number of clusters

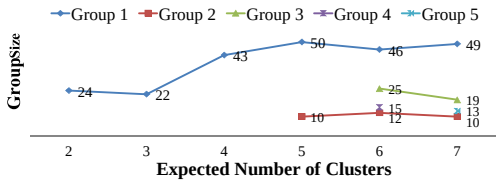

Fig. 5 Change of groups' size in 2010

Based on categorical data clustering we found five groups with diverse opinions from the 2010 survey. Those five groups are used as the primary groups for the longitudinal study.[1] We applied the same approach for other consecutive years to identify those groups.

In a longitudinal study finding an exact group based on the clustering process is a bit tricky. It is highly unlikely that a group found in a particular year will reappear exactly (in terms of statistics) in other years. On the other hand, even with some significant difference in some question, similar groups can be located based on statistical closeness. For categorical data, frequency distribution can be a good indicator. For numerical data mean can be used for that purpose but standard deviation needs to be checked.

In table 1 we show overall data distribution of survey questionnaires for years from 2010 to 2013. In the subsequent sections, we will show each of those five groups in different years. We analyze them to identify their consistency as well as their changes over time.

4.1 Group 1

This is the biggest group with an opinion that contradicts with the general conclusion of the survey. Most of the people of this group are confident in their requirement as well as the testing process. In Figure 6 we show the size of this group over study period (from 2010 to 2013). In Table 2 we show the question wise data distribution of this group in different years.

Some noticeable characteristics of this group:

1. This group consistently contains around one-third of the survey population during the study period. It may suggest a strong and persistent alternative opinion exist among the survey participants.
2. Around 90% show confidence in their requirement as well as testing process (i.e. choose either c or d in SQ2 or SQ3) throughout the study period, except in 2011. In 2011, 78% show confidence in requirement process which is still much higher than the general population.
3. Compared to the general population a higher percentage of the people in this group implies frequent interaction with the customer (chose e or f in SQ8) over the study period. 2011 is an exception where it is lower than the general population.

1 For the longitudinal study, we slightly modified the survey data, like in Question 1's answer if someone put a different process model name which is outside of the first three options then we replaced the answer with "Other".

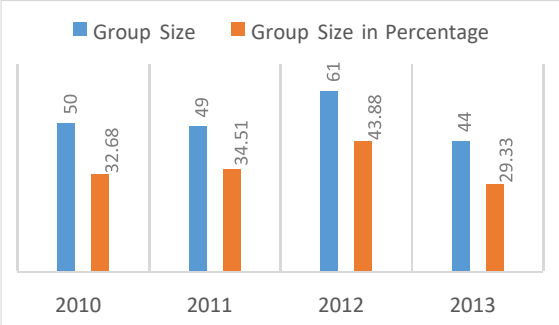

Fig. 6 Size of Group 1 from 2010 to 2013

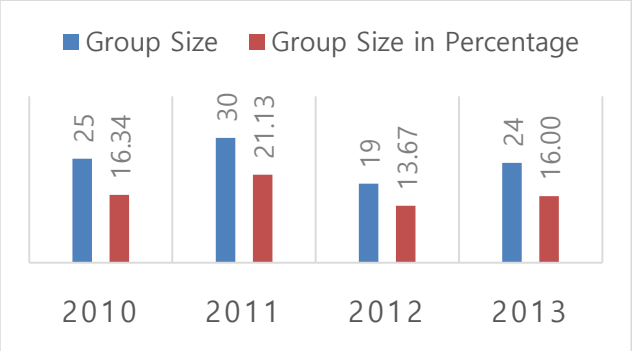

Fig. 7 Size of Group 3 from 2010 to 2013

4.2 Group 2

Members are highly confident regarding their requirements as well as the testing process in this tiny group. During the study period, we found this group in 2010 and 2012 only. In Table 3 we present the data distribution of this group.
Some noticeable characteristics of this group:

1. They are highly experienced compare to the general population.
2. As they suggest their professional competence level is very high; eight out of twelve members (67%) chose d in SQ7, in the general population it is 33%.[1]

4.3 Group 3

They are less confident regarding their requirement as well as their testing process. In Table 4 we present the data distribution of this group.
Some noticeable characteristics of this group:

1. Across the years 80% to 100% of the population suggests less confidence in their testing process (combined *a* and *b* in SQ3) which is only around 30% in the general population.
2. Regarding requirements, they also show lesser confidence compare to overall population - more than 80% compared to 50-60% in general population across the years. Probably the people of this group are less involved in their testing process as they have a lower mean compared to the general population in SQ5.

[1] This question was removed after 2010 survey.

Table 1: Data distribution in general survey population- 2010 to 2013

SQ	Answers	2010	2011	2012	2013	2010(%)	2011(%)	2012(%)	2013(%)
	Traditional	45	38	31	31	29.41	26.76	22.3	20.67
	Agile	21	25	26	27	13.73	17.61	18.71	18
	Blend	79	71	76	86	51.63	50	54.68	57.33
SQ1	Others	7	8	5	5	4.58	5.63	3.6	3.33
	Very Low	16	26	20	19	10.46	18.31	14.39	12.67
	Low	69	60	54	72	45.1	42.25	38.85	48
	High	54	45	56	51	35.29	31.69	40.29	34
SQ2	Very High	12	11	8	7	7.84	7.75	5.76	4.67
	Very Low	4	7	4	10	2.61	4.93	2.88	6.67
	Low	40	37	34	30	26.14	26.06	24.46	20
	High	85	65	75	83	55.56	45.77	53.96	55.33
SQ3	Very High	23	33	25	26	15.03	23.24	17.99	17.33
	<20%	92	77	73	74	60.13	54.23	52.52	49.33
	>30%	4	4	5	6	2.61	2.82	3.6	4
SQ4	Mean	15.26	16	16.664	16.788				
	<20%	47	33	39	33	30.72	23.24	28.06	22
	>30%	18	20	13	17	11.76	14.08	9.35	11.33
SQ5	Mean	22.37	23.853	22.029	23.377				
	<1 Year	9	8	13	11	5.88	5.63	9.35	7.33
	1-3 Years	27	24	14	14	17.65	16.9	10.07	9.33
	3+ Years	31	29	22	25	20.26	20.42	15.83	16.67
	5+ Years	35	29	40	40	22.88	20.42	28.78	26.67
	10+ Years	51	52	50	59	33.33	36.62	35.97	39.33
SQ6	No Answer	0	0	0	1	0	0	0	0.67
	Very Low	0				0			
	Low	6				3.92			
	High	87				56.86			
	Very High	58				37.91			
SQ7	No Answer	2				1.31			
	Never	11	13	12	5	7.19	9.15	8.63	3.33
	In the beginning	8	9	9	11	5.23	6.34	6.47	7.33
	At the end	6	2	9	7	3.92	1.41	6.47	4.67
	At the beginning and the end	32	14	17	25	20.92	9.86	12.23	16.67
	Continuous and on several occasions throughout the project	82	88	78	85	53.59	61.97	56.12	56.67
	Daily throughout the project	12	14	13	16	7.84	9.86	9.35	10.67
SQ8	No Answer	2	2	1	1	1.31	1.41	0.72	0.67
	Total Population	153	142	139	150				

Table 2: Data distribution in Group 1 - 2010 to 2013

SQ	Answers	2010	2011	2012	2013	2010(%)	2011(%)	2012(%)	2013(%)
SQ1	Traditional	11	23	11	8	22	46.94	18.03	18.18
	Agile	7	8	10	1	14	16.33	16.39	2.27
	Blend	31	18	38	34	62	36.73	62.3	77.27
	Others	1	0	2	1	2	0	3.28	2.27
SQ2	Very Low	0	3	5	5	0	6.12	8.2	11.36
	Low	3	8	0	0	6	16.33	0	0
	High	46	38	56	37	92	77.55	91.8	84.09
	Very High	1	0	0	2	2	0	0	4.55
SQ3	Very Low	1	0	0	0	2	0	0	0
	Low	0	4	8	0	0	8.16	13.11	0
	High	36	42	41	35	72	85.71	67.21	79.55
	Very High	13	3	12	9	26	6.12	19.67	20.45
SQ8	Never	3	6	4	0	6	12.24	6.56	0
	In the beginning	2	4	4	2	4	8.16	6.56	4.55
	At the end	1	1	1	1	2	2.04	1.64	2.27
	At the beginning and the end	6	6	10	4	12	12.24	16.39	9.09
	Continuous and on several occasions throughout the project	32	25	34	26	64	51.02	55.74	59.09
	Daily throughout the project	6	6	8	11	12	12.24	13.11	25
	Not Answered	0	1	0	0	0	2.04	0	0
	Group Size	50	49	61	44				

4.4 Group 4

This group is similar to group 3 but smaller in size. All the members follow a traditional development process. This group only existed in 2010 and 2013. Some noticeable characteristics of this group:

1. All members chose *a* in SQ1 which suggests they follow the traditional development process.
2. Across the years 80% to 100% of the population suggests less confidence in their testing process
3. Regarding requirements, they also show lesser confidence compare to overall population - more than 75% compared to 50-60.
4. cross the years 60% of the members do not frequently interact with the customer. Compare to the general population this is very high (in the general population around 35% chose *a* to *d* in SQ8).

4.5 Group 5

This group is similar to group 1, but modest in size and contains people who follow the same development process. Some noticeable characteristics of this group:

1. More than 95% of the members chose the same type of development process in SQ1. Initially, it was the traditional approach which moved to Blend in 2012 and in 2013 it became Agile. Probably suggests a transitional group.

2. Across the years the overwhelming majority of the population suggests good confidence both on their requirement as well as their testing process.
3. They are more involved in the testing process compared to the general population. For SQ5 they have a higher mean (29% vs 22%).

It is reasonable to suggest that we may find more distinctive properties if we consider all the omitted questions. Nevertheless, in this study our main objective is to show that the persistent opinion difference exists in longitudinal studies and clustering approach can identify those alternative opinions.

Table 3: Data distribution in Group 2 - 2010 & 2012

SQ	Answers	2010	2012	2010 (%)	2012 (%)
SQ1	Traditional	3	1	25	10
	Agile	4	5	33.3	50
	Blend	3	3	25	30
	Others	2	1	16.7	10
SQ2	Very Low	2	1	16.7	10
	Low	0	0	0	0
	High	0	0	0	0
	Very High	9	8	75	80
SQ3	Very Low	2	0	16.7	0
	Low	0	0	0	0
	High	0	0	0	0
	Very High	9	9	75	90
SQ6	<1 Year	0	0	0	0
	1-3 Years	1	1	8.33	10
	3+ Years	3	0	25	0
	5+ Years	1	2	8.33	20
	10+ Years	7	7	58.3	70
SQ7	Very Low	0		0	
	Low	0		0	
	High	3		25	
	Very High	8		66.7	
	Group Size	12	10		

5. Discussion

Anderberg indicated that it is very hard to comprehend possible partitioning from a dataset by human ability alone. He gave an example that even to group 25 observations into 5 groups can be huge (exactly 2,436,684,974,110,751) [20].

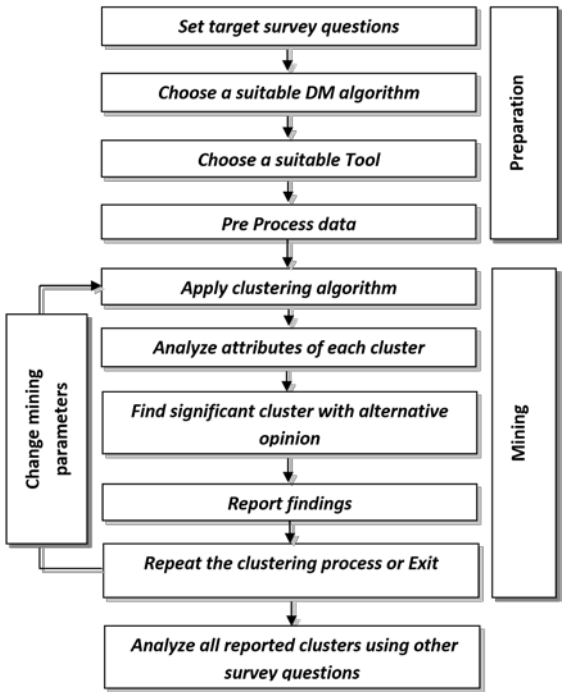

Fig. 8 Process Diagram- Longitudinal Study

So, it is very difficult to thoroughly partition the population manually and investigate their characteristics for even a small survey. It can be even more complex for a longitudinal study due to additional yearly data. Alternatively, similar problems in other domains can be solved by the use of clustering.

In this study, the survey populations are systematically partitioned using clustering techniques. Then the significant groups that show alternative opinions are separated and analyzed. To analyze longitudinal study we locate those groups in different years and compare them to understand their changes over time. Figure 8 shows the process flow graph which discusses some impeding important factors of the process.

Before we start the mining process, the data has to be prepared. The empty records have to be removed or to be loaded with some appropriate data in order to distinguish them from others because some clustering algorithms are not fit enough in handling the empty data field. In this study very few participants refrain from answering the questions we have chosen. In a longitudinal study, some changes in

Table 4: Data distribution in Group 3 - 2010 to 2013

SQ	Answers	2010	2011	2012	2013	2010(%)	2011(%)	2012(%)	2013(%)
	Traditional	0	5	3	5	0	16.67	15.79	20.83
	Agile	5	8	1	0	20	26.67	5.26	0
	Blend	19	17	13	17	76	56.67	68.42	70.83
SQ1	Others	1	0	2	2	4	0	10.53	8.33
	Very Low	3	7	0	3	12	23.33	0	12.5
	Low	15	19	19	19	60	63.33	100	79.17
	High	6	3	0	2	24	10	0	8.33
SQ2	Very High	1	1	0	0	4	3.33	0	0
	Very Low	0	0	0	4	0	0	0	16.67
	Low	25	30	16	20	100	100	84.21	83.33
	High	0	0	0	0	0	0	0	0
SQ3	Very High	0	0	3	0	0	0	15.79	0
	<20%	14	10	9	7	56	33.33	47.37	29.17
	>30%	0	4	2	2	0	2.82	1.44	1.33
SQ5	Mean	16	21.28	18.95	20.83				
	Group Size	25	30	19	24				

questions set are expected which may impact the analysis process. In our case, we focus on a common set of questions across the years.

Occasionally some participants provide nonstandard information based on their perception. In this study, for the survey conducted in the year 2010, 4 participants produced process names which were not on the list. They gave the answer under "Other" category and gave their own answer. Initially, we did not modify the original data. But in some other scenario, if the number of nonstandard input is higher, it might give a better clustering when we revise them under a common category. After identifying interesting groups, we modify those exception data under a common category like "Other" in SQ1 and revised the clustering analysis.

6. Conclusion and Future work

It is not common to apply data mining in opinion based surveys in software engineering. Lack of data due to the small number of participants may discourage the empirical researcher community to use DM as an analysis tool. Usually, this data is analyzed using descriptive statistical tool which produces overall conclusions, so minority opinions lost their voice. On the other hand, those alternative opinions suggested by smaller groups may lead to future success or give a warning against failure.

Since finding the diversity among the different groups is challenging in traditional methods especially in longitudinal studies, we suggest to use a data mining approach. In our study, we show that diversity can be revealed and tracked over a long period of time with the help of some common clustering tools and techniques.

We applied the suggested approach on a sample LS survey which reveals groups with consistent diverse opinions over a long period of time. We analyze those distinct groups and observe their changes in terms of size and characteristics over time. In the future, the existing survey design strategies will be examined from a mining point of view that can produce some recommendations for collecting sound and meaningful datasets for clustering. It may possible to use clustering to deconstruct the overall conclusion of many contemporary surveys.

Acknowledgments

Thanks to QTEMA, Professor Robert Feldt and Professor Tony Gorschek of Blekinge Institute of Technology for sharing the data and providing feedback. This research was supported by Deanship of Scientific Research, Project No. 1155-coc-2016-1-12-S, Qassim University, Saudi Arabia.

References

[1] T. Xie, J. Pei, and A. Hassan, "Mining software engineering data," in Software Engineering Companion, 2007. ICSE 2007 Companion. 29th International Conference on, May 2007, pp. 172–173.

[2] M. M. Hassan and M. Blom, "Applying clustering to analyze opinion diversity," in Proceedings of the 19th International Conference on Evaluation and Assessment in Software

Engineering, ser. EASE '15. New York, NY, USA: ACM, 2015, pp. 8:1–8:10. [Online]. Available: http://doi.acm.org/10.1145/2745802.2745809

[3] S. L. Pfleeger and B. A. Kitchenham, "Principles of survey research: Part 1: Turning lemons into lemonade," SIGSOFT Softw. Eng. Notes, vol. 26, no. 6, pp. 16–18, Nov. 2001. [Online]. Available: http://doi.acm.org/10.1145/505532.505535

[4] B. Kitchenham and S. L. Pfleeger, "Principles of survey research part 6: Data analysis," SIGSOFT Softw. Eng. Notes, vol. 28, no. 2, pp. 24–27, Mar. 2003. [Online]. Available: http://doi.acm.org/10.1145/638750.638758

[5] ——, "Principles of survey research: Part 5: Populations and samples," SIGSOFT Softw. Eng. Notes, vol. 27, no. 5, pp. 17–20, Sep. 2002. [Online]. Available: http://doi.acm.org/10.1145/571681.571686

[6] T. Hall, Longitudinal Studies in Evidence-Based Software Engineering. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 41–41. [Online]. Available: https://doi.org/10.1007/978-3-540-71301-2 14

[7] C. X. Ling and C. Li, "Data mining for direct marketing: Problems and solutions." in KDD, vol. 98, 1998, pp. 73–79.

[8] M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger, "Pulse: Mining customer opinions from free text," in Advances in Intelligent Data Analysis VI. Springer, 2005, pp. 121–132.

[9] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima, "Mining product reputations on the web," in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2002, pp. 341–349.

[10] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in Mining text data. Springer, 2012, pp. 415–463.

[11] B. A. Kitchenham, S. L. Pfleeger, L. M. Pickard, P. W. Jones, D. C. Hoaglin, K. El Emam, and J. Rosenberg, "Preliminary guidelines for empirical research in software engineering," Software Engineering, IEEE Transactions on, vol. 28, no. 8, pp. 721–734, 2002.

[12] J. Moses, "Benchmarking quality measurement," Software Quality Journal, vol. 15, no. 4, pp. 449–462, 2007.

[13] J. Moses and M. Farrow, "Tests for consistent measurement of external subjective software quality attributes," Empirical Software Engineering, vol. 13, no. 3, pp. 261–287, 2008.

[14] J. Moses, "Should we try to measure software quality attributes directly?" Software Quality Journal, vol. 17, no. 2, pp. 203–213, 2009.

[15] T. Gorschek, E. Tempero, and L. Angelis, "On the use of software design models in software development practice: An empirical investigation," Journal of Systems and Software, vol. 95, pp. 176–193, 2014.

[16] D. J. Hand, "Data mining: statistics and more?" The American Statistician, vol. 52, no. 2, pp. 112–118, 1998.

[17] C. Aasheim, J. Shropshire, L. Li, and C. Kadlec, "Knowledge and skill requirements for entry-level it workers: A longitudinal study," vol. 23, pp. 193–204, 2012.

[18] M. J. Shaw, C. Subramaniam, G. W. Tan, and M. E. Welge, "Knowledge management and data mining for marketing," Decision support systems, vol. 31, no. 1, pp. 127–137, 2001.

[19] QTEMA. (2014) Qtema official homepage. [Online]. Available: http://www.qtema.se

[20] M. R. Anderberg, "Cluster analysis for applications," DTIC Document, Tech. Rep., 1973.