An Empirical Approach for Validation of Inter-Rater Reliability of Identified Candidate Aspects

J.S. Aravindan[1*], K. Vivekanandan[2]

[1]Department of Computer Science, Bharathiar University, Coimbatore, India
[2] BSMED, Bharathiar University, Coimbatore, India

[*]*Corresponding Author: soonuaravindan@gmail.com, Tel.: +91-98942-90740*

Abstract— The development of technologies for observing the relevancy of identified candidate aspects of the products and validating the reliability of filed experts is a challenging process. If the identified candidate aspects is taken into consideration for decision making process, the technology has to meet certain feasibility criteria like acceptance of aspects by product experts (Judges). Questionnaires with extracted aspects were distributed to Judges and the candidate aspects are identified. This paper describes the content validity evaluation of identified candidate aspects. The validity of the instrument was evaluated by two field experts who were asked to score each aspect on a 4-point Likert scale (1. Not Relevant, 2. Item need some revision, 3. Relevant but need minor revision, 4. Very Relevant) on relevance of 5 product (Apex AD2600 Progressive-scan DVD player, Canon G3, Creative Labs Nomad Jukebox Zen Xtra 40GB, Nikon Coolpix 4300 and Nokia 6610) datasets and the Kappa Score is calculated.

Keywords—Kappa Score, Aspects, Relevance, Reliability

I. INTRODUCTION

With the advent of Web 2.0, posting online reviews of products has become increasingly popular and effective way for the people to share their opinions and sentiments towards products and services with other online users. E-Commerce websites providing facilities for people to post and comment their review has become a very common practice [1,2].

Those online reviews presents rich and influencing information of products and if it is properly utilized, can provide vendors to facilitate their business to higher levels.

In this paper, Aspects are extracted using Frequency based approach where 1% of the frequent nouns is taken for the study. Extracted aspects for the 5 datasets are presented in the table 2, 4, 6, 8, 10 respectively.

Experiments were conducted on the real world consumer's product review datasets taken from [3]. To evaluate the performance of validation technique, the author examined the inter-rater reliability of aspects using kappa score.

The rest of the paper is organized as follows. Section 2 presents literature survey of previous research related to inter-rater reliability. In Section 3, proposed algorithm for calculating kappa score is presented. It is followed by Section 4 in which experimental analysis and results on

datasets are given. Section 5 summarize the contribution made in this paper.

II. RELATED WORK

In this section, related work of validating inter-rater reliability of candidate aspect is presented.

Cohen, J. [4] says kappa coefficient is the proportion of agreement corrected for chance between two judges assigning cases to a set of k categories, which is offered as a measure of reliability.

Keyton, et al, [5] defined inter-rater reliability as an extent or degree to which the method followed for collecting information is being collected in a consistent manner. That is, the information collecting process and the procedures being used is good enough that the same results can be obtained repeatedly and not be left to chance, either.

Krippendorf, [6] says statistical measures are being used to validate inter-rater reliability in order to provide a logistical proof that the similar answers collected are more than simple chance.

III. METHODOLOGY

In this paper, 2 Judges were asked to rate the extracted aspects of 5 products as

1 = not relevant
2 = item need some revision

3 = relevant but need minor revision

4 = very relevant

and the results were presented in the tables 2 - 11.

Table 1. Algorithm for Calculating Kappa Score

For each Dataset
Create a Contingency Table for responses given by 2 different Judges
Calculate Row and Column totals for the items
Calculate the sum of Row totals and Column totals
Add all of the agreement cells (diagonal) from the contingency table together
Calculate the Expected Frequency for each agreement cells
For i in 1 to n number of rows
Row(i) Total * Column (i) Total / Sum of Row and Column Totals
Add all the Expected Frequencies together
Compute Kappa
Agreement Cell Total – Expected Frequencies Total / Total Number of items –
Expected Frequencies Total
If Kappa Score is > 0.7
    Then Consider it as Satisfactory
Else
    Reject the Score

IV. RESULTS AND DISCUSSION

Table 2. Candidate Aspects of Apex AD2600 Progressive-scan DVD player Dataset

	Apex	Judge 1	Judge 2
1	amazon	1	1
2	apex	4	3
3	brand	1	1
4	button	3	3
5	buttons	3	3
6	christmas	1	1
7	company	1	1
8	contact	1	1
9	control	3	4
10	customer	1	1
11	day	1	1
12	days	1	1
13	disc	4	4
14	discs	4	4
15	disk	4	3
16	display	3	3
17	dvd	4	4

18	dvds	4	4
19	everything	1	1
20	feature	1	1
21	features	1	1
22	fine	1	1
23	formats	1	1
24	gift	1	1
25	guess	1	1
26	hours	1	1
27	line	1	1
28	machine	1	1
29	menu	3	4
30	model	1	1
31	money	1	1
32	month	1	1
33	months	1	1
34	movies	2	1
35	number	1	1
36	output	1	1
37	picture	3	3
38	play	4	3
39	player	3	3
40	players	3	3
41	plays	1	2
42	price	1	1
43	problem	1	2
44	problems	1	1
45	product	1	2
46	purchase	1	1
47	quality	1	2
48	return	1	1
49	reviews	1	2
50	screen	3	3
51	service	1	1
52	something	1	1
53	sound	3	4
54	support	1	2
55	thing	1	1
56	time	1	1
57	unit	2	1
58	use	1	1
59	vcd	3	3

60	video	3	3
61	watch	1	1
62	way	2	1
63	week	1	1
64	weeks	2	1
65	work	1	1
66	years	1	2

Table 3. Contingency Table of Apex AD2600 Progressive-scan DVD player Dataset

		Judge 1				Total
		1	**2**	**3**	**4**	
Judge 2	**1**	**36**	4	0	0	40
	2	7	**0**	0	0	7
	3	0	0	**9**	3	12
	4	0	0	3	**4**	7
Total		43	4	12	7	**66**

Table 4. Candidate Aspects of Canon G3 Dataset

	Canon	**Judge 1**	**Judge 2**
1	auto	1	1
2	battery	4	3
3	buy	1	1
4	camera	4	3
5	cameras	4	3
6	canon	1	1
7	card	3	3
8	control	1	1
9	exposure	4	3
10	features	1	1
11	film	3	1
12	flash	3	3
13	focus	4	4
14	hands	1	1
15	image	3	2
16	images	3	2
17	lcd	3	3
18	lens	4	4
19	life	1	1
20	lot	1	1
21	love	1	1
22	megapixel	4	4

23	metz	1	2
24	mode	1	2
25	moment	2	1
26	nikon	3	1
27	olympus	2	1
28	options	1	1
29	photography	3	2
30	photos	3	3
31	picture	3	3
32	pictures	3	3
33	point	1	1
34	powershot	3	4
35	price	1	1
36	problem	2	1
37	quality	2	1
38	range	1	2
39	research	1	1
40	resolution	4	3
41	results	1	1
42	review	1	2
43	reviews	1	1
44	screen	4	3
45	settings	2	1
46	shoot	2	1
47	shots	2	1
48	shutter	3	3
49	slr	4	3
50	software	1	1
51	something	1	1
52	thing	1	1
53	time	1	1
54	use	1	1
55	view	1	1
56	viewfinder	3	4
57	way	1	1
58	zoom	3	3

Table 5. Contingency Table of Canon G3 Dataset

		Judge 1				Total
		1	**2**	**3**	**4**	
Judge 2	**1**	**22**	7	2	0	31
	2	4	**0**	3	0	7

3	0	0	**8**	7	15
4	0	0	2	**3**	5
Total	26	7	15	10	**58**

Table 6. Candidate Aspects of Creative Labs Nomad Jukebox Zen Xtra 40GB Dataset

	Creative	**Judge 1**	**Judge 2**
1	album	2	1
2	apple	1	1
3	archos	2	1
4	artist	1	1
5	battery	4	4
6	bit	1	2
7	button	3	4
8	buttons	4	3
9	buy	1	1
10	capacity	3	4
11	car	1	2
12	case	3	3
13	cds	4	3
14	collection	1	2
15	computer	2	1
16	cons	1	1
17	controls	4	3
18	data	1	1
19	day	1	1
20	device	1	2
21	disk	3	4
22	display	4	3
23	drive	4	3
24	eax	1	1
25	everything	1	1
26	explorer	1	2
27	fact	1	1
28	feature	1	1
29	features	1	2
30	file	1	1
31	files	2	1
32	headphones	4	4
33	hours	2	1
34	interface	2	1
35	ipod	3	4
36	jack	1	1
37	jukebox	3	4
38	life	1	1
39	line	1	1
40	look	1	1
41	lot	1	1
42	mediasource	2	1
43	memory	4	4
44	minutes	1	2
45	month	1	1
46	music	1	2
47	name	1	1
48	navigation	3	4
49	need	1	1
50	nomad	2	1
51	people	2	1
52	play	1	1
53	player	3	3
54	players	4	3
55	playlists	1	2
56	pod	1	1
57	price	3	4
58	problem	1	1
59	problems	1	1
60	product	1	1
61	program	1	1
62	pros	1	1
63	purchase	1	1
64	quality	3	3
65	replacement	1	1
66	review	1	1
67	reviewers	1	1
68	reviews	1	1
69	screen	4	3
70	scroll	1	1
71	size	1	1
72	software	4	1
73	something	1	1
74	song	1	1
75	songs	1	1
76	sound	1	1
77	storage	3	4
78	support	1	1

79	tags	1	1
80	thing	1	1
81	things	1	1
82	time	1	1
83	times	1	1
84	title	1	1
85	track	2	1
86	tracks	2	1
87	transfer	1	1
88	unit	1	1
89	usb	3	4
90	use	1	1
91	user	1	1
92	warranty	1	1
93	way	1	1
94	weeks	1	1
95	wheel	1	1
96	windows	1	1
97	work	1	1
98	xtra	1	1
99	zen	1	1

Table 7. Contingency Table of Creative Labs Nomad Jukebox Zen Xtra 40GB Dataset

		Judge 1				Total
		1	**2**	**3**	**4**	
Judge 2	**1**	**56**	11	0	1	68
	2	9	**0**	0	0	9
	3	0	0	**3**	7	10
	4	0	0	9	**3**	12
Total		65	11	12	11	**99**

Table 8. Candidate Aspects of Nikon coolpix 4300 Dataset

	Nikon	**Judge 1**	**Judge 2**
1	auto	1	1
2	autofocus	3	4
3	battery	4	4
4	camera	4	4
5	cameras	4	4
6	canon	1	1
7	cap	1	1
8	card	3	2
9	cards	3	2

10	coolpix	1	1
11	date	1	1
12	experience	1	1
13	features	1	2
14	flash	3	3
15	hands	1	1
16	images	3	3
17	lens	4	4
18	life	2	1
19	lot	1	1
20	love	1	1
21	macro	2	1
22	manual	1	1
23	memory	3	4
24	mode	1	1
25	models	1	1
26	modes	2	1
27	money	1	2
28	months	1	1
29	nikon	1	2
30	photos	4	4
31	pics	4	4
32	picture	4	4
33	pictures	3	4
34	price	3	4
35	problem	1	1
36	quality	2	1
37	reader	1	1
38	resolution	4	3
39	scene	1	1
40	settings	1	1
41	shots	2	1
42	size	3	3
43	slr	3	3
44	software	1	1
45	speed	4	4
46	thing	1	1
47	time	4	4
48	use	1	2
49	way	1	2
50	zoom	3	4

Table 9. Contingency Table of Nikon coolpix 4300 Dataset

		Judge 1				Total
		1	**2**	**3**	**4**	
Judge 2	**1**	**19**	5	0	0	24
	2	5	**0**	2	0	7
	3	0	0	**4**	1	5
	4	0	0	5	**9**	14
Total		24	5	11	10	**50**

Table 10. Candidate Aspects of Nokia Dataset

	Nokia	Judge 1	Judge 2
1	access	1	1
2	battery	4	4
3	cable	3	3
4	camera	4	3
5	cell	4	4
6	color	3	4
7	customer	1	1
8	days	1	1
9	email	2	1
10	feature	3	4
11	features	3	4
12	function	1	1
13	games	2	3
14	gsm	1	3
15	hands	1	1
16	headset	3	4
17	hours	3	1
18	internet	4	1
19	life	1	1
20	lot	4	1
21	menu	1	1
22	motorola	2	1
23	music	1	1
24	name	1	1
25	nokia	2	4
26	number	1	1
27	option	1	1
28	options	1	1
29	people	1	1
30	phone	3	4
31	phones	1	1
32	plan	1	1
33	pocket	1	1
34	problem	1	1
35	problems	1	1
36	quality	1	1
37	radio	3	1
38	reception	1	1
39	ringtones	2	1
40	screen	4	4
41	service	2	1
42	signal	4	4
43	size	4	4
44	speakerphone	4	4
45	thing	1	1
46	things	1	1
47	time	1	1
48	tmobile	1	1
49	tones	2	2
50	tzones	1	1
51	use	1	1
52	voice	3	3
53	volume	3	4
54	way	1	1
55	work	1	1
56	years	1	1

Table 11. Contingency Table of Nokia Dataset

		Judge 1				Total
		1	**2**	**3**	**4**	
Judge 2	**1**	**29**	4	2	2	37
	2	0	**1**	0	0	1
	3	1	1	**2**	1	5
	4	0	1	6	**6**	13
Total		30	7	10	9	**56**

Table 12. Kappa Score of All Datasets

Dataset	# Extracted Aspects	Kappa Score
Apex AD2600 Progressive-scan DVD player	66	**0.891**
Canon G3	58	**0.792**
Creative Labs Nomad Jukebox Zen Xtra 40GB	99	**0.801**
Nikon coolpix 4300	50	**0.886**
Nokia 6610	56	**0.729**

Since all the kappa score value of Table 12 is greater than 0.7 the inter-rater reliability rate is found to be satisfactory [4] for all the 5 datasets.

V. CONCLUSION

In this paper, we proposed an algorithm for calculating the Kappa score for the 5 product review datasets to validate the inter-rater reliability of the two field experts (Judges). The Kappa Score value for all the 5 datasets is found to be greater than 0.7 and the result is satisfactory. Hence, the identified candidate aspects of the 5 product review datasets can be taken into consideration for further opinion extraction process. Understanding inter-rater reliability is important for researchers because they should realize whether the aspects they extract for their studies are suitable for the products or not.

REFERENCES

[1] M. Alwan, B. Turner, S. Kell, K. Penberthy, W. Cohn, R. Felder, "*Development of Survey Instruments to Guide the Design of Health Status Monitoring Systems for the Elderly: Content Validity Evaluation*", In the Proceedings of the 2006 IEEE International Conference on Information & Communication Technologies (ICTTA 2006), Syria, pp.793-797, 2006.

[2] V. Bhambri, "*Data Mining as a Solution for Data Management in Banking Sector*", International Journal of Computer Sciences and Engineering, Vol.1, Issue.1, pp.20-25, 2013.

[3] M. Hu, B. Liu, "*Mining and summarizing customer reviews*", In the Proceedings of 2004 ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 2004), New York, pp.168-177, 2004.

[4] J. Cohen, "*A Coefficient of Agreement for Nominal Scales*" Educational and Psychological Measurement, Vol.20, Issue.1, pp.37-46, 1960.

[5] J. Keyton, T. King, NM. Mabachi, J. Manning, LL. Leonard, D. Schill, "*Content analysis procedure book*", Lawrence KS: University of Kansas, Kansas, pp.22-37, 2004.

[6] M. Shelley, K. Krippendorff, "*Content Analysis: An Introduction to its Methodology*", Journal of the American Statistical Association, Vol.79, Issue.385, pp.240-144, 1984.

Authors Profile

Mr. J. Soonu Aravindan pursed Bachelor of Science from P.S.G College of Technology, Coimbatore in 2007 and Master of Computer Applications from College of Engineering, Guindy in year 2010. He is currently pursuing Ph.D. in Department of Computer Sciences, Bharathiar University. He is a member of Computer Society of India since 2013. His main research work focuses on Data Mining, Data Analytics and Opinion Mining.

Dr. K. Vivekanandan, received the Ph.D. degrees in Computer Science and Management from Bharathiar University and M.Sc. degree in Applied Mathematics from P.S.G. College of Technology. His Area of Specialization includes Management Information System, Research Methodology, E Marketing, Data Mining, Web Mining and Opinion Mining. He has 25 Years of Teaching Experience and 17 Years of Research Experience. He has published more than 80 research papers in national and international journals. Currently he is the Director of School of Management, Bharathiar University.