ISSN No. 0976-5697

Volume 8, No. 5, May-June 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

An Empirical Approach for PruningAspects inProduct Review Datasets using Content Validity for Identifying Relevant Candidate Aspects

J. Soonu Aravindan
Research Scholar (Ph.D.)
Bharathiar University
Coimbatore, India

Dr. K. Vivekanandan
Professor
Bharathiar University
Coimbatore, India

Abstract: Due to the large amount of opinions available on the websites, consumers are often overwhelmed with information and find it extremely difficult to use the available information to make a decision about the products they purchase.Opinions of these consumers are classified as expressing Positive, Negative or Neutral sentiments about the aspects of the products. From the reviews of the consumers, Candidate Aspects are extracted using feature extraction methods. Later, Aspect Pruning is performed in order to remove the aspects which are not relevant to the product. Finally, theIdentified Candidate aspects are validated by the experts using Content Validity. This paper validates the candidate aspects 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.

Keywords: Opinion, Content Validity, Aspect Pruning

1. INTRODUCTION

A decade ago, when an individual needed to make a decision, He/she typically asked for opinions from friends, neighbours and families. Similarly, when an organization wanted to find the opinions about its products and services, it conducted opinion polls, surveys, and focus groups.

In the last few years, volumes of opinionated text have grown rapidly and are also publically available. Social media plays a major role by allowing people to share and express their opinion on products, events, topics, individuals, and organizations in the form of comments, reviews, blogs, tweets, status updates, etc. instantly. Therefore, it is quite obvious that people always prefer to hear others opinion before making a decision.

Aspect extraction is the major task in aspect based opinion mining. Aspects are Features (e.g., Features of Camera - Battery, Lens, Viewfinder, Weight, Microphone) of the products. In the recent years, majority of aspect extraction methods have been proposed in product domain. These methods used different techniques to extract aspects from the product reviews. These methods are classified into four main categories [1].

- Rule Based Methods
- Seed Based Methods
- Sequence Model Based Methods
- Topic Model Based Methods

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 listed in the table 1-5.

Experiments were conducted on the real world consumer'sproduct review datasets taken from [2]. To evaluate the performance of validation technique, the author examined the number of aspects obtained after pruning using content validity ratio.

The rest of the paper is organized as follows. Section 2 presents an overview of previous research related to content validation. In Section 3, we present our proposed validationmethod for aspect pruning. 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.

2. RELATED WORKS

In this section, related work of validating candidate aspect is presented.

Rubio, et al [3] demonstrated how to conduct content validity study and how to calculate content validity index, factorial validity index and an interrater reliability index for interpreting these indices.

Yaghmale F [4]developed a content validity scale of 38 items with experts asked to rate each item based on relevance, clarity, simplicity and ambiguity on a four point scale and Content Validity Index (CVI)was determined for assessingexperience with computer usage.

VahidZamanzadeh,et al [5]examined the content validity of the patient-centred communication instrument through a two-step process (development and judgment).First stage performed domain determination, sampling and instrument formation and the second stage performed content validity ratio, content validity index and modified kappa statistics.

Newman, et al [6]defined Content validity as the ability of the selected items to reflect the variables of the construct in the measure.This type of validity addresses the degree to which items of an instrument sufficiently represent the content domain.

3. PROPOSED METHOD

In this paper, 5 Judges were asked to rate the extracted aspects for 5 products as essential or not necessary and the results were presented in the tables 1 – 5.
Lawshe [7] developed a formula termed the content validity ratio:

$$CVR = (n_e - \frac{N}{2})/(\frac{N}{2})$$
$$CVI = (n_e/N)$$

Where CVR = Content Validity Ratio, CVI = Content Validity Index
n_e = number of subject matter expert panellists indicating "essential",
N= total number of subject matter expertpanellists.
This formula yields values which range from +1 to -1
Positive values indicate that at least half the subject matter expert panellistsrated the item as essential.
Only the aspects containing CVI value greater than 0.5 and positive CVR values wereconsidered as final aspects and remaining aspects were removed.

4. EXPERIMENTAL ANALYSIS AND RESULTS

Table 1: Candidate Aspects of Apex AD2600 Progressive-scan DVD player Dataset

Sr.	Extracted Aspects	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	CVR	CVI
1	amazon	0	0	1	0	0	-0.60	0.20
2	apex	1	1	1	1	1	1.00	1.00
3	brand	0	1	0	0	0	-0.60	0.20
4	button	0	1	1	1	1	0.60	0.80
5	buttons	1	0	1	1	1	0.60	0.80
6	christmas	0	0	1	0	0	-0.60	0.20
7	company	0	1	0	0	0	-0.60	0.20
8	contact	0	1	0	0	0	-0.60	0.20
9	control	1	1	1	1	1	1.00	1.00
10	customer	0	0	0	1	0	-0.60	0.20
11	day	0	0	1	0	0	-0.60	0.20
12	days	0	0	0	1	0	-0.60	0.20
13	disc	1	0	1	1	1	0.60	0.80
14	discs	0	0	1	1	1	0.20	0.60
15	disk	1	1	1	0	1	0.60	0.80
16	display	1	1	1	0	1	0.60	0.80
17	dvd	1	1	1	1	1	1.00	1.00
18	dvds	1	0	0	1	1	0.20	0.60
19	everything	0	0	0	0	0	-1.00	0.00
20	feature	0	0	0	0	0	-1.00	0.00
21	features	0	0	0	1	0	-0.60	0.20
22	fine	0	0	0	0	0	-1.00	0.00
23	formats	0	0	1	0	0	-0.60	0.20
24	gift	0	0	0	0	0	-1.00	0.00
25	guess	0	1	0	0	0	-0.60	0.20
26	hours	0	0	0	0	0	-1.00	0.00
27	line	0	1	0	0	0	-0.60	0.20
28	machine	0	0	0	1	0	-0.60	0.20
29	menu	0	1	1	1	1	0.60	0.80
30	model	0	0	0	0	0	-1.00	0.00
31	money	0	0	0	0	0	-1.00	0.00
32	month	1	0	1	0	0	-0.20	0.40
33	months	0	1	0	0	0	-0.60	0.20
34	movies	0	0	0	1	0	-0.60	0.20
35	number	0	0	1	0	0	-0.60	0.20

36	output	1	1	0	0	0	-0.20	0.40
37	picture	1	1	1	0	1	0.60	0.80
38	play	1	1	0	0	1	0.20	0.60
39	player	1	1	0	0	1	0.20	0.60
40	players	1	1	1	1	1	1.00	1.00
41	plays	0	0	0	0	0	-1.00	0.00
42	price	1	1	0	0	0	-0.20	0.40
43	problem	0	0	0	0	0	-1.00	0.00
44	problems	0	0	1	0	0	-0.60	0.20
45	product	0	0	1	0	0	-0.60	0.20
46	purchase	0	0	0	0	0	-1.00	0.00
47	quality	0	0	0	1	0	-0.60	0.20
48	return	0	0	0	0	0	-1.00	0.00
49	reviews	0	1	0	1	0	-0.20	0.40
50	screen	1	0	1	1	1	0.60	0.80
51	service	1	0	0	0	0	-0.60	0.20
52	something	0	0	1	0	0	-0.60	0.20
53	sound	1	0	1	1	1	0.60	0.80
54	support	0	0	0	0	0	-1.00	0.00
55	thing	0	0	1	1	0	-0.20	0.40
56	time	1	0	0	0	0	-0.60	0.20
57	unit	0	1	0	0	0	-0.60	0.20
58	use	0	0	0	1	0	-0.60	0.20
59	vcd	1	1	1	1	0	0.60	0.80
60	video	0	1	1	0	1	0.20	0.60
61	watch	1	0	1	0	0	-0.20	0.40
62	way	0	0	0	0	0	-1.00	0.00
63	week	0	0	1	0	0	-0.60	0.20
64	weeks	0	0	0	0	0	-1.00	0.00
65	work	0	1	0	0	0	-0.60	0.20
66	years	0	0	0	0	0	-1.00	0.00

Table 2. Candidate Aspects of Canon G3 Dataset

Sr.	Extracted Aspects	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	CVR	CVI
1	auto	0	1	0	0	1	-0.20	0.40
2	battery	1	1	1	1	1	1.00	1.00
3	buy	1	1	0	0	0	-0.20	0.40
4	camera	1	1	1	0	1	0.60	0.80
5	cameras	0	1	1	1	1	0.60	0.80
6	canon	1	0	0	0	0	-0.60	0.20
7	card	0	1	1	1	0	0.20	0.60
8	control	1	1	0	0	0	-0.20	0.40
9	exposure	1	1	0	1	1	0.60	0.80
10	features	0	1	0	0	0	-0.60	0.20
11	film	0	1	0	0	0	-0.60	0.20
12	flash	1	0	1	1	1	0.60	0.80
13	focus	1	1	1	1	1	1.00	1.00
14	hands	0	1	0	0	0	-0.60	0.20

15	image	0	0	0	0	0	-1.00	0.00
16	images	0	0	0	0	0	-1.00	0.00
17	lcd	1	1	1	1	1	1.00	1.00
18	lens	1	1	1	1	1	1.00	1.00
19	life	0	0	1	0	0	-0.60	0.20
20	lot	0	0	0	0	1	-0.60	0.20
21	love	0	1	0	0	1	-0.20	0.40
22	megapixel	1	1	1	1	1	1.00	1.00
23	metz	0	0	0	0	0	-1.00	0.00
24	mode	0	0	0	0	0	-1.00	0.00
25	moment	0	0	0	0	0	-1.00	0.00
26	nikon	0	0	0	0	0	-1.00	0.00
27	olympus	0	0	0	0	0	-1.00	0.00
28	options	0	0	0	0	0	-1.00	0.00
29	photography	1	0	0	0	0	-0.60	0.20
30	photos	1	1	1	0	0	0.20	0.60
31	picture	1	0	1	1	1	0.60	0.80
32	pictures	1	1	1	1	1	1.00	1.00
33	point	0	0	0	0	0	-1.00	0.00
34	powershot	0	1	1	1	1	0.60	0.80
35	price	0	0	0	0	0	-1.00	0.00
36	problem	0	1	0	0	0	-0.60	0.20
37	quality	0	1	0	0	0	-0.60	0.20
38	range	1	0	0	0	0	-0.60	0.20
39	research	0	0	0	0	0	-1.00	0.00
40	resolution	1	1	1	1	1	1.00	1.00
41	results	0	0	0	0	0	-1.00	0.00
42	review	0	0	0	0	0	-1.00	0.00
43	reviews	1	0	0	0	0	-0.60	0.20
44	screen	1	1	1	1	1	1.00	1.00
45	settings	0	1	0	0	0	-0.60	0.20
46	shoot	0	1	0	0	0	-0.60	0.20
47	shots	0	1	0	0	0	-0.60	0.20
48	shutter	1	1	0	1	1	0.60	0.80
49	slr	1	1	1	1	1	1.00	1.00
50	software	0	0	0	0	1	-0.60	0.20
51	something	0	1	0	0	0	-0.60	0.20
52	thing	0	1	0	0	0	-0.60	0.20
53	time	1	0	0	0	0	-0.60	0.20
54	use	1	0	0	0	0	-0.60	0.20
55	view	1	0	1	0	0	-0.20	0.40
56	viewfinder	1	1	1	1	1	1.00	1.00
57	way	1	1	0	0	0	-0.20	0.40
58	zoom	1	1	0	1	1	0.60	0.80

Table 3. Candidate Aspects of Creative Labs Nomad Jukebox Zen Xtra 40GBDataset

Sr.	Extracted Aspects	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	CVR	CVI
1	album	1	1	0	0	0	-0.20	0.40
2	apple	1	1	0	0	0	-0.20	0.40
3	archos	1	1	0	0	0	-0.20	0.40
4	artist	1	1	0	0	0	-0.20	0.40
5	battery	1	1	1	1	1	1.00	1.00
6	bit	1	0	0	0	0	-0.60	0.20
7	button	1	1	1	1	1	1.00	1.00
8	buttons	1	1	1	1	1	1.00	1.00
9	buy	0	1	0	0	0	-0.60	0.20
10	capacity	1	1	1	1	1	1.00	1.00
11	car	1	0	0	0	0	-0.60	0.20
12	case	1	0	1	1	0	0.20	0.60
13	cds	1	1	1	0	0	0.20	0.60
14	collection	0	0	0	0	1	-0.60	0.20
15	computer	0	1	1	0	0	-0.20	0.40
16	cons	1	0	0	0	0	-0.60	0.20
17	controls	1	1	1	1	0	0.60	0.80
18	data	0	1	0	0	0	-0.60	0.20
19	day	1	0	0	0	0	-0.60	0.20
20	device	1	0	0	0	0	-0.60	0.20
21	disk	1	1	0	1	1	0.60	0.80
22	display	1	1	1	1	0	0.60	0.80
23	drive	1	1	1	1	0	0.60	0.80
24	eax	1	0	0	0	0	-0.60	0.20
25	everything	1	1	0	0	0	-0.20	0.40
26	explorer	1	0	0	0	0	-0.60	0.20
27	fact	0	0	0	0	1	-0.60	0.20
28	feature	1	0	1	0	0	-0.20	0.40
29	features	1	0	0	0	0	-0.60	0.20
30	file	1	0	0	0	0	-0.60	0.20
31	files	0	1	0	0	0	-0.60	0.20
32	headphones	1	1	1	1	0	0.60	0.80
33	hours	0	1	0	0	1	-0.20	0.40
34	interface	0	1	0	1	0	-0.20	0.40
35	ipod	1	0	1	1	1	0.60	0.80
36	jack	1	0	0	0	0	-0.60	0.20
37	jukebox	1	1	1	0	1	0.60	0.80
38	life	0	1	0	0	0	-0.60	0.20
39	line	1	0	0	0	0	-0.60	0.20
40	look	1	0	0	0	0	-0.60	0.20
41	lot	1	1	0	0	0	-0.20	0.40
42	mediasource	0	1	0	0	0	-0.60	0.20
43	memory	1	1	1	1	1	1.00	1.00
44	minutes	1	0	0	0	0	-0.60	0.20
45	month	1	1	0	0	0	-0.20	0.40
46	music	1	0	0	0	1	-0.20	0.40

47	name	1	1	0	0	0	-0.20	0.40
48	navigation	1	1	1	1	1	1.00	1.00
49	need	1	0	0	0	1	-0.20	0.40
50	nomad	0	1	0	0	0	-0.60	0.20
51	people	0	1	0	0	0	-0.60	0.20
52	play	1	1	0	0	0	-0.20	0.40
53	player	1	1	1	0	0	0.20	0.60
54	players	1	1	1	1	0	0.60	0.80
55	playlists	1	0	0	0	0	-0.60	0.20
56	pod	1	1	0	0	0	-0.20	0.40
57	price	1	1	1	1	1	1.00	1.00
58	problem	1	1	0	0	0	-0.20	0.40
59	problems	1	0	0	0	0	-0.60	0.20
60	product	1	0	0	0	0	-0.60	0.20
61	program	1	1	0	0	0	-0.20	0.40
62	pros	1	0	0	0	0	-0.60	0.20
63	purchase	1	0	0	0	0	-0.60	0.20
64	quality	1	1	0	0	1	0.20	0.60
65	replacement	1	0	0	0	0	-0.60	0.20
66	review	1	0	0	0	0	-0.60	0.20
67	reviewers	0	0	0	0	0	-1.00	0.00
68	reviews	0	0	0	0	0	-1.00	0.00
69	screen	1	1	1	1	1	1.00	1.00
70	scroll	1	1	0	0	0	-0.20	0.40
71	size	1	1	0	0	0	-0.20	0.40
72	software	4	1	0	0	0	-0.60	0.20
73	something	1	1	0	0	0	-0.20	0.40
74	song	1	1	0	0	0	-0.20	0.40
75	songs	1	1	0	0	0	-0.20	0.40
76	sound	1	1	0	0	0	-0.20	0.40
77	storage	1	1	1	1	1	1.00	1.00
78	support	1	0	0	0	0	-0.60	0.20
79	tags	1	1	0	0	0	-0.20	0.40
80	thing	1	1	0	0	0	-0.20	0.40
81	things	1	1	0	0	0	-0.20	0.40
82	time	1	1	0	0	0	-0.20	0.40
83	times	1	1	0	0	0	-0.20	0.40
84	title	1	1	0	0	0	-0.20	0.40
85	track	0	1	0	0	0	-0.60	0.20
86	tracks	0	1	0	0	0	-0.60	0.20
87	transfer	1	1	0	0	0	-0.20	0.40
88	unit	1	1	0	0	0	-0.20	0.40
89	usb	1	1	1	1	1	1.00	1.00
90	use	1	0	0	0	0	-0.60	0.20
91	user	1	0	0	0	0	-0.60	0.20
92	warranty	1	0	0	0	0	-0.60	0.20
93	way	1	0	0	0	0	-0.60	0.20
94	weeks	1	1	0	0	0	-0.20	0.40
95	wheel	1	0	0	0	0	-0.60	0.20

96	windows	0	0	0	0	0	-1.00	0.00
97	work	0	0	0	0	0	-1.00	0.00
98	xtra	1	0	0	0	0	-0.60	0.20
99	zen	0	1	0	0	0	-0.60	0.20

Table 4. Candidate Aspects of Nikon Coolpix 4300 Dataset

Sr.	Extracted Aspects	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	CVR	CVI
1	auto	1	1	0	0	0	-0.20	0.40
2	autofocus	1	0	1	1	1	0.60	0.80
3	battery	1	1	1	1	1	1.00	1.00
4	camera	1	1	1	1	1	1.00	1.00
5	cameras	1	1	1	1	1	1.00	1.00
6	canon	1	0	0	0	0	-0.60	0.20
7	cap	1	0	0	0	0	-0.60	0.20
8	card	0	0	0	0	0	-1.00	0.00
9	cards	0	0	0	0	1	-0.60	0.20
10	coolpix	1	1	0	0	0	-0.20	0.40
11	date	1	0	0	0	0	-0.60	0.20
12	experience	1	0	0	0	0	-0.60	0.20
13	features	1	0	0	0	0	-0.60	0.20
14	flash	1	1	1	1	1	1.00	1.00
15	hands	1	0	0	0	0	-0.60	0.20
16	images	1	1	1	1	1	1.00	1.00
17	lens	1	1	1	1	1	1.00	1.00
18	life	0	1	0	0	0	-0.60	0.20
19	lot	1	1	0	0	0	-0.20	0.40
20	love	1	1	0	0	0	-0.20	0.40
21	macro	0	1	0	0	0	-0.60	0.20
22	manual	1	1	0	0	0	-0.20	0.40
23	memory	1	1	1	1	1	1.00	1.00
24	mode	0	1	0	0	0	-0.60	0.20
25	models	1	0	0	0	0	-0.60	0.20
26	modes	0	1	0	0	0	-0.60	0.20
27	money	1	0	0	0	0	-0.60	0.20
28	months	1	1	0	0	0	-0.20	0.40
29	nikon	1	0	0	0	0	-0.60	0.20
30	photos	1	1	1	1	1	1.00	1.00
31	pics	1	1	1	1	1	1.00	1.00
32	picture	1	1	1	1	1	1.00	1.00
33	pictures	0	1	1	1	1	0.60	0.80
34	price	0	1	1	1	1	0.60	0.80
35	problem	1	1	0	0	0	-0.20	0.40
36	quality	0	1	0	0	0	-0.60	0.20
37	reader	1	1	0	0	0	-0.20	0.40
38	resolution	1	1	1	1	1	1.00	1.00
39	scene	1	1	0	0	0	-0.20	0.40
40	settings	1	1	0	0	0	-0.20	0.40
41	shots	0	1	0	0	0	-0.60	0.20

42	size	1	1	1	1	1	1.00	1.00
43	slr	1	1	0	1	1	0.60	0.80
44	software	1	1	0	0	0	-0.20	0.40
45	speed	1	1	1	1	1	1.00	1.00
46	thing	0	1	0	0	0	-0.60	0.20
47	time	1	1	1	1	1	1.00	1.00
48	use	1	0	0	0	0	-0.60	0.20
49	way	1	0	0	0	1	-0.20	0.40
50	zoom	1	1	1	1	1	1.00	1.00

Table 5: Candidate Aspects of Nokia Dataset

Sr.	Extracted Aspects	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	CVR	CVI
1	access	1	1	0	0	0	-0.20	0.40
2	battery	1	1	1	1	1	1.00	1.00
3	cable	1	0	1	1	1	0.60	0.80
4	camera	1	1	1	1	1	1.00	1.00
5	cell	1	0	1	1	1	0.60	0.80
6	color	1	1	1	1	1	1.00	1.00
7	customer	1	1	0	0	0	-0.20	0.40
8	days	1	0	0	0	0	-0.60	0.20
9	email	0	1	0	0	0	-0.60	0.20
10	feature	0	1	1	1	1	0.60	0.80
11	features	1	0	1	1	1	0.60	0.80
12	function	0	0	0	0	0	-1.00	0.00
13	games	0	0	0	0	0	-1.00	0.00
14	gsm	1	0	0	0	0	-0.60	0.20
15	hands	1	1	0	0	0	-0.20	0.40
16	headset	0	1	1	1	1	0.60	0.80
17	hours	0	1	0	0	0	-0.60	0.20
18	internet	0	0	1	1	0	-0.20	0.40
19	life	1	1	0	0	0	-0.20	0.40
20	lot	0	1	0	0	0	-0.60	0.20
21	menu	1	1	0	0	0	-0.20	0.40
22	motorola	0	1	0	0	0	-0.60	0.20
23	music	1	1	0	0	0	-0.20	0.40
24	name	1	0	0	0	0	-0.60	0.20
25	nokia	1	1	0	0	0	-0.20	0.40
26	number	1	1	0	0	0	-0.20	0.40
27	option	1	1	0	0	0	-0.20	0.40
28	options	1	1	0	0	0	-0.20	0.40
29	people	1	0	0	0	0	-0.60	0.20
30	phone	0	1	1	1	1	0.60	0.80
31	phones	1	1	0	0	0	-0.20	0.40
32	plan	1	1	0	0	0	-0.20	0.40
33	pocket	1	1	0	0	0	-0.20	0.40
34	problem	0	1	0	0	0	-0.60	0.20
35	problems	0	1	0	0	0	-0.60	0.20
36	quality	0	1	0	0	0	-0.60	0.20

37	radio	0	1	0	0	0	-0.60	0.20
38	reception	1	1	0	0	0	-0.20	0.40
39	ringtones	0	1	0	0	0	-0.60	0.20
40	screen	1	1	1	1	1	1.00	1.00
41	service	1	1	0	0	0	-0.20	0.40
42	signal	1	1	1	1	1	1.00	1.00
43	size	1	1	1	1	1	1.00	1.00
44	speakerphone	1	1	1	1	1	1.00	1.00
45	thing	1	1	0	0	0	-0.20	0.40
46	things	1	1	0	0	0	-0.20	0.40
47	time	1	1	0	0	0	-0.20	0.40
48	tmobile	1	1	0	0	0	-0.20	0.40
49	tones	0	0	0	0	0	-1.00	0.00
50	tzones	1	1	0	0	0	-0.20	0.40
51	use	1	1	0	0	0	-0.20	0.40
52	voice	1	0	1	1	1	0.60	0.80
53	volume	1	1	1	1	1	1.00	1.00
54	way	1	0	0	0	0	-0.60	0.20
55	work	1	1	0	0	0	-0.20	0.40
56	years	1	1	0	0	0	-0.20	0.40

Table 6.Relevant Aspects after pruning

Dataset	# Extracted Aspects	# RelevantAspects
Apex AD2600 Progressive-scan DVD player	66	**19**
Canon G3	58	**20**
Creative Labs Nomad Jukebox Zen Xtra 40GB	99	**22**
Nikon coolpix 4300	50	**19**
Nokia 6610	56	**15**

Table 7. Percentage of Irrelevant Aspects

Dataset	% of Irrelevant Aspects
Apex AD2600 Progressive-scan DVD player	**71.21**
Canon G3	**65.52**
Creative Labs Nomad Jukebox Zen Xtra 40GB	**77.78**
Nikon coolpix 4300	**62.00**
Nokia 6610	**73.21**

Table 6 shows the number of aspect terms retrieved after pruning the extracted aspects. Table 7 shows that more than 60 % of the aspect terms extracted are irrelevant aspects which has to be removed for better accuracy for analysing the sentiments of the aspects.

5. CONCLUSION

In this paper, we proposed a validation system that validate relevant candidate aspects from the extracted aspects of consumer product reviews. Evaluation experiments were designed to run on real world datasets taken from product reviews. The proposed system has identified and removed the irrelevant aspects constituting 60% from the 5 datasets which leads to better accuracy in extracting opinions and emotions of the relevant aspects.

REFERENCES

1. Zhang, L., & Liu, B. (2014). Aspect and Entity Extraction for Opinion Mining. Studies in Big Data, 1-40.
2. Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04.
3. Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. Social Work Research, 27(2), 94-104.
4. Yaghmale F. (2003). Content Validity and its estimation, Journal of Medical Education, Vol.3, No.1, 25-27.

5. Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., &Nikanfar, A. (2015). Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. Journal of Caring Sciences, 4(2), 165-178.
6. Newman, I., Lim, J., & Pineda, F. (2013). Content Validity Using a Mixed Methods Approach. Journal of Mixed Methods Research, 7(3), 243-260.
7. LAWSHE, C. H. (1975). A QUANTITATIVE APPROACH TO CONTENT VALIDITY. Personnel Psychology, 28(4), 563-575.