Smart Innovation, Systems and Technologies 51

Suresh Chandra Satapathy
Swagatam Das *Editors*

Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 2

KES International

Springer

Smart Innovation, Systems and Technologies

Volume 51

About this Series

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

More information about this series at http://www.springer.com/series/8767

Suresh Chandra Satapathy · Swagatam Das
Editors

Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 2

Springer

Editors
Suresh Chandra Satapathy
Department of Computer Science
and Engineering
Anil Neerukonda Institute of Technology
and Sciences
Visakhapatnam
India

Swagatam Das
Indian Statistical Institute
Jadavpur University
Kolkata
India

ISSN 2190-3018 ISSN 2190-3026 (electronic)
Smart Innovation, Systems and Technologies
ISBN 978-3-319-30926-2 ISBN 978-3-319-30927-9 (eBook)
DOI 10.1007/978-3-319-30927-9

Library of Congress Control Number: 2016936566

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

This SIST volume contains the papers presented at the ICTIS 2015: International Conference on Information and Communication Technology for Intelligent Systems. The conference was held during November 28–29, 2015, Ahmedabad, India and organized communally by Venus International College of Technology, Association of Computer Machinery, Ahmedabad Chapter and supported by Computer Society of India Division IV—Communication and Division V—Education and Research. It targeted state-of-the-art as well as emerging topics pertaining to ICT and effective strategies for its implementation for engineering and intelligent applications. The objective of this international conference is to provide opportunities for the researchers, academicians, industry persons, and students to interact and exchange ideas, experience and expertise in the current trend and strategies for information and communication technologies. Besides this, participants were enlightened about vast avenues, current and emerging technological developments in the field of ICT in this era and its applications were thoroughly explored and discussed. The conference attracted a large number of high-quality submissions and stimulated the cutting-edge research discussions among many academic pioneering researchers, scientists, industrial engineers, and students from all around the world and provided a forum to researcher. Research submissions in various advanced technology areas were received and after a rigorous peer-review process with the help of program committee members and external reviewer, 119 (Vol-I: 59, Vol-II: 60) papers were accepted with an acceptance ratio of 0.25. The conference featured many distinguished personalities like Dr. Akshai Aggarawal, Hon'ble Vice Chancellor, Gujarat Technological University, Dr. M.N Patel, Hon'ble Vice Chancellor, Gujarat University, Dr. Durgesh Kumar Mishra, Chairman Division IV CSI, and Dr. S.C Satapathy, Chairman, Division V, Computer Society of India, Dr. Bhushan Triverdi, Director, GLS University, and many more. Separate invited talks were organized in industrial and academia tracks during both days. The conference also hosted few tutorials and workshops for the benefit of participants. We are indebted to ACM Ahmedabad Professional Chapter, CSI Division IV, V for their immense support to make this conference possible in

such a grand scale. A total of 12 sessions were organized as a part of *ICTIS 2015* including nine technical, one plenary, one inaugural, and one valedictory session. A total of 93 papers were presented in the nine technical sessions with high discussion insights. The total number of accepted submissions was 119 with a focal point on ICT and intelligent systems. Our sincere thanks to all sponsors, press, print, and electronic media for their excellent coverage of this conference.

November 2015

Suresh Chandra Satapathy
Swagatam Das

Conference Committee

Patrons

Mr. Rishabh Jain, Chairman, VICT, Ahmedabad, India
Dr. Srinivas Padmanabhuni, President, ACM India

General Chairs

Dr. A.K. Chaturvedi, Director, VICT, Ahmedabad, India
Mr. Chandrashekhar Sahasrabudhe, ACM India
Dr. Nilesh Modi, Chairman, ACM Ahmedabad Chapter

Advisory Committee

Dr. Chandana Unnithan, Victoria University, Australia
Dr. Bhushan Trivedi, Professional Member, ACM
Mustafizur Rahman, Endeavour, Agency for Science Technology and Research, Australia
Mr. Bharat Patel, Professional Member, ACM
Dr. Durgesh Kumar Mishra, Chairman, Division IV, CSI
Mr. Hardik Upadhyay, GPERI, India
Dr. Dharm Singh, PON, Namibia
Mr. Ashish Gajjar, Professional Member, ACM
Hoang Pham, Professor and Chairman, Department of ISE, Rutgers University, Piscataway, NJ, USA
Mr. Maitrik Shah, Professional Member, ACM
Ernest Chulantha Kulasekere, Ph.D., University of Moratuwa, Sri Lanka
Ms. Heena Timani, Professional Member, ACM
Prof. S.K. Sharma Director PIE, PAHER, Udaipur, India

Shashidhar Ram Joshi, Ph.D., Institute of Engineering, Pulchowk Campus, Pulchowk, Nepal
Mr. Bhaumik Shah, Professional Member, ACM
Subhadip Basu, Ph.D., Visiting Scientist, The University of Iowa, Iowa City, USA
Prof. Lata Gohil, Professional Member, ACM
Abrar A. Qureshi, Ph.D., University of Virginia's College at Wise, One College Avenue, USA
Dr. Aditya Patel, Professional Member, ACM
Dr. Muneesh Trivedi, Professor, ABES, Ghaziabad, India
Dr. Ashish Rastogi (CSI), IT Department, Higher College of Technology, Muscat-Oman
Dr. Kuntal Patel, Professional Member, ACM
Dr. Aynur Unal, Standford University, USA
Prof. Rutvi Shah, Professional Member, ACM
Dr. Malaya Kumar Nayak. Director IT Buzz Ltd., UK
Dr. Nisarg Pathak, Professional Member, ACM

Organising Committee

Chair
Prof. Jayshree Upadhyay, HOD, CE, VICT, Ahmedbad, India

Secretary
Mr. Mihir Chauhan, VICT, Ahmedabad, India

Members
Prof. Mamta Jain, VICT, Ahmedabad, India
Mr. Zishan Shaikh, VICT, Ahmedabad, India
Mr. Jalpesh Ghumaliya, VICT, Ahmedabad, India
Prof. Nehal Rajput, VICT, Ahmedabad, India
Prof. Murti Patel, VICT, Ahmedabad, India
Prof. Jalpa Shah, VICT, Ahmedabad, India
Ms. Lajja Vyash, VICT, Ahmedabad, India
Ms. Ruchita Joshi, VICT, Ahmedabad, India
Mr. Aman Barot, VICT, Ahmedabad, India
Mr. Maulik Dave, VICT, Ahmedabad, India
Ms. Priyadarshini Barot, VICT, Ahmedabad, India

Technical Program Committee

Chair
Dr. Suresh Chandra Satapathy, Chairman, Division V, CSI

Co-Chair
Prof. Vikrant Bhateja, SRMGPC, Lucknow

Members
Dr. Mukesh Sharma, SFSU, Jaipur
Prof. D.A. Parikh, Head, CE, LDCE, Ahmedabad
Dr. Savita Gandhi, Head, CE, Rolwala, GU, Ahmedabad
Dr. Jyoti Parikh, Associate Professor, CE, GU, Ahmedabad
Ms. Bijal Talati, Head, CE, SVIT, Vasad
Dr. Harshal Arolkar, Member ACM
Dr. Pushpendra Singh, JK Lakshimpath University
Dr. Sanjay M. Shah, GEC, Gandhinagar
Dr. Chirag S. Thaker, GEC, Bhavnagar, Gujarat
Mr. Jeril Kuriakose, Manipal University, Jaipur
Mr. Ajay Chaudhary, IIT Roorkee, India
Mr. R.K. Banyal, RTU, Kota, India
Mr. Amit Joshi, Professional Member, ACM
Dr. Vishal Gour, Bikaner, India
Mr. Vinod Thummar, SITG, Ahmedabad, India
Mr. Nisarg Shah, 3D Blue Print, Ahmedabad, India
Mr. Maulik Patel, Emanant TechMedia, Ahmedabad, India

CSI Apex Committee

Prof. Bipin V. Mehta, President, CSI
Dr. Anirban Basu, Vice-President, CSI
Mr. Sanjay Mohapatra, Hon-Secretary, CSI
Mr. R.K. Vyas, Hon-Treasurer, CSI
Mr. H R Mohan, Immediate Past President, CSI
Dr. Vipin Tyagi, RVP, Region III, CSI

Conference Track Managers

Track#1: Image Processing, Machine Learning and Soft Computing—Dr. Steven Lawrence Fernandes
Track#2: Software and Big Data Engineering—Dr. Kavita Choudhary
Track#3: Network Security, Wireless and Mobile Computing—Dr. Musheer Ahmad

Contents

Part I
ICT Based Security and Privacy Applications

POS Word Class Based Categorization of Gurmukhi Language Stemmed Stop Words

Kaur Jasleen and R. Saini Jatinderkumar

Abstract Literature in Indian language must be classified for its easy retrieval. In Punjabi literature classifier, five different categories: nature, romantic, religious, patriotic and philosophical, are manually populated with 250 poems. These poems are pre-processed through data cleaning, tokenization, bag of word, stop word identification and stemming phases. Due to unavailability of Punjabi stop words in public domain, manual collection of 256 stop words are done from poetry and articles. After stemming, 184 unique stemmed words are identified. Based on part of speech tagging, 184 stop words are categorized into 98 adverbs, 7 conjunctions, 43 verbs, 24 pronouns and 12 miscellaneous words. These unique 184 stemmed words are being released for other language processing algorithm in Punjabi. This paper concentrates on providing better and deeper understanding of Punjabi stop words in lieu of Punjabi grammar and part of speech based word class categorization.

Keywords Adverb · Conjunction · Verb · Pronoun · Part of speech · Punjabi · Stop word

1 Introduction

With the advent of World Wide Web and Unicode encoding, Indian language content is increasing on the web day by day. In today's internet era, people prefer to use their regional language or mother language to communicate their thoughts. So this data must be classified for its easy retrieval and usage. Text Classification is an act of assigning natural language text into predefined categories [1]. India, being a

K. Jasleen (✉)
Uka Tarsadia University, Bardoli, Gujarat, India
e-mail: sidhurukku@yahoo.com

R. Saini Jatinderkumar
Narmada College of Computer Application, Bhaurch, Gujarat, India
e-mail: saini_expert@yahoo.com

S.C. Satapathy and S. Das (eds.), *Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 2*, Smart Innovation, Systems and Technologies 51, DOI 10.1007/978-3-319-30927-9_1

multilingual country, consists of wide number of languages and rich literature. Out of these languages, 22 languages are recognized as regional languages [2]. Punjabi, one of them, is widely spoken language in Punjab (India) as well as in Pakistan [3]. Punjabi Language belongs to Indo-Aryan Language Family. Punjab is known for its rich culture and literature. Poem is one form of literary art. Poetry is always imaginative in nature with a message to its reader [4]. Poetry always has a strong association with feelings, thoughts and ideas. An automatic poetry classification is a text classification problem. Input to classifier is poem in Punjabi language and classifier will assign a category on the basis of its content. This paper is focused on the analysis of stop words from grammatical point of view.

2 Indian Language Based Text Classifier

India is a multilingual country. Many languages are being used in India. Indo-Aryan (consists of Hindi, Gujarati, Bengali, Punjabi, Marathi, Urdu, and Sanskrit) and Dravidian (Telugu, Tamil, Kannada) are major language families spoken in India [5]. Brief survey about the text classification works done in Indian languages is given below.

2.1 Text Classification in Indo-Aryan Language Family

Statistical techniques using Naïve Bayes and Support Vector Machine are used to classify subjective sentences from objective sentences for Urdu language. As Urdu language is morphological rich language, this makes the classification task more difficult. The result of this implementation shows that accuracy, performance of Support Vector Machines is much better than Naïve Bayes classification techniques [6]. For Bangla text classification, n-gram based algorithm is used and to analyze the performance of the classifier. Prothom-Alo news corpus is used. The result show that with increase in value of n from 1 to 3, performance of the text classification also increases, but from value 3 to 4 performance decreases [7]. Sanskrit text documents have been classified using Sanskrit Word net. Semantic based classifier is method is built on lexical chain of linking significant words that are about a particular topic with the help of hypernym relation in Word Net [8]. Very few works in literature are found in field of text classification in Punjabi Language. Domain based text classification is done by Nidhi and Vishal [9]. This classification is done on sports category only. Two new algorithms, Ontology based classification and Hybrid approach are proposed for Punjabi text classification. The experimental results conclude that Ontology based classification (85 %) and Hybrid approach (85 %) provides better results. Sarmah et al. [10] presented an approach for classification of Assamese documents using Assamese WordNet. This approach has accuracy of 90.27 % on Assamese documents.

2.2 Text Classification in Dravidian Language Family

Naïve Bayes classifier has been applied to Telugu news articles to classify 800 documents into four major classes. In this, normalized term frequency-inverse document frequency is used to extract the features from the document. Without any stop word removal and morphological analysis, at the threshold of 0.03, the classifier gives 93 % precision [11]. For morphologically rich Dravidian classical language Tamil, text classification is done using vector space model and artificial neural network. The experimental results show that Artificial Neural network model achieves 93.33 % which is better than the performance of Vector Space Model which yields 90.33 % on Tamil document classification [12]. A new technique called Sentence level classification is done for Kannada language; in this sentences are analyzed to classify the Kannada documents. This Technique extended further to sentiment classification, question answering, text summarization and also for customer reviews in Kannada blogs [13].

3 Pre-processing Steps Involved in Punjabi Poetry Classifier

Before classification, data must be pre processed to remove unwanted words and noise [14]. Pre-processing phase of poetry classifier consists of Data Cleaning, Feature Extraction and Feature Selection. As Punjabi is resource scarce language, there is no publicly available corpus, so manual collection of poetry is done. Initially, Data is collected into 5 different categories: ਕੁਦਰਤ [*kudarata*] 'Nature', ਪ੍ਰੀਤ [*prīta*] 'Romantic', ਧਾਰਮਿਕ [*dhāramika*] 'Religious', ਦੇਸ਼ਭਗਤੀ [*dēśabhagatī*] 'Patriotic' and ਦਾਰਸ਼ਨਿਕ [*dāraśanika*] 'Philosophical'. Initially, these categories are populated with 50 poems in each category. Implementation of various subphases (as discussed below) is done in Visual Basic.Net using Microsoft Visual Studio 2010 as front end and Microsoft Access 2007 at back end using Unicode characters [15].

3.1 Cleaning

Preprocessing step is involved to remove the noise from data so that this noisy data don't penetrate into the next higher levels. It includes special symbol deletion. Symbols like: comma (,), dandi (।), double dandi (॥), sign of interrogation (?) and sign of exclamation (!) are present in poems. In case of Punjabi language, dandi (।) is used in place of full stop. Double dandi (॥) is generally used in ancient Punjabi writings like religious poetry.

3.2 Feature Extraction and Feature Selection

Feature Extraction phase consists of tokenization, unique words and its term frequency calculation, stop word removal and stemming. In tokenization, each poem is tokenized and 'bag of word' model is created. Data structures used for implementation are hash tables, files and arrays. Unique words are identified from poems and its frequency is calculated from tokenized words. After this, stop words are eliminated from unique words. Stop words are most common words occurring in the text which are not significant for classifier. 256 stop words are identified from poetry, news articles and other Punjabi stories. Stemming is way of converting a written text into its root form [16]. Gupta [17] developed different rules for handling stemming for verbs, adverbs and pronouns. These stemming rules are manually applied to 256 identified stop words. After stemming, 184 unique stemmed stop words are identified and presented in Table 1. This table consists of Columns: word in Punjabi (C1), its transliteration in English (C2) and its meaning in English (C3) [18, 19].

Table 1 List of stemmed stop words

S. no.	C1	C2	C3	S. no.	C1	C2	C3
1	ਇਸ	[isa]	This	2	ਜਸਿ	[jisa]	Who, what, which
3	ਵਚਿ	[vica]	In the	4	ਨ	[na]	No
5	ਤਕ	[taka]	Up	6	ਹੁਣ	[huṇa]	Now
7	ਵੀ	[vī]	Too	8	ਜਨਿਾਂ	[jināṁ]	Whom
9	ਉਤੋ	[othon]	Upon	10	ਨਾਲ	[nāla]	With
11	ਨਹੀ	[nahīṁ]	No	12	ਚਾਹੇ	[cāhē]	Either
13	ਭੀ	[bhī]	Too	14	ਕਸਿ	[kisa]	What
15	ਵਲੋ	[valōṁ]	By	16	ਪਛਿ	[pichōṁ]	After
17	ਇਹ	[iha]	This	18	ਏਧਰ	[ēdhara]	Around
19	ਏ	[iha]	This	20	ਨੂੰ	[nū]	To
21	ਜਦੋ	[jadōṁ]	When, while	22	ਅਜਹਿ	[ajihē]	Such
23	ਕਈ	[ka'ī]	Many	24	ਹੀ	[hī]	Only
25	ਤੱਦ	[tada]	Then	26	ਕੇ	[kē]	By
27	ਅੰਦਰ	[andar]	Within	28	ਹਾਂ	[hain]	Yes
29	ਉੱਤੇ	[utē]	Upon	30	ਬਹੁਤ	[bahuta]	Much
31	ਸਾਬੁਤ	[sābuta]	Complete	32	ਕਾਫ਼ੀ	[kāfī]	Enough
33	ਕਦੀ	[kadī]	Sometime	34	ਹੁਣੇ	[huṇē]	Now
35	ਨੈ	[nēṁ]	The	36	ਲਈ	[la'ī]	For
37	ਜੀ	[jī]	Respect	38	ਕਂ	[ki]	That
39	ਕਸਿ	[kisē]	Someone	40	ਮਗਰ	[magara]	Behind
41	ਪੂਰਾ	[pūrā]	Complete	42	ਦਾ	[dā]	Of
43	ਨੇ	[nē]	The	44	ਤਰ੍ਹਾਂ	[tar'hāṁ]	Like
45	ਹੋਵੇ	[hovē]	If	46	ਫੇਰ	[phēra]	Later

(continued)

Table 1 (continued)

S. no.	C1	C2	C3	S. no.	C1	C2	C3
47	ਜੇਕਰ	[jēkar]	Just in case	48	ਵੇਲੇ	[vēlē]	Times
49	ਦੇ	[dē]	Of	50	ਉੱਥੇ	[othē]	There
51	ਜਹਿੜਾ	[jēhara]	Which	52	ਕਤਿ	[kitē]	Somewhere
53	ਬਾਅਦ	[bā'ada]	After	54	ਇੱਥੇ	[ithē]	Here
55	ਸਾਰਾ	[sārā]	all,whole	56	ਜਨ੍ਹਿੰ	[jinhanu]	Whom
57	ਚੋ	[cho]	Out	58	ਜਦ	[jad]	When
59	ਕਦੀ	[kadē]	Never	60	ਵਾਂਗ	[vāṅga]	Like
61	ਸਭ	[sab]	All	62	ਦੌਰਾਨ	[doraan]	During
63	ਤਾਂ	[tan]	When	64	ਵਰਗਾ	[varagā]	Like
65	ਕਂਿ	[ki]	That	66	ਜੋ	[jō]	That
67	ਲਾ	[la]	To attach	68	ਕਰਕੇ	[karkē]	Because
69	ਪੂਰਾ	[pura]	Complete	70	ਬਲਿਕੁਲ	[bilkul]	Absolutely
71	ਨਾਲੇ	[naale]	Also	72	ਐਹੋ	[eho]	Such
73	ਤੋ	[ton]	From	74	ਕੌਣ	[kaun]	Who
75	ਹੋਣਾ	[hona]	Be	76	ਫਰਿ	[pher]	Then
77	ਪਾਸੋ	[paso]	From	78	ਤਦ	[tad]	Then
79	ਜਹਿਾ	[jeha]	Little	80	ਕੋਲੋ	[kolon]	From
81	ਏਸ	[ēs]	This	82	ਕਨਿਾ	[kina]	How much
83	ਜਨ੍ਹਿਾਂ	[jina]	Who	84	ਜਵੇਿ	[jivē]	Such as
85	ਕੁਝ	[kujh]	Some	86	ਹੇਠਾਂ	[hethan]	Below
87	ਦੁਆਰਾ	[dobara]	By	88	ਸਾਰੇ	[sarē]	All
89	ਸਦਾ	[sada]	Forever	90	ਜੱਥੇ	[jithē]	Where
91	ਏਥੇ	[ethē]	Here	92	ਕੋਈ	[koi]	Someone
93	ਬਾਰੇ	[barē]	About	94	ਕੀ	[ki]	What
95	ਕਦ	[kad]	When to	96	ਜੀ	[je]	Please
97	ਕਦੇ	[kadē]	Never	98	ਦੀਆਂ	[dī'āṁ]	Of
99	ਹੋਏ	[hoye]	Happen	100	ਚਲਾ	[chala]	Goes
101	ਰਹੇ	[rahē]	Are	102	ਲੈ	[lai]	Take
103	ਬਣੋ	[bano]	Become	104	ਆਖ	[aakh]	Say
105	ਦੇਣੀ	[dēṇī]	Give	106	ਬਣ	[baṇa]	Made
107	ਪਆਿ	[pi'ā]	Lying	108	ਕਰ	[kara]	Do
109	ਹੋਇਆ	[hō'i'ā]	Happened	110	ਪੈਣ	[pain]	Falling
111	ਗਈ	[ga'ī]	Gone	112	ਕਹਂਿ	[kēh]	Say
113	ਲਗ	[laga]	Seem	114	ਚੁਕੇ	[chukē]	–
115	ਹੁੰਦਾ	[hudā]	Happen	116	ਕਹਿਾ	[keha]	Said
117	ਜਾਂਦਾ	[jāndā]	Going	118	ਕਰਵਾਈ	[karvayei]	Conducted
119	ਵੇਖ	[vēkha]	See	120	ਬਣਾਏ	[banaye]	Created
121	ਸੁਣ	[suṇa]	Hear	122	ਕੀਤਾ	[kitta]	Carried out
123	ਆਈ	[ā'ī]	Occurred	124	ਜਾਵਣ	[javan]	Going
125	ਸਕਦੇ	[sakdē]	Can	126	ਦੇਖ	[dēkh]	See

(continued)

Table 1 (continued)

S. no.	C1	C2	C3	S. no.	C1	C2	C3
127	ਜਾਵੇ	[javē	Go	128	ਆਦਿ	[ādi]	So on
129	ਜਾਂਦਾ	[janda]	Going	130	ਲਿਆ	[li'ā]	Taken
131	ਕਰਣ	[karana]	Doing	132	ਆ	[ā]	Come
133	ਲਗਾਉਦਾਂ	[lagoda]	Not involving	134	ਰਹਿਾ	[reha]	Going
135	ਆਵੇ	[aavē]	Arrives	136	ਗਿਆ	[geya]	Been
137	ਕਰੀ	[kari]	Do	138	ਉਠ	[otha]	Arise
139	ਲਾਇਆ	[laeya]	Attach	140	ਰਹੀ	[rahi]	Been
141	ਰਹਿ	[reh]	Living	142	ਉਸਨੇ	[usnē]	He
143	ਉਹ	[uha]	He, she	144	ਤੁਸੀ	[tusi]	You
145	ਸਾਂ	[sāṁ]	Was	146	ਮੇਰਾ	[mera]	My
147	ਸਭ	[sabha]	All	148	ਉਸਦੀ	[usdi]	His
149	ਹਨ	[hana]	Are	150	ਤੇਰਾ	[tera]	Your
151	ਤੂੰ	[tu]	You	152	ਉਸ	[us]	His
153	ਸੀ	[si]	Was	154	ਉਏ	[oyē]	Person
155	ਹੋ	[ho]	Are	156	ਆਪ	[aap]	you
157	ਤੈਨੂੰ	[tēnu]	You	158	ਸਨ	[san]	Was
159	ਤੁਸਾਂ	[tusa]	You	160	ਮੈ	[mein]	I
161	ਹੈ	[hain]	Are	162	ਤੁਸੀ	[tusi]	You
163	ਹੈ	[hai]	Is	164	ਅਸੀ	[assi]	We
165	ਆਪਣਾ	[apna]	My	166	ਪਰ	[par]	but
167	ਜੇ	[jē]	If	168	ਤੇ	[tē]	And
169	ਅਤੇ	[aatē]	And	170	ਤਾਂ	[tāṁ]	So
171	ਜਾਂ	[jāṁ]	Or	172	ਭਾਵੇ	[bhāvēm]	Although
173	ਕੁਲ	[kal]	Total	174	ਅਗਲੀ	[aagali]	Next
175	ਵਗ਼ੈਰਾ	[vaġairā]	Etc.	176	ਵਰਗ	[varg]	Category
177	ਰੱਖ	[rakh]	Put	178	ਆਮ	[āma]	Common
179	ਲੱਗ	[laag]	Take	180	ਲਾ	[lā]	Apply
181	ਗੱਲ	[gal]	Thing	182	ਹਾਲ	[hāla]	Condition
183	ਪੀ	[pī]	Drink	184	ਇੱਕ	[ek]	One

On lieu of Punjabi Grammar and Part of Speech (POS) based word class categorization, stop words are categorized into 4 different word classes: Adverbs [20], Conjunctions [21], Verbs [20], Pronouns [20] and other miscellaneous words. Any word which is not suitable for first four categories is assigned to miscellaneous one. 98 different adverb forms, 43 different verbs, 24 pronouns, 7 conjunctions are identified from 184 stemmed stop words. And remaining 12 stop words are assigned to miscellaneous category.

Adverb forms in Punjabi language are classified into 2 categories: by function and by form [20]. By function, adverb clauses are categorized into following

subclasses: Adverb clause of time: ਜਦ [*jad*] 'when', ਅੱਜ [*ajj*] 'today'. Adverb clause of place: ਉੱਪਰ [*uppar*] 'upon', ਉੱਤੇ [*uttē*] 'over'. Adverb clause of purpose: ਨੂੰ [*nu*] 'to', ਲਈ [*laii*] 'for'. Adverb clause of manner: ਜਵਿ [*jive*] 'as', ਉਵੇ [*ove*] 'custom'. Condition clause: ਅਗਰ [*agar*] 'if', ਜੇਕਰ *[jekar]* 'in case'. Result clauses: ਨਾਲੋ [*naalo*] 'concurrent', ਤੋ [*to*] 'from'. Adverb clause of degree: ਬਹੁਤ [*bahut*] 'much', ਕਾਫੀ [*kafi*] '*enough*', ਸਾਬੁਤ [*sabut*] 'complete'. By form, adverb clause is divided into subgroups like derived adverbs, pure adverbs, phrasal adverbs, clausal adverbs, reduplicated adverbs and particles. Few examples are like ਇਥੇ [*ethe*] 'here', ਉਥੇ [*othe*] 'there', ਕਥਿ [*kithe*] 'where', ਜਥਿ [*jitho*] 'where', ਕਥਿ [*kitho*] 'where'. List of Adverbs are shown from serial number 1–98 in Table 1.

Verbs found among stop words are shown in Table 1 from serial number 99–141. For example: ਆਉਣਾ [*aaouna*] 'to come', ਜਾਣਾ [*jaana*] 'to go'.

Pronouns are shown from serial number 142–165. For example ਉਸਦਾ [*usda*] 'his', ਅਸੀ [*assi*] 'we'.

Conjunctions are used to join words, phrases, and clauses [21]. For example, ਅਤੇ [*atē*] 'and', ਜਾਂ [*jāṃ*] 'or'. Serial number 166–172 presents conjunction list.

Miscellaneous words are present from serial number 173–184. For example: ਵਰਗ [*varg*] 'category', ਵਗੈਰਾ [*vagera*] 'etc'.

4 Conclusion

An automatic poetry classifier is used to classify poems according to its content. Before classification starts, these poetries must have to pass through various pre-processing phases. 256 stop words are identified from poetry and news articles written in Punjabi. These 256 stop words are stemmed to its root form using Punjabi stemming rules. Analysis of 184 stemmed stop words from grammatical point of view is discussed in this paper. These stop words are categorized into adverbs, pronouns and conjunctions. In this paper, 184 stemmed stop words are presented for future use in other NLP task in Gurmukhi script. This paper provides enhanced understanding of stop words in light of part of speech tags in Punjabi language.

References

1. Sebastiani, F.: Machine learning in automated text categorization. ACM Comput. Surv. **34**, 1–47 (2002)
2. Languages of India.: http://en.wikipedia.org/wiki/Languages_of_India#Prominent_languages_of_India
3. Punjabi Language.: http://en.wikipedia.org/wiki/Punjabi_language
4. Poem.: http://oxforddictionaries.com/definition/english/poem
5. Kaur, J., Saini, J.R.: A study and analysis of opinion mining research in Indo-Aryan, Dravidian and Tibeto-Burman Language families. Int. J. Data Mining Emerg. **4**(2), 53–60 (2014)

6. Ali, R.A., Maliha, I.: Urdu text classification. In: 7th International Conference on Frontiers of Information Technology, ACM New York, USA, (2009). ISBN 978-1-60558-642-7, doi:10.1145/1838002.1838025
7. Mansur, M., UzZaman, N., Khan, M.: Analysis of N-Gram Based Text Categorization for Bangla in a Newspaper Corpus. Center for Research on Bangla Language Processing. BRAC University, Dhaka, Bangladesh (2006)
8. Mohanty, S., Santi, P.K., Mishra, R., Mohapatra, R.N., Swain, S.: Semantic based text classification using wordnets: Indian language perspective. In: 3rd International Wordnet Conference (GWC 06). pp. 321–324 (2006). doi:10.1.1.134.866
9. Nidhi., Gupta, V.: Domain based classification Punjabi text documents. In: International Conference on Computational Linguistics, pp. 297–304 (2012)
10. Sarmah, J., Saharia, N., Sarma, S.K.: A novel approach for document classification using assamese wordnet. In: 6th International Global Wordnet Conference, pp. 324–329 (2012)
11. Murthy, K.N.: Automatic Categorization of Telugu News Articles. Department of Computer and Information Sciences, University of Hyderabad, Hyderabad (2003). doi:202.41.85.68
12. Rajan, K., Ramalingam, V., Ganesan, M., Palanive, S., Palaniappan, B.: Automatic classification of Tamil documents using vector space model and artificial neural network. Expert Syst. Appl. **36**(8), 10914–10918 (2009)
13. Jayashree, R.: An analysis of sentence level text classification for the Kannada language. In: International Conference of Soft Computing and Pattern Recognition, pp. 147–151 (2011)
14. Gupta, V., Lehal, G.S.: Preprocessing phase of Punjabi language text summarization. In: International Conference on Information System for Indian languages, vol. 139, pp. 250–253 (2011)
15. Unicode Table. http://www.tamasoft.co.jp/en/general-info/unicode-decimal.html
16. Stemming. http://en.wikipedia.org/wiki/Stemming
17. Gupta, V.: Automatic stemming of words for Punjabi language. In: Advances in Signal Processing and Intelligent Recognition systems, Advances in Intelligent Systems and Computing, vol. 264, pp. 73–84 (2014)
18. Google Translation. https://translate.google.co.in/#auto/en/%E0%A8%AA%E0%A8%8F
19. Transliteration and Translation. http://www.shabdkosh.com/pa/
20. Bhatia, T.K.: Punjabi: a cognitive-descriptive grammar. Rout ledge Descriptive Grammar Series (1993)
21. Overview of Punjabi Grammar. http://punjabi.aglsoft.com/punjabi/learngrammar/?show=conjunction

Techniques and Challenges in Building Intelligent Systems: Anomaly Detection in Camera Surveillance

Dinesh Kumar Saini, Dikshika Ahir and Amit Ganatra

Abstract Security is tedious, complex and tough job in today's digitized world. An attempt is made to study and propose an intelligent system for surveillance. Surveillance camera systems are used for monitoring and controlling the security. Anomaly detection techniques are proposed for designing the intelligent control system. In the paper challenges in detection and processing of anomaly in surveillance systems are discussed and analyzed. Major components related to an anomaly detection technique of camera control system are proposed in the paper. Surveillance data is generated through camera, and then this data is transmitted over the network to the storage. Processing is to be done on real time basis and if there is any anomaly detected, the system must produce an alert. This paper is an attempt to study soft computing approaches for anomaly detection.

Keywords Surveillance · Systems · Camera · Control · Anomaly · Detection

1 Introduction

Surveillance cameras are used extensively for security purposes. Access and control of these cameras through a remote computer over the web improves the security aspects of the system. The camera control system consists of a set of cameras located at different locations and the cameras are controlled remotely. Multimedia is

D.K. Saini (✉)
Faculty of Computer and Information Technology, Sohar University,
Sohar, Oman
e-mail: dinesh@soharuni.edu.om

D. Ahir · A. Ganatra
Charotar University of Science and Technology, Changa,
Gujarat, India
e-mail: 14pgce001@charusat.edu.in

A. Ganatra
e-mail: amitganatra.ce@charusat.ac.in

S.C. Satapathy and S. Das (eds.), *Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 2*,
Smart Innovation, Systems and Technologies 51, DOI 10.1007/978-3-319-30927-9_2

the combination of different media like text, still images, audio, video, animation and graphics [1]. Generally multimedia content is bulky, so the media storage and the cost of transmission are noteworthy. To solve this, media are compressed in file for both streaming and storage. Streaming is a means of sending multimedia information over the internet so that the recipient plays it as it is being transmitted. Multimedia involves buffering mechanism (temporary storage). In that the limited segments of the streamed information is temporary stored for continuous play. Streaming avoids the copy and save/store an entire file. Without store entire file user can play data [2]. In the camera control systems anomaly detection can be used as one of the mechanism for building the intelligent system.

2 Anomaly Detection

Anomaly detection in data is the identification of patterns that do not match with normal behavior. And these anomalous patterns are also referred to as exceptions, outliers, novelties, noise, deviations, discordant observation in different application domain [3]. Anomaly detection in data is important because many times the detected anomalies translate the important, actionable information in many application domains. Here are some examples, in computer network is abnormal traffic pattern is found then it mean that sensitive information is transfer by hack computer to an unauthorized receiver [4].

An anomaly means something unusual, irregular or unexpected behavior that does not match to the normal behavior. Because of many kind of reasons anomalies could be occur in data. For example network intrusion, cyber intrusion, or terrorist activity and break down of a system.

Figure 1 shows a simple example of anomaly in two-dimension data set. In which R1 and R2 are two normal regions. The points which are outside normal region are considered as anomalies. Here point's a1, a2 and points in region a3 are anomalies.

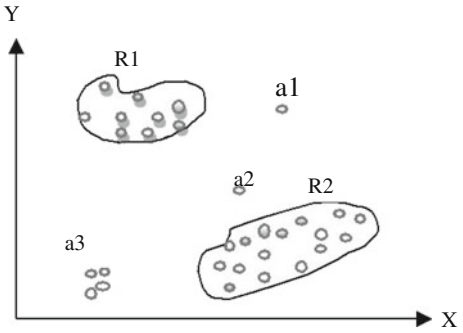

Fig. 1 Anomalies in a two-dimensional data set [5]

2.1 Challenges in Anomaly Detection

A straight forward approach to anomaly detection is defining a region that represents a normal behavior and remaining part that not belong to normal region declared as anomalies. But some factors make this approach very difficult [5].

- It's hard to define a region that comprehends all probable normal behavior. Because the boundary of anomalous and normal behavior is not much accurate. Thus an anomaly observation that situated close to the boundary may be normal or abnormal.
- While anomaly is aeries due to the malicious action, these malicious rivals often accommodate themselves to appear abnormal behavior as normal so the task of labeling normal behavior region becomes more difficult.
- By the time in various application domains, normal behavior is keep growing so the current conviction of normal behavior is may not be enough in future.
- Another challenge is, for different application domain the exact view of anomaly is dissimilar. For example, a small variation in normal reading might be diseases in medical domain. Whereas, in stock market domain small variation might be consider as normal. This makes it complex to apply a particular domain technique to another domain.
- When using recorded and real world datasets, the major issue is labeled data availability.

Sometimes noise is similar to anomaly and it is difficult to distinguish between noise and an anomaly because the data contains noise is analogous to actual anomalies hence it is hard to recognize and remove. Anomalies and noise removal are related to but both are two different things. Since the difference between noise and anomalies is lies in the interest of analyst. Noise can be refer as something unwanted and obstacle to data analyst. While anomalies are consider as something meaningful to data analyst.

3 Different Aspect of Anomaly Detection Issues

3.1 Input Data Nature

Input data nature is core aspect of Anomaly detection. In general the input consists of set of data instances which can call as sample, objects, record, point, vector, entity, pattern, event or case. And all data instance can be defined by a set of attributes. This attributes are also referred as variable, characteristics, feature, field or transmission. There are different types of attributes such are binary, categorical or continuous [5].

In choice of anomaly detection technique the characteristics of attributes have a major impact. For example, in statistical techniques the underlying model is

depends on whether the attributes data types is continuous or categorical. Likewise, in nearest-neighbor-based technique, the distance measure used in technique is determined by the characteristics of attributes.

There is another way of categorizing the input data. It is based on the relationship between the data instance. Mostly the anomaly detection technique deals with single point data. In point data there is no relationship among the data instance.

3.2 Data Labels

Data labels describe whether the data instance is normal or anomalous. Data labeling is expensive and usually it is done by human expert.

The systems are becoming extremely complex and dynamic and to understand the systems behavior it requires exploratory data analysis and descriptive modeling.

3.3 Types of Anomaly

3.3.1 Point Anomaly

Most common type of anomaly is point anomaly and has been focus on most of research. Point anomaly can be defined as an individual entity which is considered as abnormal with regards to other data. In Fig. 1 point a1, point a2 and points in region a3 are point anomalies as they are exterior to the normal region.

3.3.2 Contextual Anomaly

Contextual anomaly is also known as conditional anomaly. It can be defined as in some specific context if a data instance is anomalous then it is contextual anomaly. These types of anomalies are usually found out in spatial data and time series data.

3.3.3 Collective Anomaly

Collective anomaly can be defined as a set of related data instances which is anomalous with regard to remaining data in data set. Individual entities may not be anomalies by themselves in collective anomalies but their occurrence to gather is considered anomalous.

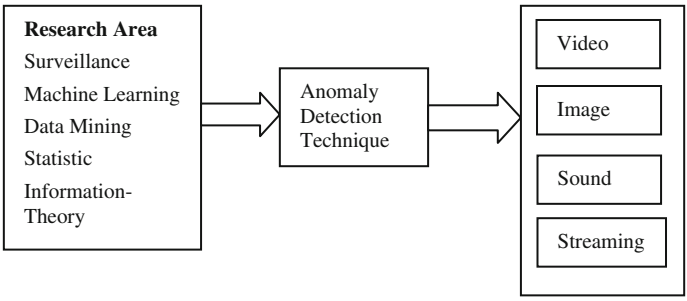

Fig. 2 Key component related to an anomaly detection technique of surveillance camera

4 Application Domain of Anomaly Detection

There are varieties of application domain for anomaly detection; such are network intrusion detection, industrial damage detection, machine learning, statistic, fraud detection, image processing, video surveillance, intrusion detection, medical science and public health. Figure 2 illustrates the component of camera control system associate with anomaly detection techniques [6]. In camera control system Input data for anomaly techniques are video data, image data, sound, or streaming. How this system work is described in next section.

5 Camera Control System

Figure 3 shows the basic block diagram for camera control system. Source in surveillance system is camera and the target is the entity or entities. The target consists of entities in which anomaly detection technique is used to find anomalies. Example of target can be crowds, road traffic, individuals, or network traffic. Data is taken from source and stored at server. From server streaming (streaming is transferring of data) of data is next part. Data can be type of text, image, voice, animation, or video. Streaming can be done in real-time/online or offline.

At server processing of data is done, this data can be recording or imaging. But this system is only focus on image and video part. Image processing is one part of this system. Processing of image is done by mathematical operations. At receiver side data preprocessing is done [7]. Data preprocessing includes editing, cleaning, modify, and feature extraction. Feature extraction is an essential preprocessing step. Feature extraction is used to reduce the required amount of resource to