Assessing Emotions in a Cross-Cultural Context

Maria Alejandra Quiros-Ramirez
Onisawa Lab., Graduate School of Systems
and Information Engineering University of Tsukuba,
1-1-1 Tennodai, Tsukuba, 305-8573 Japan
alejandra@fhuman.esys.tsukuba.ac.jp

Takehisa Onisawa
Graduate School of Systems
and Information Engineering University of Tsukuba,
1-1-1 Tennodai, Tsukuba, 305-8573 Japan
onisawa@iit.tsukuba.ac.jp

***Abstract*—Emotion recognition systems could support professionals in a wide range of areas. Several work in emotion recognition has been carried out in the last decades, yet few attention has been payed to cross-culture context emotion recognition. Multimodal emotional expressions from 36 subjects with different cultural backgrounds were collected. In the experiment, participants observed and assessed emotional images in a 5 point positive and negative emotional valence scale. This information was used as ground truth for the recorded information. The dataset was segmented for all the participants and partially labeled for 8 of them, for a total of 160 segments. Recognition of positive and negative emotions was obtained from the dataset suggesting agreement points in expression of emotion between cultures.**

***Index Terms*—emotion encoding, emotion recognition, cross-cultural, context, emotion database, universality, specificity, valence**

I. Introduction

Communication is the basis of our daily interaction with other people and a complex process to transmit personal ideas to another individual. This process becomes even more challenging when the cultural background is different among interacting people. Emotions are basic components of the communication process as well. Emotional messages through non-verbal behavior support and modify our communication [1].

Automatic understanding and assessment of emotions could bring strong benefits to a wide variety of areas [2]. Since the beginnings of Affective Computing [3], the field has advanced quickly: with initial attempts of recognition of emotion from face-only pictures evolving to the current complex signal arrangements and multicue recognition systems.

Yet, not much focus has been given to cross-cultural recognition of emotions. It is still an open question whether or not emotions are universal [4]. Universality of emotions can represent a problem in trying to build a single detector that fits any individual despite cultural background.

A. Universality Background

Universality of emotions has been debated since Darwin [5]. In his work, Darwin presented correlations between facial expressions and emotions in different subjects. Ekman's work [6] in multiculturality has backed up the universality of emotions through studies carried out within different ethnic groups. Russell [7] presented strong evidence to disprove Ekman's theory of universality and recent evidence of differences in emotion perception [8] questions the universality of facial expressions.

It is important to point out that most of the work that has been done in order to assess universality of emotions, has been carried out through the study of facial cues. Furthermore, even though there are several theories of emotion available [9], these studies mainly focus on the discrete categorization of emotions, e.g. happy, sad, angry.

B. Related Emotion Recognition Research

The field of affective computing has advanced a lot in the previous decades [10]. Major efforts have been carried out in each of the steps required to build an emotion recognition system: data collection, modelling, analysis and interpretation. Multimodality and continuous affect are characteristics of the most recent systems. Even though researchers have explored different theories of emotion, for example, categorical and dimensional [9], most of the available systems assume universality of emotions. Little attention is put in the design of systems that consider cultural context, thus aiming to model a system that can decode emotions of any individual without considering his or her cultural background.

Even though the majority of the systems are converging towards multimodality [10], culture related emotion recognition systems are still focused on one single cue, for example, body posture [11] or speech [12]. Very few information is found about facial expressions and gestures in automatic emotion recognition systems that consider cross-cultural context.

C. Issues related with Cross-Cultural Emotion Recognition

Scherer [4] describes expressions of emotion as a mix of psychobiological, sociocultural and epochal factors. In his study he presents evidence of the ongoing debate about universality versus specificity of emotions. His findings suggest that emotion encoding and decoding depend on the context of the interaction. Finally, he recommends *multimodality* in order to study further cross-cultural emotions.

Another latent problem at the time of studying cross-cultural interactions is language. Haidt et al. [13] describe emotion words as *poor anchors* for cross-cultural comparisons. They point as well the need of looking beyond the six most common emotions for this type of comparisons.

TABLE I
SUBJECTS BREAKDOWN BY REGION

Region	Sub-Region	#	Total
Africa	North	3	4
	Central	1	
America	Caribbean	2	8
	Central	3	
	South	3	
Asia	East	5	15
	Central	4	
	West	1	
	South-East	3	
	South	2	
Europe	West	2	7
	South	3	
	East	2	
Oceania	Australia	1	2
	Melanesia	1	

D. *Scope of the paper*

Currently, interaction among people from different cultures is not rare. Based on the above mentioned studies, it is clearly necessary to include and understand cross-cultural context in future emotion recognition systems. Such a consideration could create emotional agents that successfuly support the professionals in the tasks that require interpersonal communication and assessment of affect in real life situations. Also, a cross-cultural context inclusive system could help to build a bridge in communication between people of different cultural backgrounds.

The purpose of this study is to provide a cross-cultural emotion dataset that allows the analysis of multimodal emotion. Thus, a system based on this data would aim to recognize emotions from subjects with different cultural backgrounds.

Instead of considering discrete emotions, the analysis is focused on understanding and assessing positive and negative emotions: thinking of a continuous space, instead of labeling emotions with words, we define them as states that are positive or negatives and an intermediate state which is considered neutral. For example, feelings of happiness and amusement are considered into the positive group while feelings of anger and sadness are considered into the negative group.

An emotion recognition system that works as a neutral decoder of emotions encoded by subjects raised in different cultural backgrounds is presented. Thus, it is possible to explore the encoding process of emotions in relation with culture.

In order to perform this test, an experiment to collect cross-cultural data from facial gestures, head and body motions was carried out.

The paper is organized as follows. In section 2 the proposed methodology to create an emotional dataset and the analysis to assess emotions in cross-cultural context is presented. Section 3 explains the process of data collection and labeling of the data. In section 4 the emotion recognition system is presented followed by conclusions and future work in section 5.

II. PROPOSED METHODOLOGY

The process of building an emotion recognition system starts by choosing available emotion data, otherwise collecting emotion data from scratch. Even though there are some emotion databases available [10], none of them meet our requirements: naturalistic emotions, subjects of different cultural backgrounds, none invasive devices. This last point is considered to maintain a natural scheme of interaction without attaching devices to people, emulating real life scenarios. Data available in different databases belongs mainly to subjects of a single cultural background.

In the previous section, important points that require attention while considering cross-cultural context in emotions are presented:

- *Multimodality*. An arrange of devices that allow the study of real time multimodality is prepared. For this study, the cues of interest are head, facial expressions and body gestures.
- *Theories of emotion*. Even though most of the research in universality or specificity of emotions revolves around the six basic emotions outlined by Ekman [6], dimensional emotions are considered in this paper. In this case the dimensional meassure employed is *valence*, which represents how aversive (negative values) or attractive (positive values) is an interaction.
- *Language*. Assessing emotions discretely forces subjects to assign linguistic values to these emotions. In a cross-cultural environment, linguistics could be the source of encoding or decoding bias. Dimensional affect assessment eliminates the linguistic variable by exchanging it for an evaluation scale. A five point value assessement scale is used. Besides the assessment of emotions, stimulus that requires a deep understanding of a specific language could be another cause of bias. Therefore, pictures are utilized as stimulus since they do not require any linguistic explanation.

III. DATA COLLECTION

The recording devices used were two high speed cameras to capture facial and head information and two high definition cameras, one for the head and one for the body.

A special room with no windows was used in order to control the experiment's illumination settings. Three different sources of lightning were used, two from the sides of the recording array and one on top of it. All the sensors were synchronized together in order to be able to retrieve the correct samples from each of the devices accurately in time to analyse multimodal cues.

Pictures were selected as a non-linguistic stimulus to obtain spontaneous emotional displays from the participants. Images from the GAPED database [14] were selected. This database provides a value for each picture which corresponds to the emotional valence. This value was used to evaluate the picture as positive or negative, which represents the picture's *emotional ground truth*.

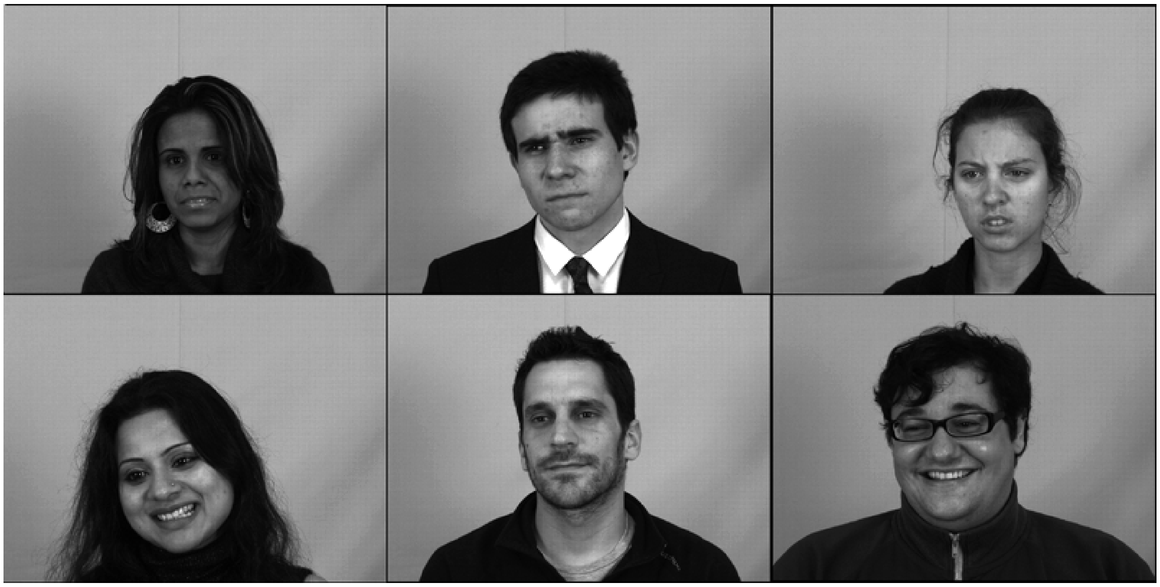

Fig. 1. Some expressions captured during the experiment. The first row presents expressions obtained while the participants viewed negative pictures. The second row presents expressions obtained while the participants viewed positive pictures. The participants in the first row come from Jamaica, France, Costa Rica; in the second row, India, Spain, Brazil, respectively. The subject's emotional self-evaluation was used as emotional tagging to avoid any bias from an external labeler.

A. Subjects

Thirty six naive people from different countries participated voluntarily in the experiment. Their ages range from 21 to 35, 14 of the participants were female and 19 male. A region breakdown of the participants can be observed in Table I. Each of them had different educational backgrounds from undergraduate to post-doctoral fellows, from the University of Tsukuba and nearby research centers. English proficiency ranged from intermediate to native. Five of the subjects wore glasses and one had beard.

B. Experimental Setup

In the first stage of the experiment, the subject was asked to observe images displayed in the screen. Twenty emotionally loaded images were selected from the GAPED affective picture database: 8 described as positive, 8 described negative and 4 described as neutral.

During the experiment, a grey screen would be displayed for 3 seconds before each picture. The picture itself would be displayed for 5 seconds. The subject was asked to look at it carefully. After each picture, the subject was instructed to evaluate aloud his or her feeling about the picture in a five point scale from -2 to 2 representing negative to positive feeling. The scale would be displayed in the screen everytime after each picture. The task took approximately two and a half minutes to be completed. The pictures were presented randomly to each participant.

Figure 1 shows some still shots obtained from the experiment. The upper row of the figure presents three different participants while they were observing images which they rated themselves as negative. The lower row of the figure presents expressions of participants who were observing pictures which they rated as positive.

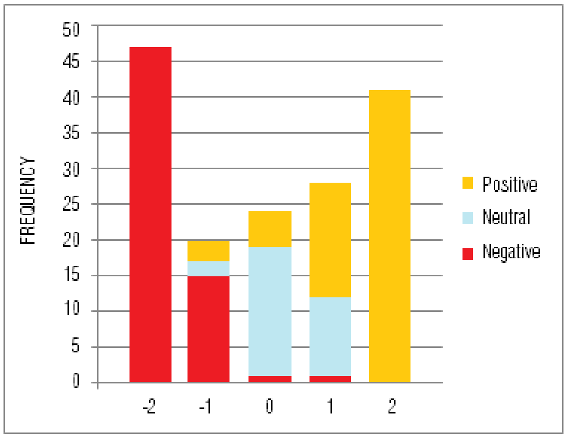

Fig. 2. GAPED database evaluation information vs subjects emotional self report of the labeled grouped. The horizontal axis represents the subject's five point evaluation; vertical axis represents the frequency. Colors represent the original value of the picture. Disagreement between the original value and the emotional assessment of the subjects can be observed mainly in the neutral pictures, which were frequently evaluated as positive and sometimes negative.

C. Data segmentation and labeling

Processing the data after collection is a challenging process [2]. In order to lighten this task, different segmentation and labeling techniques were employed.

In order to perform automatic segmentation of the collected data, a photosensor was placed in the screen where the pictures were presented to the subject. This sensor was able to capture the onset and offset of the picture's presentation in order to

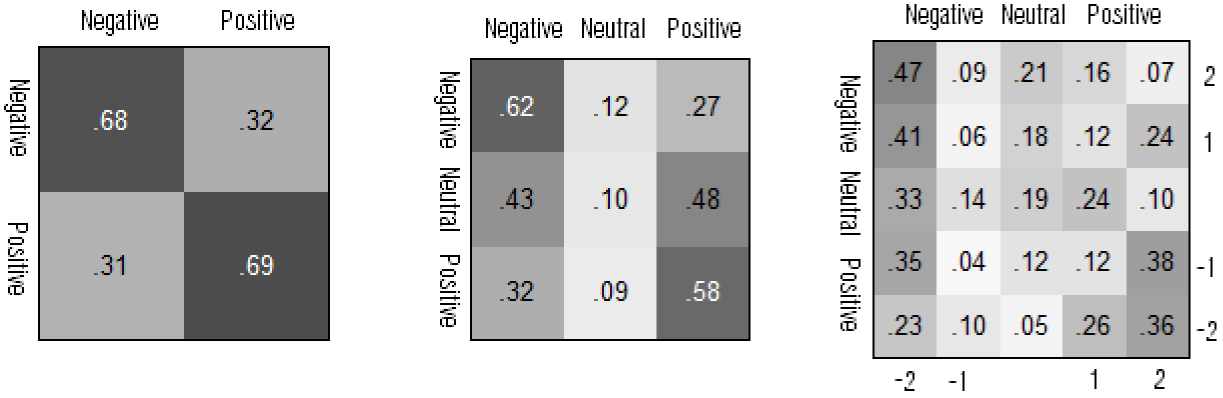

Fig. 3. Confusion Matrices. From left to right, confusion matrix of 2 point, 3 point and 5 point scale valence classification schemes. Rows represent the prediction. Columns represent the ground truth. Ground truth was obtained from self-evaluation of the participants in an original 5 point scale of their feelings (valence). For comparison purposes, this scale was separated as well in three main valence groups: positive, negative and neutral as well as two point scale without considering neutral stage.

tag in time the region of interest of our recordings. Twenty recordings of interest were obtained from each participant, for a total of 720 segments which represents approximately 30 minutes of expressions and interactions.

The subject's assessment of his or her feeling from observing the picture was used as a ground truth for training the system. This allows us to avoid labeling biases from annotators that can be inside or outside of the culture of the subject.

Twenty nine features were labeled: 19 facial features, 5 head motions and 5 body motions.

- Facial features: Inner eyebrows up, outer eyebrow raiser, eyebrow lower, frown, eyelid tightener, eyelids towards eachother, multiple blinks, smile, laugh, abnormal breathing, nose wrinkler, jaw drop, lip pressor, lip suck, lip corner puller, lip corner depressor, jaw sideways, swallow, chin raiser.
- Head features: move head, move head away, nod, say no, tilt head.
- Body features: move finger up and down, move hands, touch or scratch with the hand, press hands, move leg.

Eight participants' data was feature-labeled, for a total of 160 segments. This 160 segments are represented by 67 negative pictures (47 rated as -2 and 20 rated as -1), 24 neutral pictures and 69 positive pictures (28 rated as 1 and 41 rated as 2). Figure 2 shows a comparison between these 8 subjects evaluation and the picture's ground truth obtained from the GAPED database.

The pictures that presented most disagreement between original evaluation inside the database and the rated evaluation during this experiment were the neutral pictures: 3% of the negative pictures was classified as positive or neutral, 13% of the positive pictures were rated as neutral or negative while 41% of the neutral pictures catalogued as positive or negative. This issue contributed in lowering the expected corpus of neutral expressions during the data collection.

TABLE II
F1-SCORE AND PREDICTION RATE (ACCURACY) FOR EACH VALENCE CLASS IN THREE DIFFERENT POINT SCALES

	F1 Score					Accuracy
Valence	Negative		Neutral	Positive		
2 Points	0.68		NA	0.69		0.69
3 Points	0.58		0.111	0.59		0.53
5 Points	0.421	0.065	0.190	0.111	0.424	0.237
	-2	-1	0	1	2	

IV. EMOTION RECOGNITION MODEL

The data collection explained in the previous section represents our cross-cultural dataset. Each entry of the dataset corresponds to the interaction of the subject with each of the presented emotional pictures. This means for each subject, there are 20 entries in the dataset. Each entry contains information of the subject's face, head and body motions and the emotional value assessed by him or herself.

In order to test the capability of recognizing positive and negative emotions from the presented cross-cultural dataset, an emotion recognition system was trained and tested. Features from face, head and body motions were labeled to serve as input of the model as follows.

Even though more features were labeled at the beginning, only the most significant ones were selected to feed the model. In this ocassion, a feature is considered as significant if it is observed more than 5 times for at least two subjects' face, head and body movements, in order to avoid features that are individual expressions instead of emotion related expressions.

As explained before we consider three emotional groups: positive, negative and neutral. For the training stage, feature vectors of face, and head and body motions are inputted to the system as well as the corresponding emotion obtained from the subject's emotional self-assessment. The output of the system is an emotional value. According to the theory of universality, emotions from people with different backgrounds can be recognized without influence of their culture [4]. In this case, following this theory, the subject's nationality is not inputted to the model in order to test if emotions can be recognized equally through our subjects despite of their cultural background. Thus, our model considers cross-culturality from a data point of view (in the dataset).

An implementation of *Support Vector Machines* from SVM-KM Toolbox [15] with Gaussian kernel was employed for both training and testing. The selected standard deviation of the kernel was 10 for the classification of data separated in three and two valence groups and 5 for the classification of data separated in five valence groups.

The data of eight participants (from Jamaica, France, Costa Rica, India, Spain, Brazil, Hungary and Japan) was labeled with the features of face, head and body motions previously listed. A total of 160 samples are obtained.

A leave one out cross-validation (LOOCV) procedure was selected in order to use all the samples of the 8 participants for the model. This procedure consists of training the model with *n-1* samples and testing it with the remaining sample, where *n* represents the total amount of samples. The training is performed *n* times, excluding one different example each training. The output of this procedure is a confusion matrix obtained by using each test sample exactly once.

The LOOCV procedure is used for training and testing our data set, in order to use every sample as an independent test. This procedure has been chosen instead of partitioning the data in train-test set to avoid biasing the model towards the expressions of a single participant, considering that our samples come from only 8 different individuals.

The system was first trained using an emotional input of the three emotional groups: negative, neutral and positive. In order to assess the influence of neutral expressions, the system was then trained using only positive and negative emotional groups, excluding the vectors that belonged to neutral emotions. Finally, in order to analyze the strength difference between the emotions of the same group, the system was trained using the five categories of strength of emotion (from -2 to 2).

Table II presents the F1-score [16] and the prediction rate obtained for each of the training strategies explained. The confusion matrices for each grouping are presented in Figure 3, rows represent the predicted value and columns represent the labeled value (ground truth self-assessed by each subject).

Observing the results by class (negative - neutral - positive) the neutral gave the lowest results in both 3 and 5 emotion points groupings. The low performance in the detection of neutral expressions may be due to two factors: the original amount of segments that were supposed to collect neutral expressions were almost halved according to the self-assessment of the subjects. This creates an imbalance in the amount of samples for each class (sparseness per class). This same issue causes 5 point scale classification to have the worst prediction rate. This could be reduced by adding images for training in future data collection experiments. Second, it is possible that the expressions between valence that is close to neutral (-1, 1) are ambiguous in comparison with expressions that have higher valence (-2, 2).

Finally, considering a general classification between positive and negative emotions, it is possible to get good prediction results from the model. This can be observed in both the 3 and 2 point scales. It is clearer when the neutral group is excluded.

V. Conclusion and Future Work

In this paper the need to take in account cross-cultural context in the process of developing an emotion recognition system was introduced. A new cross-cultural emotion dataset was presented as well as segmentation and labeling techniques in order to tackle the challenging points of the postprocessing of data after the recording.

Emotion recognition of positive and negative emotions was obtained from our cross-cultural dataset. This finding suggest that it is possible to find agreement points between the expression of dimensional emotions between cultures. The neutral expressions were difficult to distinguish from mild positive and negative emotions. Subjectivity in the neutral expressions of different participants was observed. It is necessary to develop a different procedure in order to decode neutrality.

As future steps, it is necessary to automate feature extraction of the different cues of interest. Several interesting data was not analysed for this paper, but a deeper and more detailed labeling could bring better and more conclusive results about the classification of cross-cultural emotions. Analysis of a greater sample of the dataset will be performed in order to study this point. It is necessary to study further the disagreement points in expression between cultures, and consider including the cultural background as an input for the model, in order to improve the recognition rates.

The work in order to expand the dataset will continue by increasing the amount of subjects per culture and by performing comparative experiments to investigate further the principle of universality and specificity of emotion.

Acknowledgment

The authors would like to thank Senya Polikovsky and Dr. Yoshinori Kameda for their partnership and support during the development of this study. We appreciate the collaboration of the CSS laboratory, for lending us the experimental room in order to run our recordings. Special gratitude to the people who voluntarily participated as subjects for this experiment.

References

[1] A. Bartsch and S. Trewin, "Emotional Communication - a Theoretical model," *Proceedings of the IGEL 2004*, 2004.

[2] R. Cowie, "Building the databases needed to understand rich, spontaneous human behaviour," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–6.

[3] R. W. Picard, *Affective Computing*. Cambridge, Massachusetts: The MIT Press, 1997.

[4] K. R. Scherer, E. Clark-polner, M. Mortillaro, K. R. Scherer, E. Clark-polner, and M. Mortillaro, "In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion," *International Journal of Psychology*, vol. 46, no. 6, pp. 401–435, 2011.

[5] C. Darwin, *The expression of the emotions in man and animals*. London: John Murray, 1872, freeman #1141.

[6] P. Ekman, "Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique." pp. 268–87, Mar. 1994.

[7] J. a. Russell, “Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies.” *Psychological bulletin*, vol. 115, no. 1, pp. 102–41, Jan. 1994.

[8] R. E. Jack, C. Blais, C. Scheepers, P. G. Schyns, and R. Caldara, “Cultural confusions show that facial expressions are not universal.” *Current biology : CB*, vol. 19, no. 18, pp. 1543–8, Sep. 2009.

[9] I. B. Mauss and M. D. Robinson, “Measures of emotion: A review.” *Cognition & emotion*, vol. 23, no. 2, pp. 209–237, Feb. 2009.

[10] H. Gunes, B. Schuller, and M. Pantic, “Emotion representation, analysis and synthesis in continuous space: A survey,” *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pp. 827–834, 2011.

[11] A. Kleinsmith, P. De Silva, and N. Bianchi-Berthouze, “Cross-cultural differences in recognizing affect from body posture,” *Interacting with Computers*, vol. 18, no. 6, pp. 1371–1389, 2006.

[12] N. Kamaruddin, A. Wahab, and C. Quek, “Cultural dependency analysis for understanding speech emotion,” *Expert Systems with Applications*, vol. 39, no. 5, pp. 5115–5133, Apr. 2012.

[13] J. Haidt, “Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition,” *Cognition &amp; Emotion*, vol. 13, no. 3, 1999.

[14] E. S. Dan-Glauser and K. R. Scherer, “The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance.” *Behavior Research Methods*, vol. 43, no. 2, pp. 468–477, 2011.

[15] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, “Svm and kernel methods matlab toolbox,” 2005.

[16] C. van Rijsbergen, *Information Retrieval*, Butterworth, 1979.