Identifying Raga Similarity in Hindustani Classical Music through Distributed Representation of Raga Names

Joe Cheri Ross[1], Rudra Murthy[1], Kaustuv Kanti Ganguli[2], and Pushpak Bhattacharyya[1]

[1] Dept. of Computer Science & Engineering
[2] Dept. of Electrical Engineering
Indian Institute of Technology Bombay, Mumbai, India
joe@cse.iitb.ac.in

Abstract. This paper investigates a method to identify the similarity between ragas in Hindustani classical music using textual discussions on ragas. The proposed approach learn representations in the form of word vectors for raga names, which can be used to extract similarities between ragas. This can be applied for generating word vectors for any specific domain with data scarcity. The method learns word vectors of dataset specific words (absent in the general pool of word vectors) with the help of the specific dataset and pre-trained word vectors trained on general English corpus like Google word vectors. The proposed method yields better results over the baselines signifying its appropriateness to any specific dataset.

1 Introduction

Hindustani music is the north Indian style of Indian classical music and is one among the two classical music traditions in India (south Indian tradition is Carnatic music). Raga is the predominant melodic concept in Indian classical music, which can be considered as the tonal framework for composition and improvisation [1]. A raga defines the permitted notes, their order and hierarchy, style of intonation of individual notes, relative duration and their specific melodic approach [2]. A raga ontology [3] is integral to knowledge representation approaches for Hindustani classical music, especially for representing meta-data information along with the audio content information. Identification of similarities between raga is essential to build a raga ontology depicting the relationships between the ragas as well.

Generally similarity between ragas is inferred through attributes associated with the ragas. For instance, in Hindustani music, ragas can be classified based on the tonal material involved, which is termed *thaat*. There are 10 *thaats* in Hindustani music [4]. *prahar*, *jati*, *vadi*, *samvadi* etc are the other attributes on which the similarity between ragas is based on. Most of the accepted similarities between ragas encompass the similarities in many of these attributes. These similarities cannot always be derived exclusively from these attributes. Trained musicians can perceive these similarities efficiently, but cannot always be statistically inferred from compositions or performances. Many text resources with description on ragas available in the form of books, websites, forums, etc [5, 6, 7] have excellent depictions of these similarities. **This work intends to identify similarities between Hindustani ragas based on word vectors learned from textual descriptions and discussions on Hindustani ragas.** Word vectors are continuous vector representations of words, learned with very large datasets capturing the similarity between the words [8].

The available pre-trained word embeddings trained on general English domain, like Google word vectors[3] are useful for many natural language processing tasks. This serves as an excellent resource even for tasks in a specific domain, but devoid of dataset specific words. Mostly the size of these specific datasets is not sufficient enough to produce quality word embeddings for the specific words using the approach by Mikolov [8]. In this paper we propose a novel method to learn word vectors for words specific to a dataset, but absent in a general English corpus. This problem is motivated by the task of identification of raga relationships from limited available discussions and descriptions on Hindustani ragas, but not confined to it. Here, the similarity between the generated word vectors for the raga names is taken as the measure of similarity between ragas.

The proposed method trains word vectors for only the raga words (dataset-specific words) keeping the common words' word vectors constant. Ghosh *et. al.* [9] proposes a method (*Dis2Vec*) to learn domain-specific word vectors for disease words. Different from our approach, this does not attempt to resolve insufficiency of available text through utilization of pre-trained word vectors (further explained in Section 3).

The similarity between the word vectors of the ragas represents the relationship between the ragas. Identifying raga relationships has immense applications to music recommendation tasks for Indian classical music and this attempt is first of its kind to the best of our knowledge.

2 Background

Raga is the foremost and possibly the most rudimentary and fundamental concept in both the classical music traditions of India. There are manifold ways a raga can be defined, based on the scope of a specific paradigm. In principle, a raga (the closest concept in Western music is the mode) is a melodic framework where certain rules are applicable on a set of notes. Though it is highly discussed that Indian classical music is an improvisatory art music tradition, there is a strong underlying structure that an artist follows; which also gives an experienced listener a cue about the raga identity. These rules are often posed in the form of attributes in the musicology literature. Certain attributes impose constraints on the movements of the notes of a raga, while some are independent of the tonal material. The former includes: the set of notes (usually five to eight notes in an octave) that relates to the *thaat* (melodic mode in Western music), ascent and descent note sequences (*aroha-avaroh* which in turn decides the *jaati*), the resting notes (*nyas svara*), the focal note (*vadi-samvadi svaras*) etc. An example of the latter is the *prahar* (time of the day when the raga should be performed) which is one of the most discussed attributes for characterizing a raga, especially in the musicological texts.

[3] GoogleNews-vectors-negative300: https://code.google.com/archive/p/word2vec/

Raga similarity can be thought of as a linear combination of the attribute-similarity among the ragas. Though, certain attributes are relatively dominant (e.g., the *thaat* to which a raga belongs to) and hence carries more importance in the similarity function. An example of such a family of ragas will be discussed subsequently (*Kalyan thaat* ragas). Another yet important factor of raga similarity is the characteristic phrases of a raga. There are cases where a pair of ragas belonging to two different *thaats* are perceived similar due to overlapping (or high resemblence between) characteristic phrases (also known as *pakads*). Even there are examples of two ragas being annotated 'similar' with respect to one attribute, while certain other attributes are distinctly contrasting. Thus the space of raga similarity would appear different if the relative weights of the attributes are altered.

We take an example of the family of *Kalyan thaat*, where raga *Kalyan* is the first representative raga. The ragas *Suddh Kalyan, Shyam Kalyan, Anandi Kalyan* are example members of the first group above mentioned, having a suffix of *Kalyan*, confirming that the relationship with the parent is both in terms of notes and characteristic phrases. The other siblings like *Kedar, Kamod, Hameer, Chhayanat, Bhupali* are having some fuzzy membership assigned to their relation to raga *Kalyan* (based on characteristic only), at the same time they are also children of *Bilawal thaat* in terms of notes. *Bilawal thaat* has other children as varieties of (*Prakaar*) of *Bilawal* like *Alhaiya Bilawal*. Others are *Deshkar, Bihag* etc. There is a special candidate called raga *Maru Bihag* which is a child of *Bihag* in terms of its characteristic and note sequence, at the same time it has a relation to raga *Kalyan* based on the notes. Another example raga *Deshkar* is a child of *Bilawal thaat*, having a sibling *Bhupali* which is also related to raga *Kalyan* with respect to its characteristics.

There can also be hidden relationships between the ragas. For instance S N S D n G R[4] is a characteristic phrase in raga *Jaijaiwanti*. If we transpose the key to its current dominant, it becomes P M P G m N D which is a characteristic phrase in raga *Hameer*. If the tonic of a performance in *Jaijaiwanti* is perceived to be P note, the raga of the *Hameer*. Such a relationship is termed as *'murchhana'* in musicological texts, but this attribute is seldom used as a criterion for raga similarity because of its non-obviousness in the listener's ears.

3 Raga Word Vector Generation

For generating raga names word vectors, we utilize textual resources having discussions on Hindustani ragas and available pre-trained word vectors like Google word vectors. Pre-trained word vectors employed for this is referred as *base word vectors*. Google word vectors are 300 dimensional vectors trained with 3 million words and phrases. The size of available text discussing ragas is not sufficient enough to train word vectors with the available Mikolov's approach [8]. To address this, our approach introduces a modification to the training of neural network architecture by Mikolov. Mikolov's approach has a feedforward neural network, with input layer, one hidden layer and output layer. The size of the input layer is V (size of vocabulary) and hidden layer size is N

[4] In Hindustani notation system S r R g G m M P d D n N corresponds to C C$^{\#}$ D D$^{\#}$ E F F$^{\#}$ G G$^{\#}$ A A$^{\#}$ B notes in western music notation system, when the tonic is at C

(word vector dimension). The weights between input and hidden layer is represented as $WI \in \mathbb{R}^{V \times N}$ and weights between hidden layer and output layer is represented as $WO \in \mathbb{R}^{N \times V}$. For word vector training in Mikolov's approach, the weight matrices are randomly initialized and WI is taken as the word embeddings for the V words post training [8].

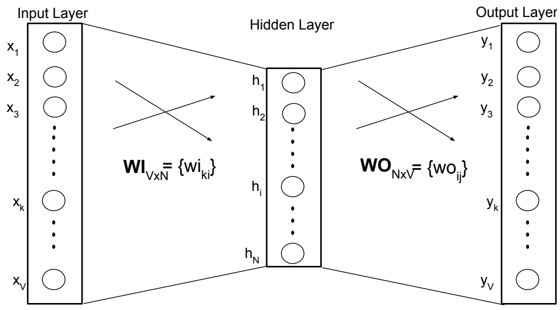

Fig. 1. Mikolov's approach neural network model [10]

In our proposed method, WI and WO are pre-loaded with base word vectors for the words available in the base word vectors. For the rest of the words including raga names, corresponding WI and WO vectors are randomized as done in Mikolov's approach. A list of raga names are made available and while training the word vectors, only the WI and WO vectors corresponding to the raga names are updated.

The word vectors learned for the common words from a larger context bear the expected meaning of the word which do not differ much in this specific domain text. But word vectors for common words trained from a small dataset may not encompass the actual meaning and will eventually affect the training of dataset specific rare words. The proposed modification to word vector training helps to prevent the common word vectors from getting changed and eventually affecting the raga names word vectors. **Pre-processing**: There is a scarcity in the availability of text content with discussions related to Hindustani music and raga. To make maximum utilization of the available content some preprocessing is essential. Spelling discrepancies in raga names severely affects the word vectors learned for a raga name, since the diversity in spelling reduces the frequency of raga names in the text. To address this, all the occurrences of a raga in the text are corrected to have the same spelling using string matching based on edit distance [11]. Also multi-word raga names are combined to single entity with the special character '_'.

4 Datasets

Table 1 shows the details of the datasets. The datasets are used in combination for different experiments. *Swaraganga* [6] and *Parrikar* [5] has raga related information extracted from 2 different websites on Hindustani music. *Wikipedia* dataset has Wikipedia pages on Hindustani ragas and *Hindustani Music In the 20th Century* is a book by Wim van der Meer [7]. *Swarganga* and *Wikipedia* have description of the ragas in a more formal way explaining the main attributes and some relations, whereas *Parrikar* and *Hindustani Music In the 20th Century* have more subjective description from musicologist's point of view. The first 2 datasets being small in size are used in combination and the rest 2 of significant size are added to this in the other experiments. 154 Hindustani ragas are considered for this study.

Dataset	**Size (#words)**
Swarganga website (*sganga*)	17912
Wikipedia (*wiki*)	24094
Hindustani Music In the 20th Century (*meer*)	99252
Parrikar website (*parrikar*)	115796

Table 1. Details of Datasets

5 Experiments

As discussed, our method updates word vectors for specific words in a dataset, preserving common words present in pre-trained word embeddings. word2vec[5] (as per Mikolov's approach) released by Google [12] is modified to accommodate the proposed design. We evaluate the performance of this method based on the quality of similarity between ragas observed through raga word vectors. The *cosine similarity* between word vectors is taken as the measure of similarity. Many of the tasks involving word vectors have identified cosine similarity as the effective similarity measure [13]. Experiments are performed with combinations of datasets mentioned in Table 1. For all our experiments with Mikolov's approach (baseline) and our approach, parameters are set to default values[6] except for *min-count*, which defines the minimum frequency required for a word to be considered. *min-count* is set to less than 4 for the experiments involving raga datasets to ensure the inclusion of meagerly occurring raga names and related words to the vocabulary.

Two baseline methods are devised with Mikolov's approach. First baseline (**BL-1**) generates word vectors from raga datasets alone, whereas the second baseline (**BL-2**) combines general English corpus with raga datasets. The general English corpus is formed combining DBPedia short abstracts and long abstracts, Europarl English monolingual corpus [14], news commentary corpus (released for WMT 2011 shared task). Table 2 shows the results of the baseline experiments.

[5] http://word2vec.googlecode.com/svn/trunk/
[6] word2vec parameters: window=10, iter=3 and negative=10

5.1 Quantitative Evaluation

The non-availability of ground truth for this task makes the evaluation challenging. We designed an evaluation method considering one strong attribute of similarity between ragas, the *thaat*. This evaluation method is based on the assumption that majority of the ragas identified similar to a particular raga will belong to the same *thaat* [2]. The evaluation method described in Algorithm 1, takes 10 most similar ragas to each raga and computes the score based on the number of similar ragas belonging to the same *thaat* as that of input raga (r). For our experiments this serves as an acceptable metric for comparison.

Algorithm 1 Compute evaluation score

Require: list of raga names ($ragas$), word vectors
Ensure: $score$
$score \leftarrow 0$
$total\_no \leftarrow 0$
for all r in $ragas$ **do**
    $sim\_ragas \leftarrow$ 10 most similar ragas to r based on cosine similarity between raga word vectors
    $t \leftarrow$ thaat of r
    $score+ = \sum_{r_s \in sim\_ragas} belongs\_to(r_s,\ t)$
    $total\_no+ =$ length(sim_ragas)
$score \leftarrow score/total\_no$

5.2 Qualitative Evaluation

Qualitative evaluation is also done with the help of a trained Hindustani musician, to evaluate to what extent the identified similarities align with perceived similarities by the musicological community.

6 Results

This section describes the results of quantitative and qualitative evaluation.

6.1 Quantitative Evaluation

Table 2 shows the baseline results with raga datasets combinations and aforementioned general English corpus. Table 3 compares the accuracy of different *min-count* with different dataset combinations taking the base word vectors as Google and general English Corpus (*Gen English*) with our approach. The experiment with all the datasets combined shows best accuracy and *min-count*= 1 gives better accuracy through out.

Compared to the baseline experiments (Table 2), our approach shows better accuracy for all the dataset combinations. BL-1 performance is affected due to insufficient text.

Datasets	Score	
	BL-1	BL-2
sganga + wiki	0.0813	0.0945
sganga + wiki + meer	0.1063	0.1023
sganga + wiki + meer + parrikar	0.1016	0.1117

Table 2. Results: Baseline experiments (score: obtained with the described evaluation method, *min-count*=1).

Baseline BL-2 is better comparable with the experiments taking *Gen English* as base word vectors in Table 3, since it uses this same general English data combined with raga datasets. Our approach has clear performance improvement over BL-2 with *Gen English* experiments. Baselines are executed with *min-count*=1. One key observation here is, word vectors generation done with *min-count*=1 adversely affects the common word vectors which in turn affects the raga word vectors (as in BL-2). Our method rectifies this with the two-step word vectors training with variable *min-count* (*min-count*=10 for generating base word vectors and *min-count* < 4 in the incremental generation step).

6.2 Qualitative Evaluation

Qualitative evaluation is done with the help of 2 Hindustani musicians. One is trained professional musician in Hindustani music and the other is an amateur with basic understanding of Hindustani music. The main analysis produced here is done by the professional musician and the amateur musician did overall analysis confirming many of the former's observations. The musicians were provided with a t-SNE visualization [15] of the raga word vectors and text listing 10 closest (based on cosine similarity) ragas to each raga. This visualization with some of the observed similarities magnified, is shown in Figure 2. The musicians evaluated for all 3 dataset combinations.

sganga+wiki+meer+parrikar and *sganga+wiki +meer* are found to be superior to *sganga+wiki* w.r.t the similarities captured. Among them, *sganga+wiki+meer+parrikar* has more relations captured. This observation has a direct correlation with the results in Table 3 proving the validity of the quantitative evaluation method. Overall many of the close ragas in the visualization belong to the same *thaat* in all the 3 cases. Since *thaat* based similarity corresponds to melodic similarity and is considered to be prime among the other raga attributes, the identified similarities sound meaningful.

Ragas from *Kafi, Asavari, Bhairavi thaats* are closely placed. These *thaats* have a minor scale with a change in a single note. A few ragas which are perceived as similar and have less commonalities w.r.t raga attributes are identified similar (eg. *Patdeep*, *Nanad* and *Tilak Kamod, Komal Rishabh Asavari*). These are perceived similar because of the similarities in their characteristic phrases. Between *Patdeep* and *Nanad*, the main characteristic phrase differ by a note. Certain pentatonic ragas *viz.*, *Bhoop, Gunkri, Arabhi, Shivaranjani* and *Malkauns* are found close to each other (refer (b) in Figure 2). Certain clusters are observed with ragas similar w.r.t group (eg. *Malhar* group ragas are found together) and *prahar*. (c) in Figure 2 shows 2 different clusters of malhar group ragas. (d) shows that the *misra ragas Lalit bhatiyar, Jogkauns, Nat Bihag, Nat Bhairav,* *Asa Mand, Puriya Kalyan, Puriya Dhanashree, Bhupal Todi, Basanti Kedar* are placed closer. Some of the ragas adopted from Carnatic music are identified closer to each other.

From the 2D visualization, the amateur musician made an interesting observation that many of the ragas belonging to the same *thaat* lie in a diagonal strip on the plane, with a negative slope. Also frequently sung ragas are found closer to the center of the visualization, where as rare ragas are seen away from the center.

Along with these positive observations there are a few undesired similarities appearing. A few pair of ragas which are supposed to be far apart are seen closer(eg. (*Basant*, *Bageshri*), (*Poorvi*, *Hameer*)). Most of the clusters of similar ragas identified, are found to be grouped together based on one of the attributes, which the musicological community also considers relevant for that grouping. The low number of invalid similarities compared to missing valid similarities, shows high precision of the system.

Experiment	**Base word vectors**	**Score**		
		min-count=1	**min-count=2**	**min-count=3**
sganga+wiki		0.0867	0.0695	0.0664
sganga+wiki+meer	Google	0.1141	0.0828	0.0844
sganga+wiki+meer+parrikar		0.1172	0.1039	0.0977
sganga+wiki		0.1016	0.0648	0.0672
sganga+wiki+meer	Gen English	0.1078	0.0945	0.0977
sganga+wiki+meer+parrikar		0.1140	0.0992	0.1047

Table 3. Results of our approach with different base word vectors and *min-count* (score: obtained with the described evaluation method).

Dataset	**Experiment**	**Homogeneity**	**Completeness**	**V-measure**	**Rand**	**Mutual Information**
US political tweets	BL-1	0.4185	**0.5439**	0.4730	0.1619	0.2802
	BL-2	0.3124	0.3519	0.3310	0.1425	0.1506
	Our Approach	**0.4583**	0.5069	**0.4813**	**0.1719**	**0.3215**
CoNLL 2003 NER dataset	BL-1	0.0088	0.0088	0.0088	0.0201	0.0085
	BL-2	0.0025	0.0197	0.0045	0.0088	0.0022
	Our Approach	**0.0120**	**0.0250**	**0.0163**	**0.0464**	**0.0117**

Table 4. Results of clustering evaluation with 2 different datasets (best results shown in bold).

7 Discussions

In the paper so far, we performed our experiments on Hindustani raga related text. To evaluate our approach better , we also experimented with generating word vectors for

2 different datasets *viz.* US political tweets [16] and CoNLL 2003 NER [17] datasets. For a set of words defined as dataset-specific words in the datasets we learn the word vectors, pre-loading the word vectors for other words in the dataset. Clustering based evaluation shows how good these dataset-specific words get clustered according to their known classes.

From US political tweets, position words belonging to 6 different political issues (*eg. gun laws, immigration, insurance etc.*) are considered as dataset-specific words. From CoNLL 2003, all the entities belonging to 4 named entity classes are taken as dataset-specific words. Here the evaluation is done by analyzing how well these words get clustered according to their known classes. In the US political tweets dataset, the position words are expected to be clustered according to the issues they belong to and in CoNLL 2003 dataset the entities are expected to be clustered according to their named entity classes.

Word vectors for these words are learned using the baseline approaches and our approach. K-means clustering is employed, defining k as the number of word categories. Our approach is compared with the baselines by evaluating the clustering of words using a set of standard metrics for clustering. *Table 4 shows that our approach performs better for both the datasets w.r.t all the metrics, with a single exception.* Along with affirming the applicability of our approach for any dataset with rare words, this evaluation also helped to perform a strong validation with tasks having solid ground truth.

8 Conclusion and Future Work

The proposed method generates word embeddings from the available limited text on Hindustani ragas, as a means to identify the similarity between ragas. Quantitative and qualitative evaluations substantiates that the similarities between the raga word vectors emulate perceived raga similarities. Comparison with the baseline methods confirms the utility of this method to get word vectors for words specific to a dataset. Further validation of the method is done by clustering evaluation performed with 2 other datasets and this confirms the practicality of the approach to any dataset. As future work, we plan to improve this method by considering the relative importance of context words variably, as certain word categories are more important for this task.

9 Acknowledgment

This work received partial funding from the European Research Council under the European Union's Seventh Framework Programme (FP72007-2013)ERC grant agreement 267583 (CompMusic).

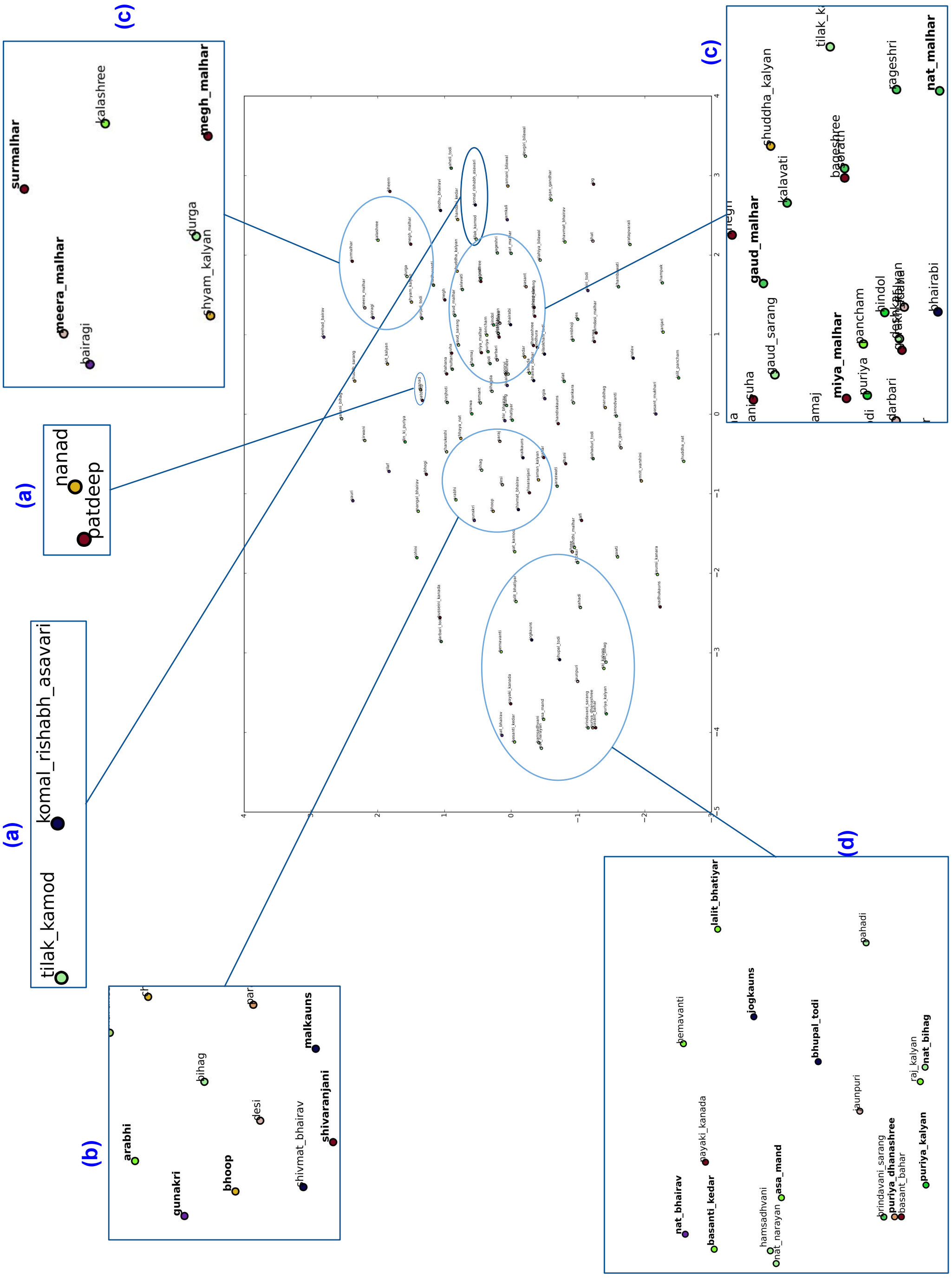

Fig. 2. TSNE visualization of raga similarity with valid similarity clusters shown in magnified windows. Relevant raga names in a magnified window are shown in bold

Bibliography

[1] Suvarnalata Rao, Joep Bor, Wim van der Meer, and Jane Harvey. *The raga guide: a survey of 74 hindustani ragas*. Nimbus Records with Rotterdam Conservatory of Music, 1999.

[2] Suvarnalata Rao and Preeti Rao. An overview of hindustani music in the context of computational musicology. *Journal of New Music Research*, 43(1):24–33, 2014.

[3] Gopala Krishna Koduri and Xavier Serra. A knowledge-based approach to computational analysis of melody in indian art music. In *International Workshop on Semantic Music and Media colocated with International Semantic Web Conference*, pages 1–10, 2013.

[4] Soubhik Chakraborty, Guerino Mazzola, Swarima Tewari, and Moujhuri Patra. *Computational Musicology in Hindustani Music*. Springer, 2014.

[5] Rajan Parrikar. parrikar.org, 2000.

[6] Swarganga. Swarganga.org, 2004.

[7] Wim van der Meer. *Hindustani music in the 20th century*. Springer Science & Business Media, 2012.

[8] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[9] Saurav Ghosh, Prithwish Chakraborty, Emily Cohn, John S. Brownstein, and Naren Ramakrishnan. Characterizing diseases from unstructured text: A vocabulary driven word2vec approach. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM '16, pages 1129–1138, New York, NY, USA, 2016. ACM.

[10] Xin Rong. word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*, 2014.

[11] William E Winkler. String comparator metrics and enhanced decision rules in the fellegi-sunter model of record linkage. 1990.

[12] Yoav Goldberg and Omer Levy. word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*, 2014.

[13] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Hlt-naacl*, volume 13, pages 746–751, 2013.

[14] Philipp Koehn. Europarl: A parallel corpus for statistical machine translation. In *MT summit*, volume 5, pages 79–86, 2005.

[15] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.

[16] Aditya Joshi, Pushpak Bhattacharyya, and Mark Carman. Political issue extraction model: A novel hierarchical topic model that uses tweets by political and non-political authors. In *WASSA, Proceedings of NAACL-HLT*, pages 82–90, 2016.

[17] Erik F Tjong Kim Sang and Fien De Meulder. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In *Proceedings of*

the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4, pages 142–147. Association for Computational Linguistics, 2003.