Australian Journal of Basic and Applied Sciences, 7(9): 35-43, 2013
ISSN 1991-8178

Predicting hearing loss symptoms from Audiometry data using FP-Growth Algorithm and Bayesian Classifier

[1]Nasir G. Noma, [1]Mohd Khanapi Abd Ghani, [2]Mohamad Khir Abdullah, [2]Noorizan Yahya

[1]Biomedical Computing & Engineering Technologies (BIOCORE Malaysia), Applied Research Group, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, 76100 Hang Tuah Jaya, Durian Tunggal, Melaka
[2]Department of Otorinolaringologi Hospital Pakar Sultanah Fatimah Muar Johor, Malaysia

Abstract: This paper presents the results of applying machine learning algorithms to predict hearing loss symptoms given air and bone conduction audiometry thresholds. FP-Growth (frequent pattern growth) algorithm was employed as a feature extraction technique. The effect of extracting naïve Bayes classifier's vocabulary from patterns generated by FP-Growth algorithm was explored. Both multivariate Bernoulli and multinomial naïve Bayes models were used with and without the feature extraction. The results were validated with repeated random sub-sampling validation performed using 5 partitions with 10, 20, 30, 40 and 50 training examples respectively averaged over 10 iterations. The multivariate Bernoulli model with feature extraction is found to be more accurate in predicting hearing loss symptoms with average error rate of only 0, 0.5, 1, 1.75 and 5.4% for the partitions with 10, 20, 30, 40 and 50 training examples respectively compared to multinomial model with feature extraction. However, the two models with feature extraction produce better results than same models without feature extraction.

Key words: Threshold, multivariate Bernoulli, multinomial, naïve Bayes, FP-Growth, audiometry

INTRODUCTION

Machine learning is starting to transform many fields from medicine and biology to business, economics, engineering and many other areas. The possibilities are obvious in IBM's Watson computer which uses machine learning technique to help doctors make better decisions by suggesting appropriate treatment when presented with a set of symptoms (Reynolds, 2011). The use algorithms to find and extract relationships in data dates back to 1950's(Guyon *et al*., 2006). Within this data lies valuable patterns and information, which can be extracted with machine learning algorithms. These patterns can become source of valuable information which can assist the clinicians in the process of diagnosis and decisions that can improves patient outcomes (Sullivan and Wyatt, 2006).

The FP-growth algorithm is an efficient way of finding frequent patterns in a dataset. The two most conventional ways of finding interesting patterns in data are association rules and frequent item sets(Cabena *et al*., 1997). FP-Growth algorithm is built from Apriori algorithm (Harrington, 2012) with different techniques of generating item sets that commonly occur together in a dataset. FP-Growth algorithm can find frequent items more efficiently than Apriori algorithm in that it scans the dataset only twice to generate item sets. The first scan counts the frequency of all the items and removes the infrequent ones. The rule is, superset containing that item will also be infrequent and the second scan builds the FP-tree data structure that is traversed to mine frequent item sets(Harrington, 2012). He *et al*.(2005) Proposed an efficient algorithm based on the FP-Growth for mining the complete set of all correlated pairs on transaction databases. FP-tree data structure has been used in combination with decision tree to generate CT scan brain images and classify the images for diagnosis (Rajendran and Madheswaran, 2010). FP-tree data structure is efficient and scalable for mining both long and short frequent patterns and is believed to be faster than any frequent pattern mining algorithm (Han *et al*., 2004). Storing of the datasets in the FP-tree structure results to faster execution time than many of the item sets generation algorithm commonly used, with two orders of magnitude better in performance than algorithm like Apriori(Harrington, 2012). FP-Growth algorithm and association rule mining were employed for the purpose finding relationships in pure-tone audiometry data set of 50 hearing loss patients with 0.1 as the minimum support and 0.7 confidence thresholds for the association rule mining. Positive correlations were found between the symptoms of tinnitus, vertigo, giddiness and some pure-tone audiometric thresholds (Noma and AbdGhani, 2012).The extension of that studycarried out on 339 patients with 0.2 (20%) as the support threshold for item set generation and 0.7 (70%) as the confidence value for association rule mining. The findings shows similar relationship between those particular symptoms with hearing thresholds earlier discovered (Noma *et al*., 2013).

Corresponding Author: Nasir G. Noma, Biomedical Computing & Engineering Technologies (BIOCORE Malaysia) Applied Research Group, Faculty of Information and Communication Technology,Universiti Teknikal Malaysia Melaka 76100 Hang Tuah Jaya, Durian Tunggal, Melaka
E-mail: nasnoma@hotmail.com

The relationship between hearing thresholds and other attribute in the patients' medical data is evidenced by several studies. Hwang *et al*.(2008) have discovered a relationship between Meniere disease and low frequency sensorineural hearing loss. Agrawal *et al*.(2009) have found a connection between high-frequency sensorineural hearing loss and noise exposure, and the association of cardiovascular risk generated by smoking and diabetes with both high and low-frequency hearing loss. Pure-tone air and bone conduction thresholds audiometry were obtained from unscreened population of older adults to observe the rate of changes over the period of 10 years. The baseline measurement predictors for that period are reported for specific age groups of both genders. The threshold changes over the said period for those at the age of 50 to 60 years old were found to be more at the higher frequencies while changes in threshold were at lower frequencies for the aged that are 80 years and above(Wiley *et al*., 2008). Age-related changes in pure-tone audiometry were also observed in a longitudinal study on hearing thresholds of 813 adult males. The results shows steady rate of hearing loss at higher frequencies and increase in the rate of hearing loss at lower frequencies (Brant and Fozard, 1990). A study on the effect of age and noise related hearing loss on high frequency thresholds was carried out on 187 industrial noise-exposed and 52 non-industrial noise-exposed subjects. The test-retest in the study shows high frequency audiometry (HFA) as a technique that is as reliable as the conventional in indicating noise-induced hearing loss and can be used more reliably to monitor hearing loss of individual cases over time period. Results from this study show both exposed and non-exposed subjects to have hearing loss at high frequencies (10-18 kHz). The effect of age seems more predominant than that of noise in the higher frequencies, this influence is not the same for the conventional frequency range (0.25-8kHz) (Ahmed *et al*., 2001). In a similar study (Gates *et al*., 2000), results suggest ear with noise-induced hearing loss does not aged at the same rate with non-noise damaged ear. The study examined 15 year change in audiometric threshold of 203 men with mean age of 64 years to determine whether high frequency notches influences auditory aging. To determine if the rate of change in pure-tone hearing thresholds is differed by sex among Korean subjects, a slope of linear regression was used to measure the rate of change in pure-tone thresholds at 0.25-8kHz. Results indicates significant sex differences in pure-tone thresholds with thresholds of women lower than men at frequencies above 2kHz and hearing impairment worse in men than women at high frequencies 4-8kHz (Kim *et al*., 2010). In another cross-sectional study to find differences in age-related audiometric results, a group of 473 subjects aged between 70 to 75 years were examined using pure-tone audiometry. Results indicate no significant differences in pure-tone thresholds between these age groups (Jönsson *et al*., 1998). To study longitudinal changes in thresholds and the effects of threshold levels on these longitudinal changes, high frequency pure-tone thresholds of 188 older adults between the ages of 60 to 81 years were analysed using the slope of linear regression to measure the rate of change in pure-tone thresholds (Lee *et al*., 2005). (Blanchet *et al*., 2008) suggests pure-tone audiometry as a reliable method to describe hearing status of a population after cross-sectional analysis of pure-tone hearing thresholds of 778 subjects of French descent aged 70 years and above. In a 5-year study, 342 diabetic veterans and 352 non-diabetic veterans who were tested on different audiometric measures, including pure-tone thresholds, results shows diabetic patients at the age of 60 years and younger have hearing loss at the high frequencies (Vaughan *et al*., 2006). De Heer *et al*.(2009) also relate age and genes to specific audiogram shapes.

The application of naive Bayes to solve text classification problems in computer science domain dates back about 5 decade ago (Lewis, 1998). Naive Bayes is a subset of Bayesian decision theory that is based on probability theory. It is named after Thomas Bayes, who was a Presbyterian minister that proved a special case of what is now referred to as Bayes' theorem (Hamelryck *et al*., 2012). A number of machine algorithms are based on probability theory. Naïve Bayes is a very fast algorithm. It is robust to irrelevant features because they tend to cancel each other without affecting results. It works well in domain where there are equally important features that turns out to be a problem to other classifiers particularly decision trees that have advantages in numeric domains but do not work well if many features are equally important. It is a simple approach that works well and can outperform more sophisticated classifiers on many datasets(Witten *et al*., 2011). Naive Bayes has been proven to have advantageous properties and effective at classification than other approaches like support vector machines, boosting, genetic algorithms, decision tree and k-means algorithms (Ghosh *et al*., 2008). The first three require iterated evaluation while decision tree will be unsuccessful because the tree building is slower than Bayesian table construction (Harrington, 2012) and k-means requires a lot of calculations compared to naïve Bayes probability calculation (Ghosh *et al*., 2008). Furthermore, naive Bayes has low storage requirement because it needs only to store token counts rather than the whole message (Harrington, 2012).

Based on two algorithms discussed above, this paper presents the results of applying these machine learning algorithms to predict hearing loss symptoms given air and bone conduction audiometry thresholds. The effect of extracting naïve Bayes classifier's vocabulary from patterns generated by FP-Growth algorithm was explored.

MATERIALS AND METHODS

Audiometric data of 399 patients across all ages (3-88 years) measured at 11 frequencies ranging from 0.125-8kHz from were collected over 10-year period from 2003 to 2012. The thresholds were the patients' first audiometry test results obtained from Hospital PakarSultanah Fatimah Muar, Johor, Malaysia file archive. Each patient's audiometric data includes the diagnosis information during the first visit. This is to avoid obtaining improved result after patient undergoes treatments. The data was collected using hand-held digital scanner that is used to scan the document. It is saved in portable document format (PDF) which is manually converted into digital format by entering every detail into a text document. The results are text data in form of diagnosis information and numeric data in the form of audiometry thresholds. The data sets that are composed of both air and bone conduction audiometry hearing thresholds and the symptoms information were processed with FP-Growth algorithm to extract frequent items sets. 0.4 (40%) was chosen as the minimum support threshold for frequent item sets generation. This is because rules are generally uninteresting if they apply to less than 10% of the dataset (Witten *et al*., 2011). In this regard, only those training examples that have item sets that pass the minimum threshold as their subset form the training set for the naïve Bayes classifier. The classifiers vocabulary is extracted from the union of those item sets rather than the conventional way of extracting the vocabulary from all the training examples.

There were 3 categories of classes of symptoms information discovered among 242 training examples which were extracted from the 399 hearing loss patients' data. There are training examples with tinnitus as their symptom, some have both tinnitus and vertigo and others with giddiness, tinnitus and vertigo. This means the data sets have 3 labels which the algorithm will use to train the classifier to automatically detect whether a symptom given the air and bone conduction audiometry thresholds is tinnitus, tinnitus and vertigo or a combination of giddiness, tinnitus and vertigo.

The features that make up the vocabulary are patterns extracted from the data set with FP-Growth algorithm. Feature extraction is a way of reducing the dimensionality of a data. It involves feature construction and selection(Guyon *et al*., 2006). One of the ways feature construction pre-processed the data is through dimensionality reduction. The features in the vocabulary are made up of item sets from the pre-processed data. This has reduced the number of the features that composed the vocabulary and at same retaining relevant information. This can reduce storage requirements, noise and computational cost thereby improving the performance of the naïve Bayes algorithm. Processing becomes faster with fewer features (Guyon *et al*., 2006). Some of the popular dimensionality reduction techniques are principal component, independent component analysis, and factor analysis. Principal component analysis is the most widely used (Harrington, 2012).

The multivariate Bernoulli model (Aggarwal andZhai, 2012)has been preferred over the multinomial model(Aggarwal andZhai, 2012). The multivariate Bernoulli model takes the vocabulary and a training example and returns a binary vector of 1s and 0s that represents the presence or absence of a hearing threshold. The length of the binary vector is the length of the vocabulary. These results to as many features as there are thresholds in the vocabulary. While the multinomial model returns the fraction of times in which a threshold value appears among all thresholds in training examples of a given class. In this study's approach, the frequency of occurrence of a threshold value is not important as the presence or absence of a threshold value in a training example is more preferred. Hence, the multivariate Bernoulli model is adopted.

In order to get features from our data sets, the algorithm split up each training example into sets of features. These features are tokens of air and bone conduction audiometry thresholds. Audiometry threshold is a point of a given level of frequency and decibel at which pure tone can be heard. Every training example is represented as a vector of tokens of 1s and 0s where 1 represents the token existing in the training example and 0 represents it absence. The classifier is trained by estimating from the training examples the class probabilities and the class conditional probabilities of a feature given the class for all classes and for all features (Tjoa and Trujillo, 2006). The performance of the classifier is measured by calculating the error rate. This is done through cross-validation technique; repeated random sub-sampling validation method that partitioned the data set into training set and validation set. The partition is done multiple times and the training examples in each partition are randomly selected. Several iterations are performed using each partition and the errors averaged (Maimon and Rokach, 2005).

FP-Growth and Bayes Classifier:

Extracting patterns in form of frequent item sets is one of the conventional way of finding interesting patterns in a data set(Cabena *et al*., 1997). The extent of relationship between items within an item set can be measured in terms of support and confidence. Support is the number of times a rule is applicable within a given dataset. An audiometry data ***di***contains item set ***S*** if ***S*** is a subset of ***di***It can be represented mathematically:

$$\sigma\,(\boldsymbol{S}) = |\{\boldsymbol{di}|\boldsymbol{S} \subseteq \boldsymbol{di},\, \boldsymbol{di} \in \boldsymbol{D}\}| \tag{1}$$

σ (***S***) denotes support for an item set ***S***. ***di*** refers to individual audiometry data with ***S*** as its subset ($\boldsymbol{S} \subseteq \boldsymbol{di}$) which means every item of ***S*** is also an item in ***di*** and ***di*** is an element of the dataset (***D***).

All the training examples ***di*** that have ***S*** as their subset forms the training data ***T***for the classification

$$\boldsymbol{T} = \sum (\boldsymbol{S} \subseteq \boldsymbol{di}) \tag{2}$$

And the vocabulary is the union of all item sets that meets the minimum threshold

$$\boldsymbol{V} = \bigcup \Big(\sum_{\delta\,(S)} (S) \Big) \tag{3}$$

Bayes rule can be mathematically represented as:

$$P(C \mid D) = \frac{P(D|C)P(C)}{P(D)} \tag{4}$$

That can be used in the classifier as:

$$C_{map} = \underset{c \in C}{\operatorname{argmax}} P(C/D) \tag{5}$$

The best class (C_{map}) is the one out of all classes that maximizes *(argmax)*$P(C/D)$. By Bayes rule which ever class maximizes equation (5), maximizes equation (6):

$$C_{map} = \underset{c \in C}{\operatorname{argmax}} P(D/C)P(C) \tag{6}$$

The most likely class will be the one that maximizes the product of $P(D/C)\ P(C)$. The aim is to look for a class which in this case is the symptom or group of symptoms whose probability is greater given certain audiometry thresholds. Equation (6) can be rewritten as:

$$Cmap = \underset{c \in C}{\operatorname{argmax}} P(x_1, x_2, x_3 \ldots\ldots\ldots x_n / C)\ P(C) \tag{7}$$

The joint probability of x_1 through x_n conditioned on a class can be represented as the product of independent probabilities $P(x_1/C) \bullet P(x_2/C) \bullet P(x_3/C) \bullet \ldots\ldots\ldots P(x_n/C)$. In order to compute the most likely class, the probability of the string of features (likelihood) is multiplied by the probability of a class (prior). It can be simplified and represented by:

$$C_{NB} = \underset{c \in C}{\operatorname{argmax}} P(C_j) \prod_{x \in X} P(x/C_j) \tag{8}$$

The best class (C_{NB}) is the class that maximizes the prior probability of the class $P(C_j)$ multiplied by every feature the probability of that feature or set of features given the class. The probability of the class given the position of every hearing threshold in the data is calculated and the highest probability assigned.

$$C_1 = P(C_1) \prod_{i} P(d_1/C_1) \tag{9}$$

$$C_2 = P(C_2) \prod_{i} P(d_2/C_2) \tag{10}$$

The algorithm can be improved by the use of logarithms to prevent underflow

$$C_{NB} = \underset{cj \in C}{\operatorname{argmax}}\ [\log P(C_j) + \sum_{i \in positions} \log P(x_i/C_j)] \tag{11}$$

To train the classifier, the frequencies in the data are calculated. For the prior probability of a training example (***t***) being in a class (C_j), the number of training examples in class (C_j) is divided by the count of all training examples.

$$\acute{P}(C_j) = \frac{tcount\ (c = Cj)}{Nt} \tag{12}$$

As for the likelihood in multinomial model, the probability of threshold of *i(t_i)* given class *(C_j)* can be determined by counting the number of times threshold of *i (t_i)* occurs in a training example of class *(C_j)* and dividing by the total number of the thresholds in all training examples of class *(C_j)*

$$\acute{P}(ti \mid cj) = \frac{count\ (ti,Cj)}{\sum count\ (t,Cj)} \tag{13}$$
w ∈ V

For the multivariate Bernoulli model, the fraction of training examples of class *(C_j)* in which the threshold appears is divided by the total number of training examples in class *(C_j)*.

$$\acute{P}(di \mid cj) = \frac{count\ (di,Cj)}{\sum count\ (d,Cj)} \tag{14}$$
w ∈ V

A test data may have some tokens which were not present in the training data sets.When a threshold value appears in a test set and does not occur in the training set, the maximum likelihood estimate for that threshold given a class will be 0. This is because this 0 probability can never be conditioned away. Therefore, the solution is additive smoothing.

$$\acute{P}(ti \mid c) = \frac{count\ (ti,C)+1}{\sum(count\ (t,C)+1)} \tag{15}$$
w ∈ V

$$\acute{P}(di \mid c) = \frac{count\ (di,C)+1}{\sum(count\ (d,C)+1)} \tag{16}$$
w ∈ V

The vocabulary size is added in equation (17) and rewritten as below. This is because 1 is added in equation (15) and (16) to every vocabulary item into the denominator:

$$\acute{P}(ti \mid c) = \frac{count\ (ti,C)+1}{(\sum count\ (t,C))+|V|} \tag{17}$$
w ∈ V

$$\acute{P}(di \mid c) = \frac{count\ (di,C)+1}{(\sum count\ (d,C))+|V|} \tag{18}$$

The calculation of these parameters for the prior can be represented as:

- From the union of all item sets that meets minimum threshold, extract vocabulary *(V)*

For each class, get the training examples that have that class

Calculate P(C_j) terms:

For each *Cj* in *C* do

Training examples *(t_j)* ← All training examples with class = C_j

$$\acute{P}(cj) = \frac{|\ tj\ |}{|\ Total\ \#\ training\ examples\ |}$$

The calculation of these parameters for the multinomial likelihood can be represented as:

- Calculate *P(t_k/C_j)* terms

Thresholdsj ←Single set containing union of all frequent items sets (vocabulary)

For each t_k in vocabulary

n_k ← # of occurrence of t_k in *training examples of Cj*

$$P(tk \mid cj) = \frac{nk + a}{n + a \mid vocabulary \mid}$$

The calculation of these parameters for the multivariate Bernoulli likelihood can be represented as:

- Calculate *P(d_k/C_j)* terms

Thresholdsj ←Single set containing union of all frequent items sets (vocabulary)

For each t_k in vocabulary

n_k ← # of training examples where t_k is present

$$P(dk \mid cj) = \frac{nk + a}{n + a \mid vocabulary \mid}$$

The alpha (α) is the smoothing parameter added to n_k

Results:

The importance of feature extraction techniques cannot be over emphasized. It is important in the success of many machine learning algorithms(Guyon *et al.*, 2006). The performances of the classifier using both multivariate Bernoulli and multinomial models with and without feature extraction are compared. This section presents the validation results of the proposed technique performances on data set with 242 training examples.

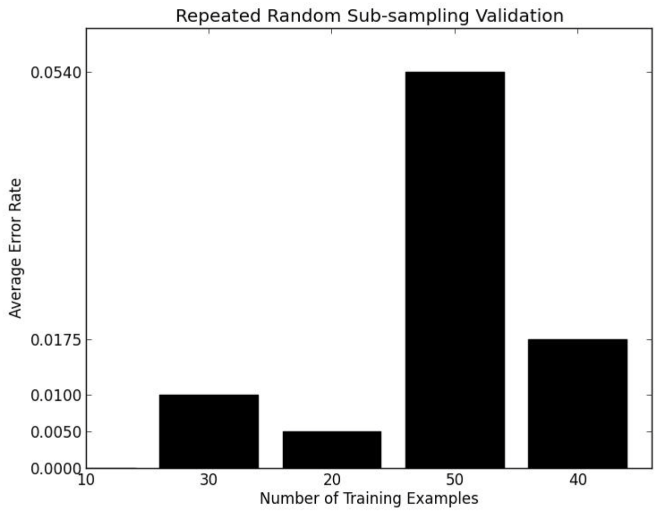

Fig. 1: Validation results using Bernoulli multivariate model with feature extraction

Figure 1 shows the average error rate from 10 iterations using 5 different partitions. There is 100% prediction accuracy using partition with 10 training examples, 99.5% accuracy with 20 training examples, 99% with 30 training examples, 98.25% with 40 training examples and 94.60% accuracy with 50 training examples. In other words, the average error rates for the 5 different partitions are, 0, 0.5, 1, 1.75 and 5.4% respectively. The classifier performs excellently with multivariate Bernoulli model with feature extraction.

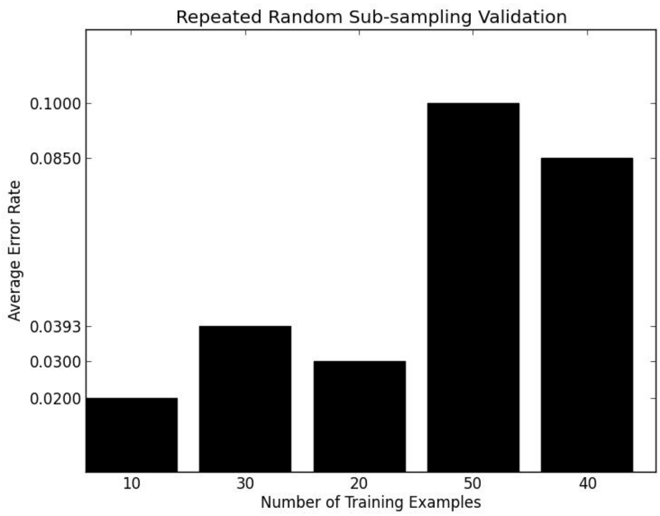

Fig. 2: Validation results using Multinomial model with feature extraction

Figure 2 visualizes the average error rate based on ten iterations using different partitions of validation set. This is the result from multinomial naïve Bayes model with feature extraction using FP-Growth algorithm. The partition with 50 training examples is the highest with 10% error rate averaged over 10 iterations. This means the prediction rate is 90% accurate. The lowest average error rate is 2% which is for partition with 10 training examples. This is followed by average error rate of 3%, 3.9% and 8.5% for 20, 30 and 40 partitions respectively.

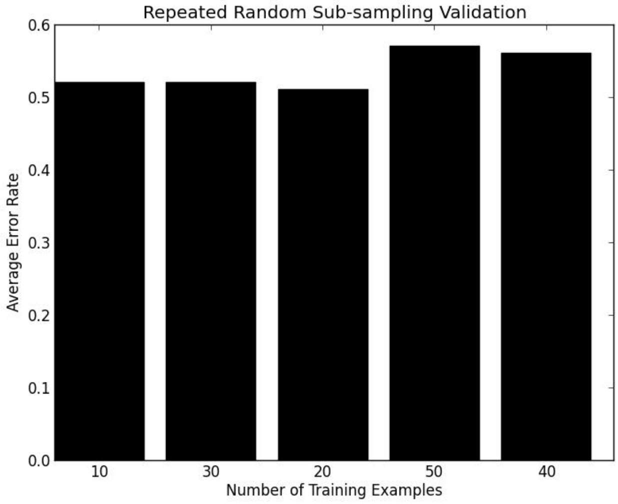

Fig. 3: Validation results using Multivariate Bernoulli model without feature extraction

Figure 3 depicts the validation results over 10 iterations using different partitions. The performance obtained from the classifier without feature extraction is contrary to the results obtained with feature extraction using both multivariate Bernoulli and multinomial models. The bar chart shows the average error rates using multivariate Bernoulli model without feature extraction. The average error rate for each partition is high. The partitions with 50 and 40 training examples have the worst average prediction inaccuracy with 57% and 56% average error rates respectively. The partitions with 10, 20 and 30 training examples have up to 50% error rate.

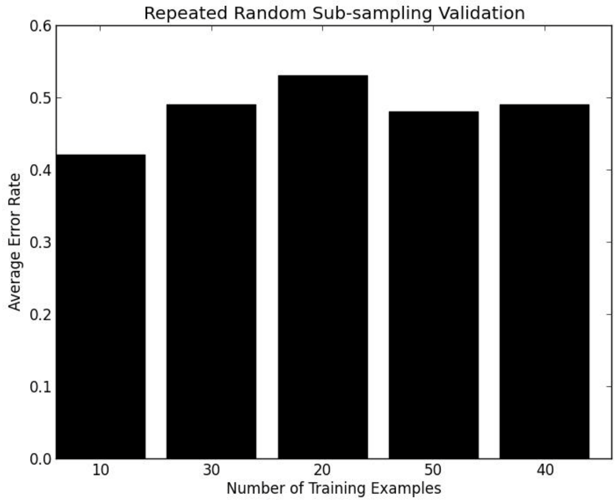

Fig. 4: Validation results using Multinomial model without feature extraction

Figure 4 shows the validation results using multinomial model without feature extraction. The partition with 10 training examples has the least average error rate of 42%. The highest is partition with 20 training examples having 53% average error rate. The partitions with 30, 40 and 50 training examples all have error rates of 48%. All the error rates were averaged over 10 iterations.

To summarise the above findings, the multivariate Bernoulli model with feature extraction gives the best resultsforall the 5 different partitions with average of error rate of 0, 0.5, 1, 1.75 and 5.4% for the 10, 20, 30, 40 and 50 partitions respectively when compared with the average error rates for multinomial model with feature extraction. The multinomial model has average error rates of 2, 3, 3.9, 8.5 and 10% for 10, 20, 30, 40 and 50 partitions respectively. However, there considerable higher average error rates estimation for both models without feature extraction. The multivariate Bernoulli average error rates over 10 iterations for the partition 10, 20, 30, 40 and 50 is 52, 50.5, 51.7, 56.3, 57.2% respectively. While that of multinomial model without feature extraction is 42, 53, 49, 48.5 and 48% for 5 different partitions of 10, 20, 30, 40 and 50 respectively.

Discussion and Conclusion:

The interesting relationships between items hidden in a data set can be discovered using the appropriate machine learning methods. These relationships may help in understanding how different items are connected with each other. This can results to discovery of new knowledge. Several studies have shown the existence of relationship between hearing thresholds and other attributes in patient’s medical record. Going by the findings in these studies, it can be justified that more relationships might exist between patients audiometry data and some attribute in the medical record. Earlier studies have used FP-Growth and association analysis algorithms to uncover relationship between hearing loss symptoms and audiometry thresholds using audiometric data of 50 hearing loss patients. Strong correlation has been found to exist between some pure-tone audiometry thresholds and the symptoms of Tinnitus, Giddiness and Vertigo. The findings for an extension of that study on a dataset with 399 audiometric data give similar results. This indicates that a connection might exist between these thresholds and those symptoms. Relevant patterns in form of item sets uncovered by FP-Growth algorithm have been used to form the naïve Bayes vocabulary. Only training examples that have a subset in the generated item sets become part of the training set. Extracting the vocabulary from the frequent item sets reduce the number features needed to form the vocabulary thereby decreasing the dimensionality of the data and minimizing storage requirements. This also increases the algorithm speed. In the experimental results, this has shown to improve the accuracy of the classifier. When 242 training examples were extracted from the 399 patients’ audiometric data to form the training set for the naïve Bayes classifier, the classes of symptoms discovered between the 242 training examples are Tinnitus, Tinnitus and Vertigo or the combination of Tinnitus, Vertigo and Giddiness. More data might reveal relationships between other types of symptoms and some hearing thresholds. The two models seem to perform better with the vocabulary extracted from patterns uncovered by FP-Growth algorithm. The multivariate Bernoulli model with feature extraction gives the best resultsforall the 5 different partitions with average of error ratesof 0, 0.5, 1, 1.75 and 5.4% for the 10, 20, 30, 40 and 50 partitions respectively over 10 iterations. However, the partition with 50 training examples and 5.4% average error rate can be considered more important because the more training examples in the validation set, the more acceptable the result will be. Increasing training data size will give more training examples that meet the minimum support threshold and thereby increasing the prediction accuracy. Future work will aim at increasing both the training data and the validation data. More data will be needed to validate these findings

Acknowledgement

The authors wish to thank the Department of Otolaryngology, Hospital Sultanah Fatimah, Muar, Johor for their support and cooperation and for making the sample medical audiology records available to us.

REFERENCES

Aggarwal, C.C., C. Zhai, (Eds.) 2012. Mining Text Data, 2012th ed. Springer, london.

Agrawal, Y., E.A. Platz, J.K. Niparko, 2009. 'Risk factors for hearing loss in US adults: data from the National Health and Nutrition Examination Survey', 1999 to 2002. Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol., 30: 139-145.

Ahmed, H.O., J.H. Dennis, O. Badran, M. Ismail, S.G. Ballal, A. Ashoor, D. Jerwood, 2001. 'High-frequency (10-18 kHz) hearing thresholds: reliability, and effects of age and occupational noise exposure', Occup. Med. Oxf. Engl., 51: 245-258.

Blanchet, C., C. Pommie, M. Mondain, C. Berr, D. Hillaire, J.-L. Puel, 2008. 'Pure-tone threshold description of an elderly French screened population', Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol., 29: 432-440.

Brant, L.J., J.L. Fozard, 1990. 'Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging', J. Acoust. Soc. Am., 88: 813-820.

Cabena, P., Hadjnian, Stadler, Verhees, Zanasi, 1997. Discovering Data Mining: From Concept to Implementation.Prentice Hall, USA.

De Heer, A.-M.R., P.L.M. Huygen, R.W.J. Collin, H. Kremer, C.W.R.J. Cremers, 2009. 'Mild and variable audiometric and vestibular features in a third DFNA15 family with a novel mutation in POU4F3', Ann. Otol. Rhinol. Laryngol., 118: 313-320.

Gates, G.A., P. Schmid, S.G. Kujawa, B. Nam, R. D’Agostino, 2000. 'Longitudinal threshold changes in older men with audiometric notches', Hear. Res., 141: 220-228.

Ghosh, A., R.K. De, S.K. Pal, (Eds.), 2008. 'Pattern Recognition and Machine Intelligence', Second International Conference, PReMI 2007, Kolkata, India, December 18-22, 2007, Proceedings, 2007th ed. Springer.

Guyon, I., S. Gunn, M. Nikravesh, L.A. Zadeh, (Eds.), 2006. Feature Extraction: Foundations and Applications, 2006th ed. Springer, Netherlands.

Hamelryck, T., K. Mardia, J. Ferkinghoff-Borg, (Eds.), 2012. Bayesian Methods in Structural Bioinformatics, 2012th ed. Springer, London.

Han, J., J. Pei, Y. Yin, R. Mao, 2004. 'Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach', Data Min KnowlDiscov, 8: 53-87.

Harrington, P., 2012. Machine Learning in Action. Manning Publications, USA.

He, Z., S. Deng, X. Xu, 2005. 'An FP-Tree Based Approach for Mining All Strongly Correlated Item Pairs', in: Hao, Y., Liu, J., Wang, Y., Cheung, Y., Yin, H., Jiao, L., Ma, J., Jiao, Y.-C. (Eds.), Computational Intelligence and Security, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp: 735-740.

Hwang, J.H., H.C. Ho, C.J. Hsu, W.S. Yang, T.C. Liu, 2008. 'Diagnostic value of combining bilateral electrocochleography results for unilateral Ménière’s disease', Audiol.Neurootol., 13: 365-369.

Jönsson, R., U. Rosenhall, I. Gause-Nilsson, B. Steen, 1998. 'Auditory function in 70- and 75-year-olds of four age cohorts.A cross-sectional and time-lag study of presbyacusis', Scand. Audiol., 27: 81-93.

Kim, S., E.J. Lim, H.S. Kim, J.H. Park, S.S. Jarng, S.H. Lee, 2010. 'Sex Differences in a Cross Sectional Study of Age-related Hearing Loss in Korean', Clin. Exp. Otorhinolaryngol., 3: 27-31.

Lee, F.-S., L.J. Matthews, J.R. Dubno, J.H. Mills, 2005. 'Longitudinal study of pure-tone thresholds in older persons', Ear Hear., 26: 1-11.

Lewis, D.D., 1998. 'Naive (Bayes) at forty: The independence assumption in information retrieval', in: Nédellec, C., Rouveirol, C. (Eds.), Machine Learning: ECML-98, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp: 4-15.

Maimon, O.Z., L. Rokach, 2005. The Data Mining and Knowledge Discovery Handbook. Springer, USA.

Noma, N.G., M.K. AbdGhani, 2012. 'Discovering pattern in medical audiology data with FP-growth algorithm', in: 2012 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES).Presented at the 2012 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), pp: 17-22.

Rajendran, P., M. Madheswaran, 2010. 'Hybrid Medical Image Classification Using Association Rule Mining with Decision Tree Algorithm', arXiv: 1001.3503.

Reynolds, G., 2011. Ethics in Information Technology, 4th ed. Cengage Learning, USA.

Sullivan, F., J. Wyatt, 2006. ABC of Health Informatics, 1st ed. BMJ Books, USA.

Tjoa, A.M., J. Trujillo, 2006. 'Data Warehousing and Knowledge Discovery', 8th International Conference, DaWaK 2006, Krakow, Poland, September 4-8, 2006, Proceedings.Springer.

Vaughan, N., K. James, D. McDermott, S. Griest, S. Fausti, 2006. 'A 5-year prospective study of diabetes and hearing loss in a veteran population', Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol., 27: 37-43.

Wiley, T.L., R. Chappell, L. Carmichael, D.M. Nondahl, K.J. Cruickshanks, 2008. 'Changes in hearing thresholds over 10 years in older adults', J. Am. Acad. Audiol., 19: 281-292; quiz 371.

Witten, I.H., E. Frank, M.A. Hall, 2011. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition, 3rd ed. Morgan Kaufmann, USA.