Data mining of audiology data to find patients who might benefit from ITE hearing aids or tinnitus maskers

Muhammad Naveed Anwar, Michael Philip Oakes

Department of Computing, Engineering & Technology, University of Sunderland, Sunderland, UK

(Naveed.Anwar@sunderland.ac.uk, Michael.Oakes@sunderland.ac.uk)

Abstract

We describe our work on the data mining of hearing aid patient data to answer the following research questions:

- Which factors influence the choice of ITE (in the ear) as opposed to BTE (behind the ear) hearing aids?
- For patients diagnosed with tinnitus, which factors influence the decision whether to fit a tinnitus masker?

Our data set is 180,000 patient records provided by the hearing aid clinic at James Cook University Hospital, Middlesbrough, UK. The records contain fixed vocabulary fields such as diagnosis, audiograms and short free text notes. The data set is unusual in that many of the patients were prescribed ITE hearing aids, which are not generally available on the National Health Service in the UK.

Using PCA (principal component analysis) and the chi-squared test, we found that flat hearing loss audiograms were associated with ITE aids and audiograms with air-bone gaps were associated with BTE aids. We also found that males tended to use ITE aids while females tended to use BTE aids. There was a positive association between ITE and age below 70. Patients with severe hearing loss tended to use BTE hearing aids while patients with a mild to moderate hearing loss tended to use ITE hearing aids. An analysis of the free text notes showed that ITE hearing aid types tended to use lacquer, had vents, required reshelling of ear impressions, had changes made to the hearing aid itself, were reviewed and the wearers were making progress.

There was no association found for gender and the use of tinnitus maskers. The keywords 'tinnitus' and 'masker' were found together in the free text notes significantly more often in patients with mild to moderate hearing losses, and significantly less often when the patient wore a BTE hearing aid or was aged 54 or less.

We combined our full set of data attributes (audiograms, gender, age, diagnosis and free text keywords), using logistic regression and a Naive Bayesian approach in two separate experiments for both of our two research questions. The resulting models can be used as the basis of decision support systems, where the inputs are hearing aid clinic data for new patients, and the outputs are either the estimated probability that the patient requires an ITE aid or that he or she requires a tinnitus masker.

Hypothesis and research question

Hypothesis:
The data mining of heterogeneous audiology data will enable us to discover new features and associations that will be useful for audiologists in their work. It will also help in developing a audiology decision support system.

Research questions:

- Which factors influence the choice of ITE (in the ear) as opposed to BTE (behind the ear) hearing aids?
- For patients diagnosed with tinnitus, which factors influence the decision whether to fit a tinnitus masker?

Clustering of audiograms by K-means – class exemplars

Free text:

	ac250	ac500	ac1K	ac2K	ac4K	ac8K
C1	73.66	73.00	74.99	80.48	91.08	108.21
C2	35.17	33.56	35.87	43.26	55.90	66.50

Diagnosis:

	ac250	ac500	ac1K	ac2K	ac4K	ac8K
C1	65.11	66.40	69.31	73.69	81.89	91.02
C2	21.88	18.39	17.83	20.87	34.85	42.94

Hearing aid type:

	ac250	ac500	ac1K	ac2K	ac4K	ac8K
C1	68.78	68.12	70.35	76.56	87.86	106.88
C2	36.98	35.77	39.10	48.44	61.20	72.13

Hearing aid types, tinnitus masker and free text keywords associated with clusters

Free text:

	Hearing aid type	
	Typical	Atypical
Cluster 1	BTE	ITE
Cluster 2	ITE	BTE

Tinnitus masker:

	Tinnitus masker	
	Typical	Atypical
Cluster 1	-	With-masker
Cluster 2	-	-

Hearing aid type:

	Hearing aid type	
	Typical	Atypical
Cluster 1	BTE	ITE
Cluster 2	ITE	BTE

Findings of Clustering of audiology data

- Patients with severe hearing loss group are associated with BTE hearing aid
- Patients with a mild to moderate hearing loss group are associated with ITE hearing aid type
- Patients with moderate to severe hearing loss group do not use maskers
- The mild to moderate hearing loss group were more concerned about tinnitus (ringing in the ears) than hearing loss
- High gain hearing aid types are associated with severe hearing loss group
- The atypical words "canc" (cancelled) and "dna" (did not attend) show that patients with severe hearing loss group were less likely to cancel (or simply fail to attend) their appointments

Principal Component Analysis (PCA) for audiograms

PCA 1 : Flat hearing loss

42 41 40 39 42 44
45 45 42 42 45

PCA 2 : Pure high tone sensorineural (inner ear) loss

37 38 48 69 82 79
46 46 55 76 89

PCA 3 : Air-bone gap (flat)

78 77 75 71 75 76
31 35 42 45 47

PCA 4 : Air-bone gap (predominant at low tone)

50 59 76 76 50 32
29 55 72 75 52

Findings of PCA

- Patients with flat hearing loss (PCA 1) audiograms were associated with ITE hearing aids
- Patients with flat audiograms with air-bone gaps (PCA 3) were associated with BTE hearing aids
- We did not find association of tinnitus with masker with any of the principal components

Discovery of associations with Chi-squared test

Categories with positive and negative keywords for ITE/BTE:

	Positive keywords (Typical)	Negative keywords (Atypical)
Age<=70	*Not found [DIAGNOSIS]	*Not found
Age>70	*Not found	*Not found
BTE	**mould, be34, map, gp, 92, audio, inf, be52, ref, staff, reqd, be36, contact [FREE TEXT]	*fta, reshel, appt, it, nn, nfa, 2001, rev, lacquer, hn, km, imp, review, 2000
ITE	**fta, reshel, appt, it, nn, nfa, 2001, rev, lacquer, hn, km, imp, review, 2000, nh, vent, progress, aid, dt, taken	**mould, be34, map, gp, 92, audio, inf, be52, ref, staff,reqd, be36, contact, tri, n, order
Male	***ITE[HEARING AID TYPE]	***BTE
Female	***BTE	***ITE
Not-tinnitus	***BTE[HEARING AID TYPE]	***ITE
Tinnitus	***Not found	***Not found

Findings of discovery of associations with Chi-squared test

- It was found that male patients tended more to use ITE hearing aids and females patients tended more to use BTE hearing aids
- The hearing aid types associated with BTE were those with high gain and had changes made to the ear mould
- ITE hearing aids types used lacquer, vents, required reshelling of ear impressions, had changes made to the hearing aid, were reviewed and the wearer were making progress
- It was also found that patients not having tinnitus were using BTE hearing aids

Logistic regression

- We used logistic regression to see which of the factors for right ear's age, gender, diagnosis, air conduction (AC250 to AC8000), and bone conduction (BC250 to BC4000) frequencies are predictive of 'behind the ear' (BTE) / 'in the ear' (ITE) hearing aids and of 'tinnitus with-masker' / 'tinnitus not-with-masker'
- The relative likelihoods of the patient needing an ITE or BTE aid are given by equation

$$\begin{aligned}\text{Log}[P(BTE/ITE)] = {} & 1.06 - 0.06(Age) + 0.00(Gender) \\ & - 0.06(Diagnosis) - 0.27(AC250) \\ & - 0.53(AC500) - 0.42(AC1000) \\ & - 0.15(AC2000) + 0.05(AC4000) \\ & - 0.04(AC8000) + 0.02(BC250) \\ & + 0.13(BC500) + 0.16(BC1000) \\ & + 0.21(BC2000) - 0.09(BC4000)\end{aligned}$$

- Then, we removed all the variables having P>0.05

Logistic regression – worked example

- The regression equation after removing the variables having P>0.05 is

$$\begin{aligned}\text{Log}[P(BTE/ITE)] = {} & 0.99 - 0.26(AC250) \\ & - 0.52(AC500) - 0.41(AC1000) \\ & - 0.14(AC2000) + 0.13(BC500) \\ & + 0.16(BC1000) + 0.20(BC2000) \\ & - 0.09(BC4000)\end{aligned}$$

- In a database record, AC250= 45, AC500=60, AC1000=70, AC2000=70, BC500=50, BC1000=65, BC2000=70, and BC4000=70
- So, we put AC250=1, AC500=2, AC1000=2, AC2000=2, BC500=0, BC1000=3, BC2000=3, and BC4000=3 from a record
- Now, Log[P(BTE/ITE)] = -0.06, which suggested behind the ear hearing aid type and in database record it was the same

Naïve Bayesian approach

- We also used Naïve Bayesian approach to see which of the factors for right ear's age, gender, diagnosis, air conduction (AC250 to AC8000), and bone conduction (BC250 to BC4000) frequencies are predictive of 'behind the ear' (BTE) / 'in the ear' (ITE) hearing aids and of 'tinnitus with-masker' / 'tinnitus not-with-masker'

Naïve Bayesian approach – worked example

- In a database record, AC250= 45, AC500=60, AC1000=70, AC2000=70, BC500=50, BC1000=65, BC2000=70, and BC4000=70
- So, we calculated likelihood ratios of BTE and ITE for all variables above and further calculated
  - for ITE: prior odds=1.15 and posterior odds=0.05,
  - for BTE: prior odds=0.87 and posterior odds=20.63
- This suggested BTE hearing aid as (posterior odds ITE) < (posterior odds BTE)

Results of decision support system (DSS)

Logistic regression:

	Human	
Machine	ITE	BTE
ITE	696	86
BTE	124	527

BTE:

Precision	0.81
Recall	0.86
F-score	0.87

ITE:

Precision	0.89
Recall	0.85
F-score	0.87

Naïve Bayesian:

	Human	
Machine	ITE	BTE
ITE	591	191
BTE	288	363

BTE:

Precision	0.56
Recall	0.66
F-score	0.60

ITE:

Precision	0.76
Recall	0.67
F-score	0.71

- Evaluated using an 80:20 (training data:testing data) split
- We had Insufficient data for tinnitus with-masker / not-with-masker