

Biometrics and Security Speaker Recognition

Jonas Marquardt und Maik Riestock
Otto von Guericke University Magdeburg

abstract
abstract
abstract
abstract
abstract

Categories and Subject Descriptors:

Additional Key Words and Phrases: Speaker Recognition, Audio Feature Extraction

1. MOTIVATION

Um Personen eindeutig zu identifizieren gibt es verschiedene Verfahren. Eines davon ist die Erkennung der Stimme. Als Aufnahmegerät ist ein handelsübliches Mikrophon ausreichend. In dieser Übung ging es darum, herauszufinden wie eine Stimmenerkennung umgesetzt wird und welche Eigenheiten dieses Verfahren mit sich bringt.

[George Doddington 1998]

2. HYKE PROJEKT

allgemein

- Bildungseinrichtung
- Rajasthan, NW India
- jede Schule hat 1-3 Lehrer
- Hauptquartier in Udaipur, 150 entfernt
- Hauptquartier in Udaipur, 150 entfernt

bisher

- visuelle Kontrolle der Anwesenheit
- 2 mal am Tag Bilder
- manuell jedes Bild verifiziert wurde
- kosten für angestellte
- jede Schule benötigt eine digitale Kamera

neu

- neuer Ansatz mit Stimmen-Erkennung
- da 75 % der Schulen eh schon ein Telefon haben
- automatische Verifikation

This report was created in the context of the course Biometrics and Security [BIOSEC] winter term 2014/15. This course was held by: Prof. Dr.-Ing. Jana Dittmann and Prof. Dr.-Ing. Claus Vielhauer; Research group Multimedia and Security, Otto-von-Guericke-University of Magdeburg, Germany. The course was supported by: Dr.-Ing. Christian Krtzer, M.Sc. Kun Qian

ergebnis

—95 % Erkennungsrate, state of the art

3. OUR APPROACH

Aufgabenstellung A

- Closed set speaker authentication on the Hyke speech database
- Compare the results achieved (in terms of authentication performance) to the results presented in [Azarias Reda 2011]
- A projection of the samples in your data set to the characters of 'Doddingtons Zoo'

def. speaker recognition [Beigi 2011] Speaker recognition, sometimes referred to as speaker biometrics, includes identification, verification (authentication), classification, and by extension, segmentation, tracking and detection of speakers. It is a generic term used for any procedure which involves knowledge of the identity of a person based on his/her voice.

Bei unseren Arbeitsschritten haben wir uns dabei an der von ... vorgestellten Prozedur gehalten

—kommt woher? [Prof. Dr. Jana Dittmann 2014]

—genaue Erklärung in den einzelnen Kapiteln

Für diese Aufgabe wurden uns die folgenden Programme bereitgestellt:

- AAFE, ist ein ... und wurde für die Feature Extraction verwendet
- WEKA, ist ein ... und wurde für die Re-post und Re-Klassifikation verwendet

4. DATABASE

- male, female
- recordings, length
- format, kompression?
- noise, telephone device

5. PRE-PROCESSING

In diesem Kapitel geht es um die Daten unserer Datenbank auf die folgenden Schritte vorzubereiten.

Da wir in unseren Ergebnissen am Ende einen möglichen Unterschied zwischen den Geschlechtern beobachten zu können... dafür wurde das Set mit beiden Geschlechtern, hier mixed set, aufgeteilt in zwei Sets mit ausschließlich Stimmen von weiblichen Sprechern, hier female set, und mit ausschließlich männlichen Sprechern, hier male set.

Außerdem benötigen die Klassifikatoren zwei Sets von Daten. Mit dem einen Set wird das Modell trainiert, hier train set, und mit dem anderen Set evaluiert, hier test set.

—Was haben wir gemacht?

—Warum?

—mit welchem ergebnis?

eine aufteilung wie folgt durchgenommen:

—mixed train set, beinhaltet ...?

—mixed test set

—female train set

—female test set

—male train set

diese aufteilung erlaubt es und zu einem die klassifikationen mit verschiedenen sets für trainings und evaluierung phase zu fñtern und zum anderen dient uns die aufteilung in mixed, female und male später einen möglichen unterschied der ergebnisse zu beobachten

6. FEATURE EXTRACTION

—Benutzung von AAFE

—Woher kommt er?

—Was macht er?

—Wie?

7. POST-PROCESSING

—Benutzung von WEKA

—Was haben wir gemacht?

—Warum?

—mit welchem ergebnis?

—Befehle aus weka?

8. KLASSIFIKATION

in diesem kapitel geht es darum klassifikatoren zu finden die bei der Klassifizierung geht es im allg um

Unsere daten sind dabei in die 6 sets unterteilen und beinhalten die feature der recordings.

Zur bestimmung der besten klassifizieren haben wir die methode des try and error verwendet. Das heißt wir haben alle möglichen klassifikatoren in den default setting auf unsere datenbasis angewandt und danach die ergebnisse verglichen.

ein gutes ergebnis bestand darin dass möglichst viele samples eines sprechers auch dem richtigen sprecher zugeordnet wurden.

Klassifizieren mit Settings:

—weka.classifiers.lazy.IBK -K 1 -W 0 -A

—weka.classifiers.trees.RandomForest -I 10 -K 0 -S 1

dabei hat sich eine klassifikator als besonders gut erwiesen, der ibk. Als vergleich haben wir das ergebnis mit dem ergebnis des klassifikators danebengestellt mit den zweit besten ergebnissen, der randomforest, zu sehen in Tabelle: I

Diese ergebnisse...

—Ansatz?

—Welche Classifier?

—Was sind das für Classifier?

—Anwendung mit train und test set

—results classifier

—Interpretation für auth

—results auth

—vergleich mit hyke

Klassifikator	Treffergenauigkeit
IBK	54.94%
RandomForest	39.14%

Table I.
Ergebnisse
der
Klassi-
fika-
tion

8.1 Authentifikation

verwendung der ergebnisse

9. DODDINGTONS ZOO

—Vorstellung des Zoos

—Anwendung

—Results

10. CONCLUSION

—past stuff

11. FUTURE WORK

—future stuff

REFERENCES

- Edward Cutrell Azarias Reda, Saurabh Panjwani. 2011. Hyke: A Low-cost Remote Attendance Tracking System for Developing Regions. *Networked System for Developing Regions* (2011).
- Homayoon Beigi. 2011. *Fundamentals of Speaker Recognition*. Springer Science+Business Media.
- Alvin Martin-Mark Przybicki Douglas Reynolds George Doddington, Walter Liggett. 1998. SHEEP, GOATS, LAMBS and WOLVES - A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation. *National Institute of Standards and Technology* (1998).
- Prof. Dr.-Ing. Claus Vielhauer Prof. Dr. Jana Dittmann. 2014. *Biometrics and Security - Lecture*. Faculty of Computer Science, Institute of Technical and Business Information Systems, Advanced Multimedia and Security Lab (AMSL).

List of Tables

APPENDIX

A. TASK DESCRIPTION

Run your prototype on the collected data and perform a performance evaluation with your prototype. The evaluation must include:

The evaluation must include:

- Closed set speaker authentication on the Hyke speech database
- Compare the results achieved (in terms of authentication performance) to the results presented in
- A projection of the samples in your data set to the characters of 'Doddingtons Zoo'