

זיהוי דיבור

מרצה: חגי אהרונוביץ'

Speech_biu@yahoo.com

www.cs.biu.ac.il/~aronowc/speech

הרצאה 1: רקע ומוטיבציה

- מבוא לזיהוי דיבור.
- שימושים שונים לזיהוי דיבור.
- תחומים משיקים לזיהוי דיבור.
- הקושי שבזיהוי דיבור.
- מבוא לזיהוי תבניות.
- עקרונות זיהוי דיבור.

מבוא לזיהוי דיבור

הגדרה: זיהוי דיבור (אוטומטי) הוא תהליך ממחושב הממיר קלט שהינו אות דיבור לפלט שהינו המידע הרלוונטי (המילולי) באות הדיבור.

לדוגמא:



,2 ,1
3
ניסיון

שימושים שונים לזיהוי דיבור

ממשק אדם-מכונה

- הכתבת מסמכים ל-PC.
- הפעלת מכשירי חשמל.
- משחקים.
- חיוג קולי.
- תחליף למערכות מענה קולי (AT&T).
- גישה למאגרי מידע (למשל נתוני טיסות, 144).
- גלישה באינטרנט.
- עזר לניווט ברכב.
- עזר למנתח.

שימושים נוספים לזיהוי דיבור

תחליף לפעילויות אנושיות

- תרגום סימולטני (עבור שיחות טלפוניות).
- האזנה אוטומטית לשיחות טלפון (משטרה, ארגוני ביון).

טלפונים סולאריים / מחשבי כף-יד

- חיוג קולי.
- הכתבת e-mail ו-SMS.
- תרגום קולי.
- גלישה באינטרנט.

תחומים משיקים לזיהוי דיבור

זיהוי דובר (speaker identification)

אימות דובר (speaker verification)

זיהוי שפה (language identification)

סנתוז דיבור (Text To Speech)

חיקוי דיבור (voice conversion)

זיהוי מבטא

ניקוי רעשים

גלאי שקר

זיהוי מצב רוח

תחומים משיקים נוספים

קידוד דיבור

דחיסת דיבור

הצפנת דיבור

גילוי דיבור

אחזור מידע

סיווג טקסט

זיהוי כתב יד

התאמת רצפים בביו-אינפורמטיקה

ראייה ממוחשבת

זיהוי פנים

פונמות

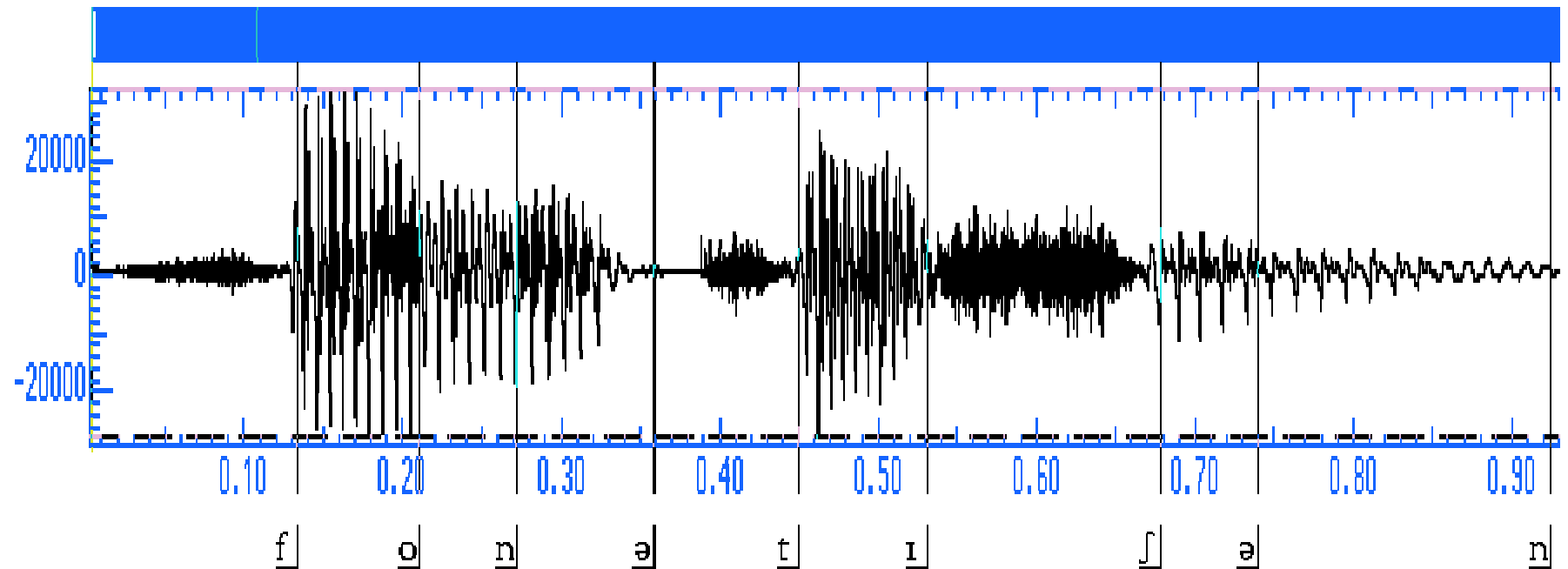
פונמה: יחידת הדיבור הבסיסית.

פונמות מורכבות מאוסף העיצורים והתנועות בשפה.

דוגמאות:

SHalom, shAlom, shaLom, shalOm, shaloM

הקושי שבזיהוי דיבור



phonetician



הקושי שבזיהוי דיבור

- זיהוי הגבולות בין הפונמות השונות.
- משך הזמן של הפונמות השונות אינו קבוע.
- אות הדיבור תלוי בפונמה, בדובר, במבטא, בהקשר, ברעש הסביבה, במצב הרוח ועוד.

ממדי קושי של זיהוי דיבור

- סוג דיבור
מילים מבודדות ← משפטים בודדים ← פסקאות שלמות
- סגנון דיבור
דיבור מוקרא ← דיבור זהיר ← דיבור ספונטני
- דוברים
דובר יחיד ← מספר דוברים מוגבל ← בלתי מוגבל
- אוצר מילים
עשרות מילים ← מאות מילים ← אלפי מילים ← עשרות אלפי מילים ← בלתי מוגבל
- דקדוק
מוגבל ע"י חוקים ← מוגבל חלקית ← לא מוגבל
- משימה
זיהוי מילולי ← הבנה (תגובה)

עוד ממדי קושי של זיהוי דיבור

- סביבה
חדר אקוסטי ← חדר שקט ← חדר עבודה ← רחוב ← מכונית
- הקלטה
מיקרופון צמוד לפה - איכותי ← מיקרופון שולחני איכותי ← טלפון
- זמן תגובה
שעות ← דקות ← זמן-אמת ← מהיר מזמן-אמת
- פלטפורמה
תחנת עבודה ← PC ← מכונית ← טלפון סלולארי ← בובת משחק
- מחיר
אלפי דולרים ← מאות דולרים ← דולרים בודדים

מבוא לזיהוי תבניות סטטיסטי

נתון

אוסף של טקסטים באנגלית ובצרפתית.

צ"ל

תוכנה היודעת להבדיל ברמה גבוהה בין טקסטים באנגלית ובצרפתית.

אלגוריתם מס' 1:

הרכב מילון אנגלי ומילון צרפתי.

בהינתן טקסט, הפרד את הטקסט למילים, וחפש כל מילה במילונים השונים.

ספור כמה מילים הן "אנגליות" וכמה "צרפתיות", והחלט עפ"י הרוב.

יתרונות: האלגוריתם יפתור היטב את הבעיה בדרך-כלל, אין צורך במידע מוקדם

בנוגע

לטקסטים.

חסרונות: יש צורך במילונים.

מה יקרה עבור מילים שלא מופיעות במילון?

מבוא לזיהוי תבניות סטטיסטי (2)

אלגוריתם מס' 2:

ניתוח תיאורטי:

עבור טקסט כלשהו, אנו רוצים לחשב את $\Pr(\text{French} | \text{Text})$ ואת $\Pr(\text{English} | \text{Text})$

ולהשוות בינם.

עפ"י חוק Bayes, מתקיים:

$$\Pr(\text{English} | \text{Text}) = \Pr(\text{Text} | \text{English}) * \Pr(\text{English}) / \Pr(\text{Text})$$

$$\Pr(\text{French} | \text{Text}) = \Pr(\text{Text} | \text{French}) * \Pr(\text{French}) / \Pr(\text{Text})$$

לפיכך, הבדיקה איזו מן ההסתברויות גדולות יותר שקולה לבדיקה האם יחס ההסתברויות

גדול מ-1:

$$\Pr(\text{English} | \text{Text}) / \Pr(\text{French} | \text{Text}) = \{ \Pr(\text{Text} | \text{English}) * \Pr(\text{English}) \} / \{ \Pr(\text{Text} | \text{French}) * \Pr(\text{French}) \}$$

מבוא לזיהוי תבניות סטטיסטי (3)

אלגוריתם מס' 2:

המשך הניתוח תיאורטי:

כעת, אנו רואים במשוואה את ההסתברויות האפרוריות של אנגלית וצרפתית, ואת ההסתברויות המותנות של הטקסט בהינתן השפות השונות.

$$\begin{aligned} P(\text{Text}|\text{English}) &= P(T_1 \dots T_n | \text{English}) \\ &= P(T_1|\text{English}) * P(T_2|T_1, \text{English}) * \dots * P(T_n|T_1 \dots T_{n-1}, \text{English}) \\ &\sim P(T_1|\text{English}) * P(T_2|\text{English}) * \dots * P(T_n|\text{English}) \end{aligned}$$

כלומר עפ"י ההנחה שהנחנו (הנחה מרקובית מסדר אפס), ההסתברות המותנית היא מכפלת

ההסתברויות של כל אחת מן האותיות בשפה.

מבוא לזיהוי תבניות סטטיסטי (4)

אלגוריתם מס' 2:

שלב האימון

סמן ידנית את שפתם של קבוצה קטנה של טקסטים (קבוצת אימון).
חשב את השכיחות של כל אחת מן האותיות הלטיניות עבור כל אחת מן השפות
מתוך קבוצת האימון (קיימות מספר שיטות).

שלב המבחן

בהינתן טקסט, חשב את ההסתברות שהוא בשפה האנגלית וחשב את
ההסתברות שהוא
בשפה הצרפתית.
בחר את השפה שלה ההסתברות הגבוהה ביותר.

יתרונות: האלגוריתם יפתור היטב את הבעיה בדרך-כלל, אין צורך מילונים.
חסרונות: יש צורך בסימון ידני של מס' טקסטים. (כמה?, האם הם מייצגים?)

מבוא לזיהוי תבניות סטטיסטי (5)

אלגוריתם מס' 3:

בדומה לאלגוריתם מס' 2, אך נעדן את ההנחה המרקובית מסדר אפס בהנחה מרקובית מסדר גבוה יותר.

יתרון יחסית לאלגוריתם מס' 2:

הנחה נכונה יותר.

חסרונות יחסית לאלגוריתם מס' 2:

דרושה קבוצה אימון גדולה יותר.
דרישות זכרון וכוח עיבוד גבוהות יותר.

מבוא לזיהוי תבניות סטטיסטי (6)

עקרונות זיהוי תבניות המשתקפים בדוגמאות שראינו

1. בחר ייצוג מתאים (אוסף מילים – אלגוריתם 1, אוסף אותיות – אלגוריתם 2,3).
2. **קדם עיבוד:**
פשט את הייצוג ככל האפשר (אותיות גדולות/קטנות, פונטים, מספרים, סימני פיסוק).
3. התאם מודל סטטיסטי לבעיה.
4. הנח הנחות מפשטות (מודל מרקובי באלגוריתמים 2,3).
5. **שלב האימון:**
חשב את הפרמטרים של המודל מתוך קבוצת אימון.
6. **שלב המבחן:**
בהינתן מודל תיאורטי והפרמטרים שחושבו, חשב את ההסתברות של כל אחת מן האלטרנטיבות בהינתן דוגמת המבחן, ובחר את זו בעלת ההסתברות הגבוהה ביותר.

עקרונות זיהוי הדיבור

עקרונות זיהוי התבניות מיושמות לבעיות רבות:

- זיהוי דיבור
- זיהוי כתב יד
- חיזוי ביצועי מניות
- חיזוי תפקודי חלבונים בתא.
- זיהוי פנים
- ועוד...

ההבדל העיקרי בין האפליקציות השונות הוא היישום הפרטני של העקרונות.

בהמשך נראה כיצד מיושמים העקרונות השונים בזיהוי דיבור:

כיצד יש לייצג את אות הדיבור?

איך ניתן לפשט את הייצוג?

אילו מודלים סטטיסטיים מתאימים לדיבור?

כיצד מאמנים את הפרמטרים של המודל?

כיצד מחשבים את ההסתברות של האלטרנטיבות השונות בהינתן המודל?

מה נלמד בקורס?

1. מודל פיזיקלי ליצירת קול ודיבור.
2. מנגנון השמיעה האנושי.
3. זיהוי תבניות.
4. חילוץ מאפיינים (feature extraction).
5. השוואת תבניות ע"י תכנות דינאמי.
6. מודלים מרקובים חבויים.
7. מודל שפה.
8. שיטות אימון:
 - Maximum Likelihood
 - Gradient Descent
 - Maximum Aopriori

ועוד...

הדגמה: MIPAD

מטלות

1. תרגילים תיאורטיים ומעשיים.
2. פרויקט / מבחן.

תרגיל מס' 1: (הגשה עד תאריך 11.11.2002)

פיצוח צופן (פשוט) באמצעות מודל שפה (פשוט).

תרגיל מס' 1

(הגשה עד תאריך 11.11.2002, בזוגות)

פיצוח צופן (פשוט) באמצעות מודל שפה (פשוט).

ב- www.cs.biu.ac.il/~aronowc/speech/ex1 ישנם הקבצים הבאים:

World192.txt: קובץ ארוך לאימון מודל שפה.

???.txt: קבצי טקסט מוצפנים (אחד לכל קבוצה)

Decode.cpp: פונקציית C לפענוח קובץ מוצפן.

המטרה: יש לכתוב תוכנית אשר מפענחת את הטקסט המוצפן (יש למצוא את

מפתח

ההצפנה).

```

void decode(char *encoded_buffer, char *decoded_buffer, int count, int key)
{
    int k[5];
    for(int j=0; j<5; j++)
        k[j] = key / (int) pow(10, 4-j) % 10;
    // decode
    for(int i=0; i<count; i++) {
        int index = i % 5;
        int dc = encoded_buffer[i] - k[index];
        if (dc < 0)
            dc += 256;
        decoded_buffer[i] = dc;
    }
}

```

מה מקבל כל זוג?

מס' סידורי של הקובץ המוצפן שאותו הוא צריך לפצח.
על הזוג לכתוב תוכנית שתפצח (או תעזור לפצח) את הקובץ המוצפן.

יש להגיש (ב-email):

1. המס' הסידורי של הקובץ המוצפן.
2. המפתח שגיליתם.
3. הסבר קצר על מה שעשיתם.
4. התוכנית שכתבתם.

מודלים מרקובים חבויים

תיאור הבעיה:

נמצאים בקזינו ורוצים להמר על תוצאת הטלת קוביה (נניח "1").
ידוע שהקזינו משתמש לעיתים בקוביה חוקית, ולעיתים בקוביה לא-חוקית.
מה לעשות?

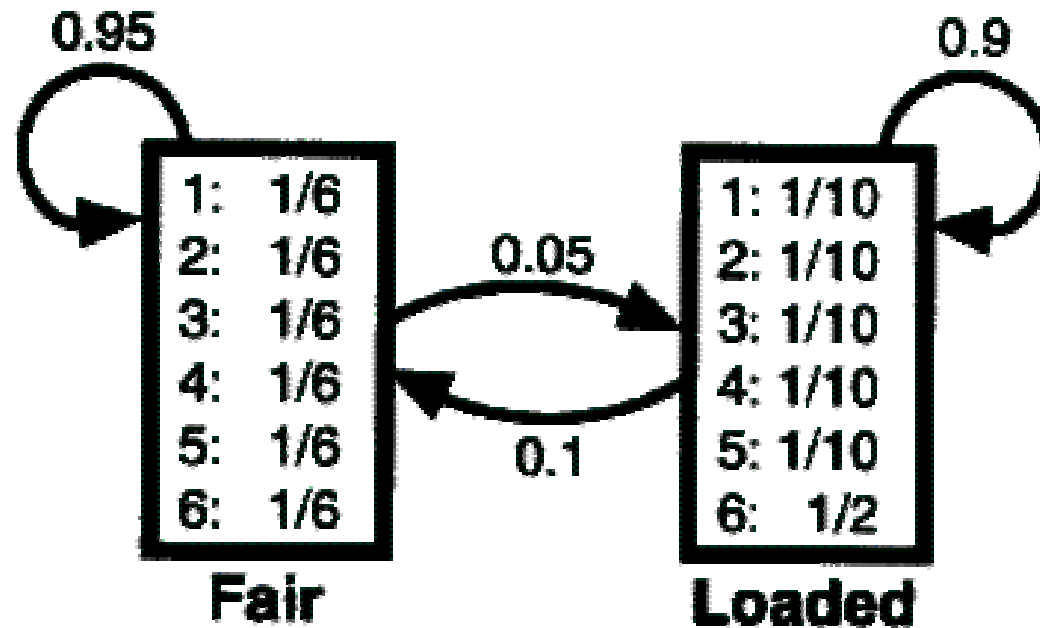
פתרון (אפשרי): מודלים מרקובים חבויים (חבויים כי לא ידוע האם הקוביה חוקית).

נניח מס' הנחות:

- מרחב האפשרויות: קוביה חוקית (סוג א') וקוביה לא חוקית (סוג ב').
- עבור קוביה מסוג מסוים, אין תלות בין הטלות שונות.
- הקזינו לא מחליף קוביה בכל הטלה, אלא נוטה לשחק זמן מה עם כל קוביה.

מודלים מרקוביים חבויים

מדריך למהמר



מודלים מרקוביים חבויים

האסטרטגיה:

1. **שערך (estimate) את הפרמטרים (הסתברויות) של המודל המרקובי החבוי.**
2. **בהינתן תצפיות (observations) – תוצאות הטלות קוביה, הערך את הסיכוי שהקוביה הנוכחית היא חוקית / איננה חוקית.**
3. **שקלל את ההסתברויות עם שיקולים אישיים כגון "שנאת סיכון".**
4. **קבל החלטה.**

חזרה

מודל שפה

נניח כי אלגוריתם זיהוי דיבור מתלבט בין שני המשפטים הבאים
(מבחינת התאמה אקוסטית לקלט).

1. אני הלכתי לישון.

2. אני גילחתי לישון.

המשפט הראשון תקין מבחינה תחבירית וסמנטית.
המשפט השני איננו תקין.

שילוב אינפורמציה כזו מאפשר לשפר את הדיוק של הזיהוי (וגם את מהירות
החיפוש).

חזרה

שיטות אימון

נחזור לבעיית הקזינו.

איך אפשר ע"י ניתוח מספר (סופי) של הטלות קוביה לשערך את הפרמטרים

של המודל המרקובי החבוי?

לשם כך צריך שיטת אימון כלשהי.

האסטרטגיה הכללית:

לאוסף של הפרמטרים המגדיר את כל הפרמטרים של המודל המרקובי החבוי נקרא

בקיצור "מודל".

כל מודל אפשרי הוא פתרון לגיטימי.

נרצה **לחפש** מודל אשר ייתן פתרון **טוב** ככל שנוכל.

שיטות האימון שונות זו מזו בהגדרת הקריטריון לטיב הפתרון ובדרכים השונות לחפשו.