

РЕЦЕНЗИЯ

на дисертационен труд
за придобиване на образователната и научна степен "Доктор" в
област на висше образование – 5. Технически науки
професионално направление – 5.3. Комуникационна и компютърна техника
докторска програма – "Компютърни системи, комплекси и мрежи"

Автор: инж. Георги Спасов Цанев, докторант в катедра КСТ на ТУ – Габрово

Тема: Изследване на методи за гласово управление

Рецензент: доц. д-р инж. Росен Стефанов Иванов, катедра КСТ на ТУ – Габрово

I. Тема и актуалност на дисертационния труд

Обект на изследване са методи и алгоритми, използвани при системи за управление чрез гласови команди с ограничен размер на речника. Постиженията в областта на дълбочинното обучение доведе до възможността системите за автоматично разпознаване на говор (ASR) да получат много широко разпространение през последните няколко години. Въпреки това остават множество нерешени проблеми. Работи се основно в областта на разработване на нови методи за по-ефективно обучение на многослойни изкуствени невронни мрежи (ANN). Повечето ASR системи не функционират надеждно при високо ниво на шума ($SNR < 5$ dB), особено ако той съдържа странични разговори. Малка част от системите могат да функционират при използването им от потребители със специфичен акцент, диалект или при използване на паразитни думи. Всички системи са все още чувствителни към пола и годините на потребителите. Темата е актуална и има своята както научна, така и практическа стойност.

II. Обзор на цитираната литература

Анализът на проблемите в областта на управлението чрез гласови команди е реализиран на базата на 164 литературни източници. От тях 31.7% (52) са публикувани през последните 5 години, 59.8% (98) са статии в чуждестранни списания, а останалите - доклади на международни конференции и книги. Липсват цитирания на български автори. В резултат на направеното литературно проучване докторантът е дефинирал следната цел на дисертационния труд: „Разработване на полиморфен модел на система за гласово управление”. За постигане на целта докторантът е работил по следните задачи:

1. Анализ на съществуващите методи за анализ.
2. Разработване на методи за класификация на речеви сигнали.
3. Създаване на генеративен модел на система за гласово управление с вариране на класификатора.
4. Разработване на алгоритми, които намаляват времето за обучение на многослойни машини на Болцман.
5. Разработване на адаптивни алгоритми за обучение на многослойни машини на Болцман.

III. Методика на изследване

Темата на дисертационния труд, както и поставената цел предполагат комплексни знания в областта на цифровата обработка на едномерни сигнали, изчислителния изкуствен интелект (обучение и разпознаване на образи чрез изкуствени невронни мрежи, генетични алгоритми), числения анализ и компютърно моделиране.

Методиката на изследване произтича от формулираните задачи:

- Създаване на бази с думи, необходими за тестване на анализирани системи за разпознаване на речеви команди.
- Разработване на алгоритми за обучение на многослойни ANN от тип машина на Болцман с цел минимизиране на грешката при обучение, както и времето за обучение.
- Експериментирание и анализиране на предложените алгоритми.

Представеният за рецензиране дисертационен труд е с обем от 186 страници от които 18 страници приложения. Изследванията в дисертационния труд са структурирани в 5 глави:

Глава 1. „Съвременни тенденции в развитието на системите за гласово управление”

В тази глава авторът достатъчно подробно описва областите на приложение на гласовото управление и достъпните за него съществуващи системи. Направен е и критичен анализ на нерешените проблеми в областта на автоматичното разпознаване на говор.

Глава 2. „Методи за класификация на данни”

В тази глава се описват едни от най-често използваните класификатори, които са приложими и при системите за управление с глас – k-means, DTW, SVM, VQ и ANN.

Главата трябва да завършва с изводи, които основно анализират приложимостта на разгледаните класификатори в системите за разпознаване на речеви команди, техните предимства и недостатъци.

Глава 3. „Генеративен модел на система за гласово управление със стохастичен класификатор”

В тази глава се описват основните модули на система за разпознаване на гласови команди. За класификатор е използвана генеративна стохастична ANN от тип ограничена машина на Болцман.

В секция 3.3 се предлага намаляване на нивото на шума чрез филтър на Винер. Не става ясно при какви видове шум е тестван филтърът и при какви стойности на SNR. От текста се остава с впечатление, че филтърът зависи от VAD, описан в секция 3.2. На практика първо се реализира намаляване на нивото на шума, а след това се активира VAD. Реализирани са експерименти с база от 32 думи. Базата е създадена от 21 участници в експеримента – 15 мъже и 6 жени. Всички експерименти са в лабораторна среда.

Глава 4. „Класифициране на речеви сигнали посредством еталон”

В тази глава са описани модулите на система за разпознаване на речеви команди посредством еталон. Системата използва алгоритъм за класификация, който е вариант на Genetic Time Wrapping (GTW) алгоритмите. Предложен е и класификатор на базата на логаритмуваната енергия на речевия сигнал в честотната област, използващ метода на крайните елементи. Предложена е дробно-квадратична прозоречна функция с цел намаляване на спектралното разсейване, поради обработката на речевия сигнал на ниво сегменти с краен брой дискрети. От представените времеви и честотни характеристики може да се заключи, че функцията може да се използва като прозоречна. Не е анализирана фазовата характеристика на функцията и закъснението което тя внася в системата. Дефинираното предимството, че предложената тегловна функция ще намали времето за предварителна обработка на речевия сигнал не е значимо, тъй като при системите за разпознаване на говор в реално време прозоречните функции се описват таблично. Част от резултатите, показани на Фиг. 4.6.1, са малко вероятни, като се има предвид, че влиянието на прозоречната функция и честотната скала (Mel, Bark или ERB) не влияят толкова силно на грешката при разпознаване, особено при ниско ниво на шума.

Глава 5. „Класифициране на речеви сигнали посредством многосвързана машина на Болцман”

Предложен е алгоритъм за обучение на многослойна ANN от тип машина на Болцман. Проведен е експеримент с цел доказване на приложимостта на предложения алгоритъм. Алгоритъмът използва интуитивен подход за промяна на стойността на скоростта на обучение на ANN в зависимост от грешката от реконструкция – ако има тенденция на намаляване на грешката, то скоростта се увеличава, а ако грешката нараства – скоростта се намалява и ѝ се променя знака. Предложен е двуточков градиентен метод за решаване на задачата за средно-полева апроксимация при обучение на ANN от тип машина на Болцман.

IV. Приноси на дисертационния труд

Постигнатите най-важни резултати, получени в дисертационния труд, имат научно-приложен и приложен характер. По моя преценка тези приноси могат да се формулират по следния начин:

1. Научно-приложни приноси:

- Предложен е класификатор на базата на логаритмуваната енергия на речевия сигнал в честотната област, използващ метода на крайните елементи. Класификацията се реализира на базата на функционал на грешката, който използва квази-Ермитов тип интерполация.
- Предложен е двуточков градиентен метод за решаване на задачата за средно-полева апроксимация при обучение на ANN от тип машина на Болцман.
- Предложен е алгоритъм за класификация по еталон, който е вариант на Genetic Time Wrapping (GTW) алгоритмите.

- Предложена е дробно-квадратична функция, която може да бъде използвана за намаляване на спектралното разсейване, поради обработката на речевия сигнал на сегменти с крайна дължина.

2. Приложни:

- Създадени са бази с думи, необходими за обучение и тестване на системи за разпознаване на речеви команди.
- Разработен е програмен код на Matlab™ с цел експериментиране на предложените алгоритми.

V. Публикации и цитирания на публикации по дисертационния труд

Авторът е представил списък на 7 научни публикации. Всички публикации са свързани с тематиката на дисертацията. От тях 3 са статии в чуждестранни списания, 1 статия в българско списание и 3 доклада на международни конференции в чужбина. Считам, че публикациите на докторанта по дисертацията съдържат основните приноси, които съм дефинирал. Това съответства на изискванията на Закона за развитие на академичния състав и на Правилника за неговото приложение за публикуване на най-съществените части от дисертационния труд.

Няма публикации с импакт фактор и импакт ранг. Липсват данни за цитиране на публикациите, представени по дисертационния труд.

VI. Авторство на получените резултати

Докторантът има 1 самостоятелна публикация. В три от публикациите е втори, в две – трети и в една – четвърти съавтор. От представените публикации по дисертационния труд оставам с впечатлението, че публикациите са дело на докторанта, под ръководството на неговите ръководители.

VII. Автореферат и авторска справка

Авторефератът отразява коректно съдържанието на дисертационния труд и приносите на докторанта. Обемът на автореферата е ненужно голям (59 страници), което представлява 35% от дисертационния труд. Авторефератът трябва да включва само математическия апарат и резултатите от експериментите, които доказват претенциите на автора.

VIII. Забележки по дисертационния труд

По време на обсъждането на дисертацията в Първичното звено, както и по време на подготовката на предварителната рецензия проведох разговори с докторанта. Същият е отстранил повечето мои забележки и е взел предвид направените от мен препоръки. Въпреки това, имам критични бележки към окончателния текст на дисертационния труд, по-важните от които са следните:

1. В Глава 3 експериментите се провеждат чрез система (Фиг. 3.4.1), която включва модули „Намаляване на шума” и VAD. На практика

- експериментите са реализирани в изолирана среда. От изложението не става ясно как тези два модула функционират.
2. Не е ясно защо базите с думи, използвани за обучение и тестване на разработените алгоритми, съдържат думи на английски, а са създадени и тествани само от българи. Самите бази трябва да са балансирани по пол на участниците и по брой записи на всеки участник. По-добрият вариант е да се използва съществуваща база с думи. Така характеристиките на предложената система може лесно да се сравнят с характеристиките на подобни системи, които използват с цел обучение и тестване същата база.
 3. Предполагам, че под „полиморфен“ модел за гласово управление авторът приема, че се избира класификатор, който дава най-добри резултати при конкретни входни данни. В дисертацията обаче не е изяснен въпроса как в реално време се подбира най-подходящият класификатор.
 4. Представеният Алгоритъм 5.2.1 (стр. 121) за обучение на машина на Болцман използва интуитивна промяна на скоростта на обучение. Обучението е с фиксиран брой на циклите. Промяната на стойността на λ при тенденция на увеличаване на грешката ($-\epsilon\lambda$) води до много силно намаляване на скоростта на обучение и има вероятност λ да попадне извън границите на зададения интервал. Избраните стойности на параметрите от които зависи скоростта на обучение и интервала в който попада тази скорост са подбрани в зависимост от конкретната структура на ANN и броя и типа на обучаващите вектори. При промяна на архитектурата на ANN и обучаващите вектори ще се наложи промяна на тези стойности.
 5. Описанието на експериментите може да бъде по-ясно, а част от получените резултати трябва да се анализират по-подробно.

IX. Заключение

Темата на дисертационния труд е актуална. Обемът на дисертационния труд и получените резултати са достатъчни за получаване на образователна и научна степен „Доктор“. Спазени са законовите изисквания от гледна точка на процедурите по докторантурата. Дисертационният труд отговаря на изисквания за получаване на образователната и научна степен „доктор“, заложили в Правилника за придобиване на научни степени и заемане на академични длъжности в Технически университет – Габрово.

Считам, че представеният дисертационен труд отговаря на изискванията на Закона за развитие на академичния състав в Република България. Постигнатите резултати ми дават основание да предложа да бъде придобита образователната и научна степен „Доктор“ от инж. Георги Спасов Цанев в област на висше образование – 5. Технически науки, професионално направление – 5.3. Комуникационна и компютърна техника, докторска програма – "Компютърни системи, комплекси и мрежи".

31.05.2018 г.
гр. Габрово

Рецензент: .

**Заличено обстоятелство,
на основание чл.2 от ЗЗЛД**