Stagii de practică în cercetare – Laboratorul SpeeD

Laboratorul de cercetare SpeeD, condus de Prof. Corneliu Burileanu propune studenţilor din anul III (seriile A, B, C şi G) o serie de stagii de practică în cercetare finanţate prin programul CASIA (http://casia.pub.ro/stagii-de-practica).

Stagiile de practică presupun lucrul la o temă de cercetare în tehnologia limbajului vorbit, timp de 3 luni (iunie-august 2013), în regim part-time (6 ore pe zi), într-un laborator de cercetare. Proiectele de practică se vor realiza sub îndrumarea lect. Andi Buzo, lect. Horia Cucu şi drd. Dragoş Drăghicescu şi se vor finaliza prin realizarea unei aplicaţii practice şi scrierea unui raport de cercetare (referat de practică).

Avantajele stagiilor de practică din cadrul SpeeD

- oportunitatea de a contribui la dezvoltarea unei aplicaţii în tehnologia limbajului vorbit sub îndrumarea unui grup de specialişti,
- posibilitatea utilizării (sau extinderii) raportului de practică pentru lucrarea de licenţă,
- subvenţie de 1100RON (platibilă în octombrie-noiembrie 2013 prin programul CASIA),
- cazare gratuită în cămin pe toată perioada stagiului (prin programul CASIA),
- un premiu de 2200RON pentru cel mai bun referat de practică (prin programul CASIA).

Temele de proiect propuse sunt enumerate la finalul acestui document.

Pe data de 29 ianuarie, ora 14:00, în Sala de Calculatoare nr. 3, le propunem studenţilor interesaţi o întâlnire, cu scopul a discuta despre aceste stagii de practică şi a le răspunde la eventualele întrebări.

Înscrierea la stagiile de practică se va face în perioada 28 ianuarie - 1 februarie, prin email la adresa andi.buzo@upb.ro, specificând:

- numele şi grupa din care faceţi parte,
- primele trei teme la care aţi vrea să participaţi (în ordinea preferinţei),
- disponibilitatea de a realiza o altă temă dintre cele propuse (în cazul în care tema preferată nu va mai fi disponibilă).

Numărul de stagii de practică finanţate prin proiectul CASIA este limitat (maxim 10), iar selecţia studenţilor interesaţi nu va mai fi făcută în urma workshop-urilor SpeeD din primăvara 2013, ci în perioada 4-11 februarie 2013.

Teme de practică propuse

1. Multi-lingual speech recognition (2 students)

   One of the new challenges in ASR[1] is dealing with multiple languages input. This condition has become extremely important for applications that target multi-lingual communities like in touristic centers, libraries, etc. One way of overcoming this issue is by implementing a Language IDentification (LID) component whose output is used to load previously trained acoustic and language models of the identified language. Another

way is by building multi-lingual acoustic and language models by using heterogeneous speech data in training.

The objective of the project is to implement both solutions and compare their performance from the accuracy and implementation complexity point of view.

2. Noise-robust voice features for ASR (1 student)

ASR systems use voice features extracted from the speech signal to transform the speech into text. State-of-the-art voice features (MFCCs, PLPs, etc.) cannot cope with additive noise in the speech signal and therefore the ASR systems performance is significantly degraded in this case. The recently proposed PNCC features seem to perform very well in the presence of noise and therefore PNCC-based ASR systems are less sensitive to additive noise.

This project aims to deeply study the PNCC features and to deploy a Java implementation of these voice features in the CMU Sphinx speech recognition toolkit.

3. ASR noise reduction techniques (1 student)

The noise present in speech recordings reduces drastically the accuracy of ASR because it creates a mismatch between the training data (usually clean speech data) and the testing data (affected by noise). Therefore, efforts are being made in order to remove the noise from the speech recordings. This process is also known as speech enhancement. There are several techniques that obtain noise reduction such as spectral subtraction, spectral estimation or uncertainty based methods.

The project consists in studying, understanding and implementing the most representative method from each category in Java. The methods will be compared by the ASR accuracy metric. The speech enhancement component will be integrated as a Data Processing component in the CMU Sphinx toolkit.

4. Spoken Web Search system (2 students)

Spoken Web Search means searching for spoken content within a speech database by using a spoken query. It makes sense for under-resourced languages for which no phonetic dictionaries are available. This problem can be approached by performing ASR for both query and speech database. Once the speech data is converted into text, the problem becomes a text searching one. The main difficulty is created by the imperfect accuracy of the ASR. Hence, one have to deal with searching an approximate text query into an approximate text database. It is obvious that a higher accuracy of ASR yields to higher search performance.

The project aims at building accurate acoustic models for under-resourced languages and providing efficient searching algorithms in approximate text databases. The evaluation of the methods will be made on the MediaEval databases.

5. Unsupervised speaker adaptation for ASR (1 student)

Inter-speaker variability is an important issue in Automatic Speech Recognition. ASR systems trained (or adapted) to recognize a single speaker (speaker-dependent ASR) have significantly better performances than ASR systems which aim at recognizing any voice. However, speaker-independent ASR systems can be backed-up with an

unsupervised, online speaker adaptation module to overcome the performance degrading caused by the inter-speaker variability.

This project requires an in-depth study of unsupervised speaker adaptation techniques and a Java implementation of such a module for an already existing speaker-independent ASR system for Romanian.

6. Unsupervised ASR domain (language model) adaptation (1 student)

Domain variability is an important issue in Automatic Speech Recognition. ASR systems designed to recognize speech from a single domain (sport news, travel dialogues, scientific speech, etc.) have significantly better performances than ASR systems which aim at recognizing speech from potentially any domain. However, general ASR systems can be backed-up with an unsupervised, online domain-adaptation module to overcome the performance degrading caused by domain variability.

This project requires an in-depth study of unsupervised domain adaptation techniques and a Java implementation of such a module for an already existing general ASR system for Romanian.

7. ASR output text restoration for increased readability (2 students)

A general ASR system has the purpose of converting the input speech signal into text. When the output text contains only a few words it is human intelligible, but when the ASR output has several pages of un-diacriticized, lowercase, punctuation-lacking text, its readability becomes a real issue.

The goal of this project is to develop an NLP (Natural Language Processing) module with the purpose of increasing the readability and eliminating the ambiguity of a long ASR output text. This NLP module should address: a) diacritics restoration, b) re-capitalization, c) text-to-digits conversion (for numbers) and d) punctuation restoration.

8. Speaker verification/identification system (1 student)

Speaker recognition has two main applications: speaker verification and speaker identification. Speaker verification means deciding whether a speech utterance belongs to a claimed speaker or not. Speaker identification means finding to which speaker belongs a given utterance. In each case the speakers must be modeled by using statistical models. The most common techniques are based on Gaussian Mixture Models (GMM). Common speech features like MFCC, PLP, etc., are used for modeling the speakers.

The objective of the project is to build a complete speaker identification system based on GMMs. It implies studying the GMMs and implementing the training and the classification algorithms.

9. Speech Similarity and Hyperlinking (2 students)

Recently the size of speech databases has increased enormously making searching in such databases very difficult. Unless the audio recordings are associated detailed metadata (relevant title, description, etc.), the searching must be done manually by listening every recording. An alternative is to classify audio recordings based on their speech content similarity. Keywords can be defined and hyperlinks can be build to create connections similar to web navigation. One possible way of obtaining this is by

using ASR. Once converted to text a similarity score can be assigned to each pair of recordings and hyperlinks can be determined.

The project aims at providing algorithms for determining the similarity in speech content between two audio recordings. The evaluation of the methods will be made on the MediaEval databases.

10. ASR errors analysis and correction (1 student)

State-of-the-art general ASR systems are far from being perfect. The average error rate on a large vocabulary speech recognition task is about 15-20%. In this context ASR error analysis is an important field of study which can bring valuable information about specific systematical errors that can be automatically corrected. Moreover, an important question, regarding the potential benefit of using human corrected ASR output to create automatic correction modules, is still unanswered in the scientific literature.

The goal of this project is to analyze and study the errors performed by an already existing ASR system for Romanian and propose new automatic correction techniques.

11. Automatic phonetization tool for dynamic-vocabulary ASR (1 student)

The acoustic model inside an ASR system does not model the words of the language, but their composing acoustic units: the phonemes. Consequently, any ASR system needs a phonetic dictionary to specify the words pronunciation (the phoneme sequence composing the word). Manual phonetization of words is only feasible for small vocabulary ASR systems, while for large vocabulary ASR systems special automatic phonetization systems have to be developed. Moreover, in dynamic-vocabulary ASR systems the words need to be phonetized on the fly and therefore an automatic phonetization tool is indispensable.

The goal of this project is to develop a dynamic-vocabulary ASR system incorporating an automatic phonetization tool.

12. Text-to-Speech Synthesis for the Romanian language (2 students)

Speech Synthesis means producing a speech signal that carries the information given by a text. The main challenge is to make the speech as natural as possible, i.e. taking into account the speaking rhythm, intonation, syllable stressing, etc. There are several algorithms that are used in speech synthesis, such as PSOLA and HMM based ones. Other issues in Speech Synthesis are related to Natural Language Processing such as diacritics restoration, division into syllables, etc.

The objective of the project is to build a simple Text-to-Speech system based on HMM. This includes speech database collection and annotation, implementation of the HMM based algorithms and the evaluation of the system with variable text input.