

ABSTRACT

A CONTINUOUS SPEECH RECOGNITION SYSTEM FOR TURKISH LANGUAGE BASED ON TRIPHONE MODEL

PATLAR, Fatma

M.S. in Computer Engineering

Supervisor: Assist.Prof. Dr. Ertuğrul SAATÇI

August 2009, 73 pages

The field of speech recognition has been growing in popularity for various applications. Such recognition embedded applications include automated dictation systems and command interfaces. Embedding recognition to a product allows a unique level of hands-free usage and user interaction. Our main goal was to develop a system that can perform a relatively accurate transcription of speech and in particular, a Continuous Speech Recognition based on Triphone model for Turkish Language. Turkish is generally different from Indo-European languages (English, Spanish, French, German etc.) by its agglutinative and suffixing morphology. Therefore vocabulary growth rate is very high and as a consequence, constructing a continuous speech recognition system for Turkish based on whole words is not feasible. By considering this fact in this thesis, acoustic models which are based on triphones, are modelled as five state Hidden Markov Models. Mel-Frequency Cepstral Coefficients (MFCC) approach was preferred as the feature vector extraction method and training is done using embedding training that uses Baum-Welch re-estimation. Recognition is implemented on a search network which can be ultimately seen as HMM states connected by transitions and Viterbi Token Passing algorithm runs on this network to find the mostly likely state sequence according to the utterance. Also to make a more accurate recognition bigram language model is constructed. MATLAB is used in processing speech and The Hidden Markov Model Tool Kit (HTK) is used to train models and perform recognition.

The performance of this thesis has been evaluated using two different databases, one of them is more commonly formed TURTEL speech database that is used for speaker independent system tests and the other one is particularly formed weather forecast reports database that is used for speaker dependent system tests.

In recognition experiments, word accuracy of speaker independent system has been measured as 59-63 percent. After finding optimum value for decision tree pruning factor by try outs, system tests have been repeated again by using the language model and the optimum pruning factor. These adjustments improved the performance by 30-33 percent and word accuracy has reached to 92-93 percent for all tests.

While the word accuracy of the speaker dependent system tests on the single person database is between 89-93 percent, usage of the language model and the optimum decision tree pruning factor has resulted with an increase in the performance and the word accuracy has reached to 95-97 percent.

Keywords: Continuous Speech Recognition, Triphone, Hidden Markov Model, Language Modelling, Bigram language model

ÖZET

ÜÇLÜ SES MODELLİ TÜRKÇE SÜREKLİ KONUŞMA TANIMA SİSTEMİ

PATLAR, Fatma

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi: Yrd.Doç.Dr. Ertuğrul SAATÇI

Ağustos 2009, 73 sayfa

Konuşma tanıma tabanlı uygulamaların popülaritesi her geçen gün daha da artmaktadır. Bu uygulamalara dikte sistemlerini ve komut arayüzlü sistemleri örnek olarak verebiliriz. Bir ürüne konuşma tanımayı entegre etmek kullanıcıya benzersiz bir kullanım kolaylığı ve etkileşim imkanı sunar. Bizimde buradaki asıl amacımız Türkçe için nispeten hassas çeviri imkanı sunacak geniş kelime dağarcıklı bir sistem tasarlamaktır. Türkçe, sondan eklemeli morfolojisiyle genel olarak Hint-Avrupa dillerinden (İngilizce, İspanyolca, Fransızca, Almanca vs.) farklıdır. Bu yapısı sözcük dağarcığında büyük bir artışa neden olmakta ve sonuç olarak Türkçe için kelime tabanlı sürekli konuşma tanıma sistemlerinin yapılabirliği pek mümkün olmamaktadır. Bu gerçeğide göz önüne alarak, bu tezde, akustik modeller, beş durumlu Saklı Markov Modelleri olarak modellenmiş üçlü-sesler temel alınarak oluşturulmuşlardır. Özellik vektörü çıkarımı için Mel Kepstral Katsayılar yaklaşımı tercih edilmiş, eğitim ise Baum-Welch yeniden tahmin algoritmasını kullanan "gömülü eğitim" yöntemi kullanılarak yapılmıştır. Tanıma işlemi bir arama ağı üzerinde işleyen Viterbi Token Passing algoritması kullanılarak gerçekleştirilmiştir. Bu arama ağı aslında model durumlarının geçişlerler birbirine bağlanmış hali olarak görülebilir. Aynı zamanda daha doğru bir tanıma yapabilmek için ikili dil modellemesi de uygulanmıştır. SMM'i, "gömülü eğitim" kullanılarak eğitilmiş; tanıma kısmında ise "Andaç geçirmeli Viterbi algoritması" kullanılmıştır.

Konuřmanın analizi ve iřlenmesinde MATLAB; modellerin eęitimde ve tanımanın gerekleřtirilmesinde ise Hidden Markov Toolkit (HTK)'den faydalanılmıřtır.

Eęitim ve testlerde iki ayrı ses veritabanı kullanılmıřtır. Genel amalı hazırlanmıř olan TURTEL veritabanı kullanıcı baęımsız testlerde, daha özel amalı oluřturulan hava durumu tahmin raporları veritabanı ise kullanıcı baęımlı sistem testlerinde kullanılmıřtır.

Konuřmacı baęımsız sistem tanıma testlerinde kelime doęruluk yzdesi 59-63 olarak hesaplandı. Sistem performansını arttırmak iin en uygun karar aęacı budama eřięi seildi ve bunun sisteme dil modeli ile uygulanmasının ardından yzde 30-33 arası artış saęlanarak doęruluk yzdesinde 92-93 arası deęerler elde edildi. Kullanıcı baęımlı olan tek kiřilik veritabanında yapılan testlerde doęruluk oranı yzde 89-93 civarında iken, en uygun karar aęacının ve dil modelinin kullanılmasının doęruluk oranını yzde 95-97'lere yzselittięi gözlemlendi.

Anahtar Kelimeler : Sreklı Konuřma Tanıma, Dil Modeli, Ulü Ses, Saklı Markov Modeli, İkili Dil Modeli