OPTIMIZACIÓN DE UN SERVICIO AUTOMÁTICO DE PÁGINAS BLANCAS POR TELÉFONO

PROYECTO IDAS

Autores: R. Córdoba, R. San-Segundo, J. Colás, J.M. Montero, J. Ferreiros, J. Macías-Guarasa, A. Gallardo, J.M. Gutiérrez, J.M. Pardo

Grupo de Tecnología del Habla. Departamento de Ingeniería Electrónica.
Universidad Politécnica de Madrid

E.T.S.I. Telecomunicación. Ciudad Universitaria s/n, 28040 Madrid

Telf: 91 5495700 ext. 343

Fax: 91 3367323

cordoba@die.upm.es

http://www-gth.die.upm.es

ÁREA IV. Servidores de aplicación y contenidos multimedia

RESUMEN

I.- INTRODUCCIÓN

El proyecto IDAS (LE4-8315), objeto de esta comunicación, es un proyecto de dos años (1998-2000), financiado por la Unión Europea en el que interviene nuestro grupo representando a la UPM. El objetivo fundamental del proyecto ha sido el desarrollo de un sistema automático mediante reconocimiento de voz capaz de dar un servicio de páginas blancas por teléfono, proporcionando números de teléfono o fax, tanto de particulares como de empresas [Leht 00].

El sistema de reconocimiento desarrollado es de habla aislada para grandes vocabularios (10000 palabras) e independiente del locutor.

Esta comunicación es la continuación de la presentada en las jornadas anteriores [San-Seg 99] y que obtuvo el premio a la mejor ponencia en su área. Describiremos los resultados finales y el trabajo realizado en el último año del proyecto, en el que cabe destacar los aspectos siguientes:

- Incremento del tamaño del vocabulario de 1.000 a 10.000 palabras. Este incremento supone un gran salto cualitativo porque: aumenta el tiempo de procesamiento y aumenta en mayor media las posibilidades de confusión entre palabras similares. Al incluir 10.000 apellidos, por ejemplo, aparecen un gran número de los mismos que se diferencian únicamente en uno o dos fonemas, lo que se dificulta la tarea del reconocimiento.
- Mejoras en el sistema de modelado. Como consecuencia del incremento en el vocabulario resultó necesario mejorar el sistema de modelado. En concreto, se han desarrollado distintos sistemas de modelado dependiente del contexto tanto semicontinuo como continuo.
- Se ha desarrollado un sistema de deletreo al que se recurre cuando el reconocimiento normal no es capaz de resolver lo pronunciado por el usuario.

Aunque en esta comunicación se hará énfasis en el módulo de reconocimiento y sus mejoras, no menos importante es el sistema completo, llamado *Servidor Vocal Interactivo (SVI)*. Un SVI no es más que un sistema capaz de proporcionar un servicio de información a través de la línea telefónica, utilizando síntesis y reconocimiento de voz. También describiremos las mejoras introducidas en este sistema desde la última comunicación.

II.- ANTECEDENTES

Esta aplicación tiene su origen en un Entorno integrado de desarrollo de aplicaciones telefónicas (TADE) desarrollado en muestro grupo. Este entorno proporciona un nuevo lenguaje, con ciertas primitivas de alto nivel, para el diseño de aplicaciones telefónicas, principalmente *SVIs*, incluyendo igualmente utilidades para cubrir todo el ciclo de vida de una aplicación: diseño, compilación y ejecución [Cas 97].

Como ejemplos de *SVIs* desarrollados con este entorno cabe destacar el sistema de atención al cliente de Hewlett Packard, que identifica al usuario a través del reconocimiento de su código de cliente (por voz o DTMF) y redirige la llamada hacia alguno de los ingenieros que esté libre en esos momentos; el servidor de notas que ofrece nuestro departamento a sus alumnos, para consultar sus calificaciones por teléfono sin más que identificarse con su DNI, y otros sistemas similares.

III.- DESARROLLO DE LA APLICACIÓN DE PÁGINAS BLANCAS

Utilizando el entorno de desarrollo aplicaciones telefónicas comentado anteriormente (TADE), realizamos el desarrollo del demostrador objeto del proyecto IDAS. Esta aplicación consta de los pasos siguientes:

- En primer lugar, se descuelga y se da un mensaje de bienvenida al usuario. Se le pregunta si desea un teléfono particular o de empresa.
- A continuación se procede con el reconocimiento de los datos del cliente del que se desea conocer el teléfono. El orden seguido es:
  1) Para particular: nombre de la ciudad, primer apellido de la persona y nombre. Si hay incertidumbre, se pregunta también el segundo apellido.
  2) Para empresa: nombre de la ciudad y nombre de la empresa.
- Para cualquiera de estos datos, el sistema de reconocimiento pide al usuario que confirme el resultado obtenido (llamado primer candidato). Si el usuario rechaza dicho resultado, se le ofrece el segundo candidato (en segunda posición del reconocimiento). Si también lo rechaza, pasa a un módulo de deletreo, donde el usuario debe deletrear el dato introducido. También se le pide al usuario que confirme el resultado de este deletreo.
- Si no se ha podido obtener alguno de los datos (el usuario rechaza todos los resultados del reconocimiento), se transfiere la llamada a un módulo de operador. Se le ofrece al operador un cuadro de diálogo donde se le permite escuchar lo dicho por el usuario y rellenar un cuadro de texto con el dato correcto. En ningún momento hay comunicación directa con el cliente, de modo que el paso a operador es transparente para el usuario.
- Una vez obtenidos los datos correctos, se accede a la base de datos y se proporciona el teléfono solicitado.

IV.- SISTEMA DE RECONOCIMIENTO

Para afrontar el reto del reconocimiento de gran vocabulario resulta necesario mejorar el sistema de modelado. Además de la mayor probabilidad de confusión, el tiempo de procesamiento aumenta. La solución a este problema es descomponer la tarea en dos módulos de complejidad diferente:

1. Una etapa de preselección. Este módulo es un sistema de reconocimiento rápido y no preciso mediante el cual se reduce la lista de candidatos a reconocer. Este módulo se basa en los resultados de TIDAISL [Mac 96] (proyecto realizado por nuestro grupo en colaboración con Telefónica I+D) y desarrollos posteriores [Fer 98] [San-Seg 99].
2. Una etapa de verificación (reconocimiento más detallado). En esta etapa se utilizan modelos más detallados y es la que determina el resultado final del reconocimiento. Describiremos en detalle los experimentos de laboratorio y los resultados obtenidos utilizando modelos HMM continuos (CHMM) y semicontinuos (SCHMM), siendo la unidad básica el alófono dependiente del contexto (alófono que tiene en cuenta los alófonos que le rodean). A continuación, se compone el modelo de la palabra aislada concatenando los modelos de cada uno de sus alófonos dependientes del contexto.
   - El mayor problema de estos sistemas es el elevado número de unidades que existen en el lenguaje que hacen imposible su entrenamiento. La solución es utilizar técnicas de agrupación de las unidades más similares para reducir su número [Cor 95] [Gav 99]. Hay distintas

posibilidades para la agrupación de dichas unidades que describiremos en detalle en la comunicación.

- En el caso de CHMM, también es crítica la elección del número de gaussianas a utilizar en cada uno de los estados del modelo, la forma óptima de incrementar el número de gaussianas correspondientes a un estado del modelo y las medidas de distancia entre gaussianas que se necesitan en los algoritmos.

3. Un módulo de deletreo. Este módulo es especialmente novedoso, utilizando un proceso en tres etapas realmente efectivo que describiremos en la comunicación.

La base de datos utilizada para entrenar y verificar los modelos HMM (tanto SCHMM como CHMM) ha sido SPEECHDAT [Mor 97]. La tasa de error de reconocimiento en laboratorio con un vocabulario de 10.000 palabras ha sido del 23.1%. En el módulo de deletreo, la tasa de error ha sido del 12.7%. En la comunicación describiremos en detalle los resultados para cada una de las opciones contempladas.

V.- EVALUACIÓN DEL SISTEMA

Para evaluar la tasa de reconocimiento obtenida en condiciones reales, se ha realizado una evaluación del sistema con usuarios finales mayoritariamente no expertos en este tipo de sistemas.

En la siguiente tabla se ofrece la tasa de reconocimiento para cada diccionario en primer (candidato 1) y segundo (candidato 2) lugar, en el deletreo (se aplica sólo cuando se falla, por lo que la tarea es más complicada, dado que se suele fallar en las peores condiciones de ruido y con las palabras que presentan mayores posibilidades de confusión), junto con la tasa global. Todos los diccionarios utilizados tienen un tamaño de 1.000 palabras, excepto el de apellidos, que es de 10.000.

	Ciudades	**Empresas**	**Apellidos**
Candidato 1	62.18%	64.37%	32.66%
Candidato 2	70,27%	70.44%	40.16%
Deletreo	52.72%	40.41%	36.27%
Global	85.95%	82.38%	69.34%

Tabla 1. Tasa de reconocimiento

La tasa global de obtención del número deseado (se combinan todos los reconocimientos) sin intervención de la operadora ha sido del 58.69%

Estos resultados son provisionales, obtenidos con un total de 990 intentos de obtener el número de teléfono. En la comunicación presentaremos los finales.

La duración media del diálogo ha sido de 82.5 seg. para teléfono particular y 61.8 seg. para teléfono de empresa. En la comunicación detallaremos los tiempos medios de obtención de cada uno de los datos y comentaremos en detalle los resultados finales, así como los resultados de la encuesta de satisfacción del usuario.

VI.- CONCLUSIÓN

El demostrador desarrollado funciona en tiempo real en un ordenador Pentium III–450Mhz. Según la evaluación realizada, la tasa de intervención de operador en las peores condiciones es menor al 42%, lo que nos da una tasa mínima del 58% de llamadas procesadas automáticamente. Podemos concluir que los *Servidores Vocales Interactivos* son una solución interesante para el camino de automatización y abaratamiento de los servicios por línea telefónica. Se reduce el número de veces que debe intervenir la

operadora y, cuando interviene, únicamente debe escuchar el fragmento de voz que no ha sido capaz de resolver el reconocimiento.

VII.- REFERENCIAS

[Cas 97] Casas, A., "Sistema telefónico multilínea con reconocimiento de voz y acceso a base de datos remota". Proyecto Fin de Carrera. Grupo de Tecnología del Habla. DIE. UPM. Madrid, 1997.

[Cor 95] Córdoba, R. "Sistemas de reconocimiento de habla continua y aislada: comparación y optimización de los sistemas de modelado y parametrización". Tesis Doctoral. Grupo de Tecnología del Habla. DIE. UPM. Madrid, 1995.

[Fer 98] Ferreiros, J., Macías-Guarasa, J., Gallardo, A., Colás, J., Cordoba, R., Pardo, J.M., Villarrubia, L. "Recent Work on Preselection Module for Flexible Large Vocabulary Speech Recognition System in Telephone Enviroment", ICSLP'98, Sidney (Australia), Nov. 98.

[Gav 99] Gavina, D., "Distintas alternativas de compartición de parámetros en modelos HMM continuos en un sistema de reconocimiento de habla aislada". Proyecto Fin de Carrera. Grupo de Tecnología del Habla. DIE. UPM. Madrid, 1999.

[Leht 00] Lehtinen, G., S. Safra, …, J.M. Pardo, R. Córdoba, R. San-Segundo, et al., "IDAS : Interactive Directory Assistance Service", VOTS-2000 Workshop, Belgium.

[Mac 96] Macías-Guarasa, J., Gallardo, A., Ferreiros, J., Pardo, J.M., Villarrubia, L. "Initial Evaluation of a Preselection Module for a Flexible Large Vocabulary Speech Recognition System in Telephone Enviroment". ICSLP'96, pag 1343-1346. Philadelphia (USA), Oct. 96.

[Mor 97] Moreno, A., *SpeechDat* [cd-rom]. Ver. 1.0. [Barcelona]: Universitat Politècnica de Catalunya <http://www.upc.es/castella/recerca/recerca.htm>, c1997. 4 cd-roms. (Spanish Fixed Network Speech Corpus)

[Par 95] Pardo, J.M., et al "Spanish text to speech: from prosody to acoustic" International Conference on Acoustic 95 vol III, 1995.

[San-Seg 99] San-Segundo, R., Colás, J., Montero, J.M., Córdoba, R., Ferreiros. J., Macías-Guarasa J., Gallardo A., Gutiérrez, J.M., Pastor, J., Pardo, J.M.: "Servidores vocales interactivos: desarrollo de un servicio de páginas blancas por teléfono con reconocimiento de voz - proyecto IDAS (interactive telephone-based directory assistance service). IX jornadas Telecom. I+D. 1999.