論 文 内 容 の 要 旨

博士論文題目

On supervised learning from sequential data with applications for speech recognition
（情報系列からの教師付き学習とその音声認識への応用）

氏　　名
シュスター・マイク

（論文内容の要旨）

（１，２００字程度）

音声認識に代表される情報系列の認識問題は工学的に非常に興味深い分野であり、その多くの場合において、入力系列 $\mathbf{x}_1^T = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, ..., \mathbf{x}_{T-1}, \mathbf{x}_T\}$ から出力系列 $\mathbf{y}_1^T = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, ..., \mathbf{y}_{T-1}, \mathbf{y}_T\}$ へのマッピングを、情報系列データから教師付きで学習する問題としての定式化が可能である。本論文では、種々の問題を一般の情報系列認識問題として統一的に扱う視点に立ち、モデリング性能、及び認識性能を向上させるための枠組みとアルゴリズムについて述べるとともに、これらを擬似的に合成したデータと実際の音声データを用いて評価した結果を示す。

リカレント・ニューラル・ネットワーク（RNN）は、従来のニューラル・ネットワークに対して情報系列を取り扱うのに適した構造を持たせた強力なモデリング手法であり、これにより時刻 t までに与えられた全ての入力情報に対する出力情報出現確率 $P(\mathbf{y}_t|\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t)$ が推定可能である。本論文の前半部では、基本的な RNN を以下に示す通り様々に拡張したものについて述べる。

a) 双方向リカレント・ニューラル・ネットワーク（BRNN）：最終時刻までの全ての入力情報に対する出力情報出現確率 $P(\mathbf{y}_t|\mathbf{x}_1^T)$ を推定可能とするものであり、単峰形関数に基づく回帰、及び識別問題に適用可能

b) 拡張 BRNN：$P(\mathbf{y}_t|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ..., \mathbf{y}_1, \mathbf{x}_1^T)$ をモデリングすることによって、出力情報系列の出現事後確率 $P(\mathbf{y}_1^T|\mathbf{x}_1^T)$ を直接推定可能とするもの

c) 多峰形の分布に従う入力情報に対して、ある出力情報系列の条件下での出現確率 $P(\mathbf{x}_t|\mathbf{y}_1^T)$ を、隣接する入力情報 $\mathbf{x}_t, \mathbf{x}_{t+1}$ が互いに独立と仮定して推定する BRNN

d) c) の拡張として、隣接入力情報間に独立性が仮定できない場合に、$P(\mathbf{x}_t|\ \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, ..., \mathbf{x}_1, \mathbf{y}_1^T)$ をモデリングすることによって、$P(\mathbf{x}_1^T|\mathbf{y}_1^T)$ を近似なしに推定可能な BRNN

また、本論文の後半部では、最も確からしい単語系列を高速に、かつ少ないメモリ量で探索可能な、音声認識用 one-pass スタックデコーダについて述べる。本デコーダは、任意次数の N-gram 言語モデル、及び単語境界音素環境も考慮した任意次数の音素環境依存音響モデルが使用可能で、現在広く用いられている日本語新聞ディクテーションタスクでの評価により最高レベルの性能が示された。

氏　名	シュスター，マイク

（論文審査結果の要旨）

本論文は、音声のような時系列情報を取り扱う認識学習問題についての研究をまとめたものである。２つの問題にテーマを絞って述べている。

まず、リカレントニューラルネットワークによる時系列情報の学習と認識について、理論および実験による検証について報告している。従来の手法としては、ある時刻以前の情報のみを利用できる学習／認識アルゴリズムが提案されていた。本論文では、ある時刻の前と後のすべての情報を同時に利用できる双方向リカレント・ニューラル・ネットワークを提案し、その有効性をTIMIT音声データベースを用いた音韻認識で実証した。この手法は、音声の調音結合を含んだ音韻のモデル化として有望と考えられる。さらに、この双方向リカレント・ニューラル・ネットワークの拡張として、３種類の拡張双方向リカレント・ニューラル・ネットワークを提案し、それらの有効性も同じTIMITデータベースで実証している。

この双方向リカレント・ニューラル・ネットワークの研究は、時系列情報の一般的かつ強力な学習／認識のパラダイムを提案している。今後の進展も大いに期待できる。これらの研究は、IEEEの信号処理に論文として掲載されただけでなく、権威のある "Encyclopedia of Electrical and Electronic Engineering" の中にも含まれている。

本論文の後半部では、効率良く時系列情報を認識するアルゴリズムの研究として、音声認識のディクテーションアルゴリズムについて述べている。そこでは、最も確からしい単語系列を高速に、かつ少ないメモリ量で探索可能な、音声認識用one-passスタックデコーダについて考察し、新しいアルゴリズムを提案し、かつ、パーソナルコンピュータに実装して、その有効性を確認している。このデコーダは、任意次数のN-gram言語モデル、及び単語境界音素環境も考慮した任意次数の音素環境依存音響モデルが使用可能で、現在広く用いられている日本語新聞ディクテーションタスクでの評価により最高レベルの性能が示されている。この結果も国際会議などで発表し、内外の注目を浴びている。

近い将来、前半の双方向リカレント・ニューラル・ネットワークと、後半の音声認識用one-passスタックデコーダが統合され、音声認識の新しい手法となることが大いに期待できる。今後の音声認識などの時系列情報の認識の研究に貢献する研究成果であると確信する。