単純階層型ニューラルネットワーク構築システム

ComsaNNBP version 1.51

標 準 取 扱 説 明 書

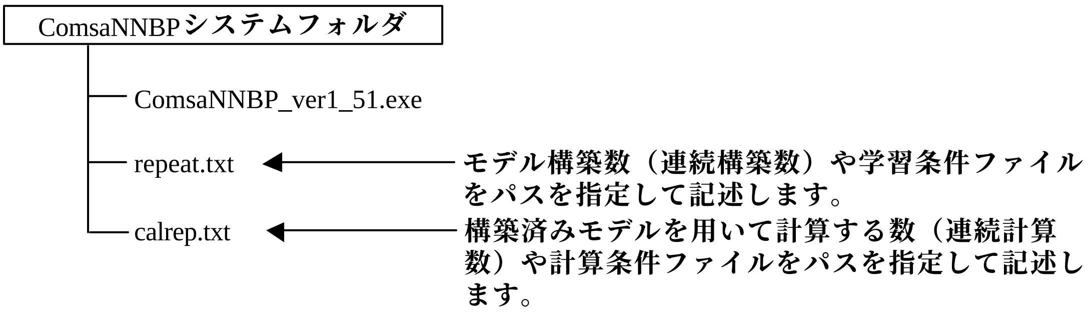
このシステムは，文字認識および画像処理，最適化問題等に用いられるニューラルネットワークモデルのうち，単純な階層構造モデルの構築を容易に行うためのシステムである．学習アルゴリズムとして誤差逆伝搬法を採用し，信号伝達関数にシグモイド関数を装備している．階層型のモデルであれば，構造は容易に設定でき，シグモイド関数の各パラメータも全て任意に指定可能である．

１．システムの構成

本システムを使用するにあたって必要となるデータファイル等について解説します．本システムをダウンロードされた場合は，sample フォルダを付属してあるので，本書とともに実作業の参考にして下さい．

１．１　システムの基本構成

本システムの基本となる構成は図－1.1 に示すとおりです．

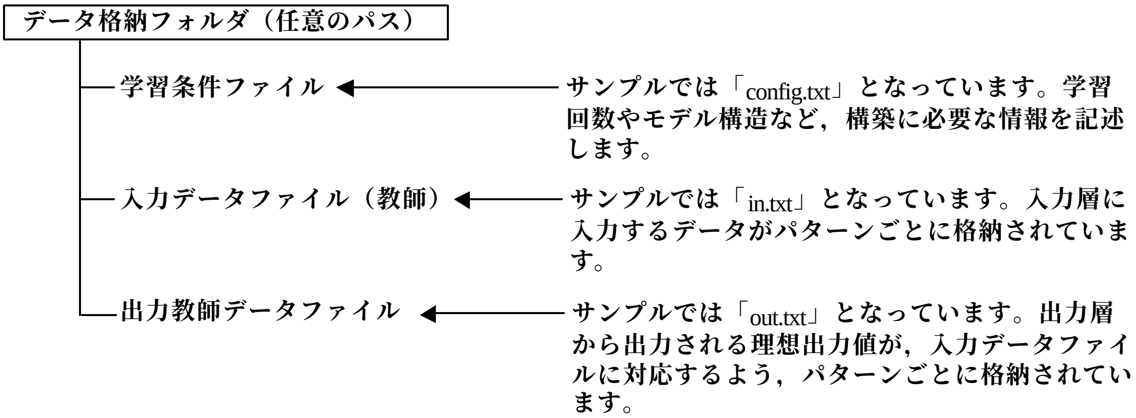


図－1.1　ComsaNNBP の基本システム構成

”repeat.txt”，”calrep.txt” の２つのファイルは，現在のバージョンでは名称を変更できません。

１．２　モデルの構築に必要なデータ

本システムを用いて階層型ニューラルネットワークモデルを構築する際に必要となるデータおよびシステム構成は図－1.2 に示すとおりです．



図－1.2　モデル構築に必要となるデータ

１．３　構築したモデルを用いた計算に必要なデータ

本システムを用いて構築した階層型ニューラルネットワークモデルを用いて計算を行う際に必要となるデータおよびシステム構成は図―1.3 に示すとおりです．

データ格納フォルダ（任意のパス）

- 計算条件ファイル ← サンプルでは「calconfig.txt」となっています。計算を行うデータ数やモデル構造，パスを含んだ重みファイル名など必要な情報を記述します。
- 入力データファイル ← サンプルでは，教師入力と同じ「in.txt」を用います。入力層に入力する計算データが全て格納されています。
- 重みファイル ← サンプルには当初は含まれません。サンプルでは，モデルを構築すると「weight.txt」が生成されます。このファイルが重みファイルです。

図－1.3　構築したモデルを用いて計算を行うのに必要なデータ

２．データファイルの詳細

前節で述べたそれぞれのファイルの詳細について解説します．

２．１　システム直下のテキストファイル

(a)　repeat.txt

本システムでは，異なる別々のモデルを連続して構築することができます．これを連続構築と呼びます．システムを起動してモデルの構築を選択すると，システムは最初に ”repeat.txt” にアクセスし，連続構築するモデルの総数と，構築するそれぞれのモデルの学習条件が記載されているファイル名を読み込んで記憶します．図―2.1 にサンプルファイルの詳細を示します．図―2.1 では，学習条件ファイルは，本システムからの相対パスを明示して指定してあります．学習条件ファイルのパスの指定は相対パスでも絶対パスでも構いませんが，かならず指定するようにしましょう．パスの指定が間違っていると正しくモデルを構築できません．また，学習条件ファイルのファイル名は，パス指定を含めて 60 字までです．ご注意下さい．

3	←	連続構築するモデルの総数
論理積¥config.txt	←	１個目の構築モデルの学習条件ファイル
論理和¥config.txt	←	２個目の構築モデルの学習条件ファイル
排他的論理和¥config.txt	←	３個目の構築モデルの学習条件ファイル

図－2.1　サンプルデータの repeat.txt の内容と解説

(b)　calrep.txt

本システムでは，モデル構築時と同様に，構築された別々の異なったモデルを用いて連続して結果を計算することができます．これを連続計算と呼びます．システムを起動して計算を選択すると，システムは ”calrep.txt” にアクセスし，連続計算を行う総数（モデル数）と，それぞれの計算条件が記載されているファイル名を読み込んで記憶します．図－2.2 にファイルの詳細を示します．”calrep.txt” も ”repeat.txt” と同様，計算条件ファイルのパスの指定は相対パスでも絶対パスでも構いませんが，かならず指定して下さい．また，こちらも計算条件ファイルのファイル名は，パス指定を含めて 60 字までです．

3	←	連続計算するモデルの総数
論理積¥calconfig.txt	←	１個目の計算モデルの計算条件ファイル
論理和¥calconfig.txt	←	２個目の計算モデルの計算条件ファイル
排他的論理和¥calconfig.txt	←	３個目の計算モデルの計算条件ファイル

図－2.2　サンプルデータの calrep.txt の内容と解説

２．２　モデル構築時のファイル

(a)　学習条件ファイル

学習条件ファイルは，学習回数や学習打ち切り誤差，教師データのデータ数，モデル構造などの基本情報の他，入出力ファイルの指定などを行います．また，シグモイド関数の各パラメータも記述します．図－2.3 にサンプルファイルの詳細を示します．

100000	←	**最大学習回数（上限）**
0.001	←	**学習打ち切り二乗誤差**
4	←	**入力教師データの数**
2	←	**入力層のニューロン数（しきい値除く）**
5	←	**中間層のニューロン数（しきい値除く）**
1	←	**出力層のニューロン数**
排他的論理和¥in.txt	←	**入力データファイル（教師）**
排他的論理和¥out.txt	←	**出力教師データファイル（理想出力値）**
排他的論理和¥weight.out	←	**構築したモデルの結合強度を出力するファイル（構築後にシステムが作成するのでファイル名を指定してください。）**
排他的論理和¥**構築情報**.txt	←	**構築したモデルの詳細な情報を出力するファイル（構築後にシステムが作成するのでファイル名を指定してください。構築したモデルの確認に役立ちます。）**
排他的論理和¥**誤差情報**.txt	←	**構築したモデルの最小二乗誤差情報を出力するファイル（構築後にシステムが作成するのでファイル名を指定してください。学習回数の確認に役立ちます。）**
1.0	←	**温度（シグモイド関数の分母** $1+\exp(-2x/u_0)$**の** u_0 **の部分）**
0.9	←	**安定化係数１（ＢＰの時の通常結合強度に乗じる値）**
0.9	←	**安定化係数２（ＢＰの時のしきい値結合強度に乗じる値）**
0.5	←	**学習率１（中間層と出力層結合強度更新時に乗じる値）**
0.5	←	**学習率２（入力層と中間層結合強度更新時に乗じる値）**

図－2.3　学習条件ファイル（サンプル config.txt ）の内容と解説

(b)　入力データファイル

入力データファイルは，モデル構築の基本になるデータです．図−2.4 にサンプルファイルの詳細を示します．データは必ず，改行以外は全て ”tab” で区切って下さい．”tab” 区切り以外でデータを区切った場合はデータの読み込みに失敗する恐れがあり，正しくモデルが構築されないことがあります．ご注意下さい．

0.000	0.000	←	1つ目のデータ。 図−2.3 によれば，入力層のニューロン数は 2つであるので，2列にわたってデータが 記載される。（データ間の区切りは「タブ」）
1.000	0.000	←	同様に2つ目のデータ。
0.000	1.000	←	同様に3つ目のデータ。
1.000	1.000	←	同様に4つ目のデータ。 図−2.3 によれば，データ数は4つであるので， 4行にわたってデータが記載される。 （データ間の区切りは「改行」）

図−2.4　入力データファイル（サンプル in.txt ）の内容と解説

(c)　出力教師データファイル

出力教師データファイルは，入力データファイルに対応し，かつ，出力層のニューロン数にも対応したものである必要があります．図−2.5 にサンプルファイルの詳細を示します．

0.000	←	1つ目の入力データに対応する理想出力値 （図−2.3 によれば，出力層のニューロン数は 1つであるので，1列にデータが並ぶ。 もし2列以上になる場合は，入力データと同様に 区切りは「タブ」とする。）
1.000	←	同様に2つ目の入力データに対応する理想出力値
1.000	←	同様に2つ目の入力データに対応する理想出力値
0.000	←	同様に2つ目の入力データに対応する理想出力値

図−2.5　出力教師データファイル（サンプル out.txt ）の内容と解説

(d)　モデル構築時に生成されるファイル

モデルの構築は以上のファイルがあれば開始できますが，本システムはモデル構築時にいくつかの「モデル情報ファイル」を生成します．表－2.1 に生成されるファイルの詳細をサンプルを例にとって示します．

表－2.1　モデル構築時に生成されるファイル（サンプルの場合）

生成ファイル	内　　容 （詳しくはサンプルデータを用いて実際にモデルを構築してみて下さい）
構築情報.txt	構築したモデルの構造や入出力教師データファイルなどの情報が記録されています。
誤差情報.txt	学習１回毎の最小二乗誤差が記録されています。
weight.out	構築したモデルの結合強度がマトリックス状に出力されます。このファイルは構築したモデルを用いて計算を行うときに必要です。

２．３　構築したモデルを用いた計算を行う時のファイル

(a)　計算条件ファイル

計算条件ファイルは，計算する入力データのデータ数，モデル構造，構築したモデルの重み情報ファイルなどを指定します．ファイル作成の際は，モデルの構造と重み情報ファイルは必ず対応がとれていることを，モデル構築情報ファイルなどから確認して下さい．図－2.6 にサンプルファイルの詳細を示します．

4	←	入力データ数
2	←	入力層のニューロン数
5	←	中間層のニューロン数
1	←	出力層のニューロン数
排他的論理和¥in.txt	←	計算するデータの入力ファイル
排他的論理和¥weight.out	←	生成された重みファイル
排他的論理和¥最終結果.txt	←	計算結果が格納されるファイル（出力ファイル）
排他的論理和¥計算情報.txt	←	使用したモデル構造等，計算の詳細を出力するファイル。自動生成されるので，名前を決定して下さい。
1.0	←	シグモイド関数の温度（構築時と同一にして下さい）

図－2.6　計算条件ファイル（サンプル calconfig.txt ）の内容と解説

(b)　入力データファイル

入力データファイルは，計算の基本になるデータです．基本的には構築時に用いたデータと作成方法は同じです．詳細は２．２(b) 節をご覧下さい．

(c)　出力ファイル

出力ファイルは，入力データファイルに対応した計算結果が出力されるファイルです．図－2.7 にサンプルファイルから生成される出力ファイル（当初はありません）を示します．

0.026215	←１つ目のデータに対する出力値
0.955613	←２つ目のデータに対する出力値
0.960223	←３つ目のデータに対する出力値
0.0600818	←４つ目のデータに対する出力値

図－2.7　出力データファイル（サンプル 最終結果.txt ）の内容と解説

(d)　計算情報ファイル

計算情報ファイルは，計算に用いたモデルや入力データ，出力先等をまとめて記載したものです．終了後に自動生成されます．計算結果の怪しい場合の確認等に利用できます．

３．プログラムの改造・転載について

本システムは，平成 14 年 6 月 1 日をもってオープンソースとしました．これは，既に NN 手法は一般的なものとなっており，幾つもの BP を用いたソースが広く出回っていることから，当システムもソースを広く公開すべきだとの観点に立ってのものです．これを機に，より使いやすく精度の高いシステムになれば・・と考えております．

したがいまして，プログラムの改造は自由とさせていただきます．

また，転載につきましては，改造前・改造後を問わず，お近くの本庄勉各機関までご連絡下さい．瀬戸内沿岸警備隊ホームページからも本庄勉各機関にメールが送信できますのでご利用下さい．

本システムが少しでも皆様のお役に立てば幸いです．

平成 14 年 6 月　　制作・著作：本庄　勉