

文書の読解を助ける補助情報を自動付与する技術

文書中の言葉に補助的な情報を自動的に付与するための技術です。例えば、漢字に読みがなを振ったり、難解な言葉に注釈や解説を加えるなど、文書の読解を助けるための用途で利用できます。

特長

- 汎用性が高いため、文書に補助情報を付与することを目的とした、さまざまな用途で利用できます。
- 自動付与しようとする情報に複数の候補がある場合には、文脈を考慮して使い分けられます。
- 自動付与の結果を手で修正した場合には、再学習を繰り返すことで性能が向上します。

利用分野

- 子供や外国人を対象とした、わかりやすい文書の作成
- 難解な言葉や専門用語に対する注釈や解説のリンク付け

キーワード 自動情報付与、機械学習、やさしい日本語

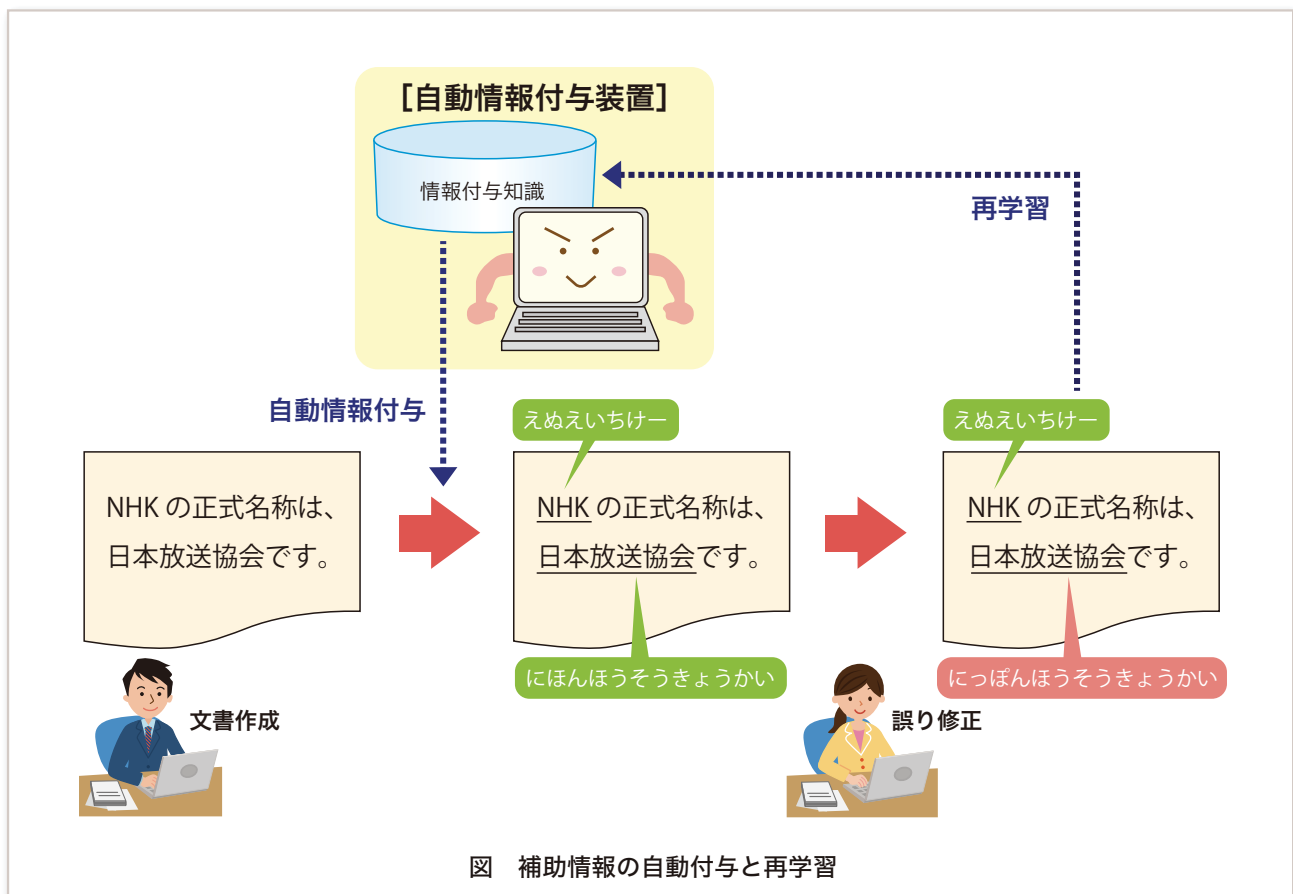


図 補助情報の自動付与と再学習

技術解説

NHKはインターネットで、外国人向けに「やさしい日本語ニュース」を提供しています。このサイトでは、一般向けのニュース原稿を平易な日本語に書き換えた後に、読解を助けるための読み、意味、注釈等の補助情報を付加した原稿を提供しています。「やさしい日本語ニュース」の制作現場では、このような補助情報の付与を、自動的に行う技術が使われています。なお、本技術は、「やさしい日本語ニュース」の制作の用途に限定されるものではなく、私たちが日常的に使う文書の作成にも適用できます。

以下、裏面の図を用いて、本技術の特徴を説明します。

(1) 補助情報の自動付与

「自動情報付与装置」の入力に、ある文書を与えると、本装置に備えた「情報付与知識」を用いて、読みがなや注釈等の補助情報が自動付与されます。「情報付与知識」によって文脈を考慮することができるため、自動付与しようとする情報に複数の候補がある場合には、その文脈にふさわしい情報を使うことができます。

(2) 情報付与知識の再学習

補助情報を自動付与した結果に対し、人手で修正した場合には、修正結果を再学習し、「情報付与知識」を更新する機能をもっています。補助情報を付与する作業を継続することによって、自動付与の性能が向上していきます。

提供可能な技術

- 補助情報を事前に付与した文書から「情報付与知識」を学習する技術
- 上記の知識を用いて、文書中の各言葉に補助情報を自動付与する技術
- 自動付与誤りの修正結果を、上記の知識にフィードバックするシステムの構築