

- ※1:入力された単語(seq\_len)のうちマスクされた不明単語がvocab(金単語)のどれかをSoftMax関数で計算する。
  ※2: 連続的に存在する意味があって関係が深い文章」の「全く関係がなく支援のつながりがない2つの文章」のどちらかを判定する。
  ※3:最後のBertLayerの出力では文章中のある単語とその周辺単語の関係性をベースにその単語の意味にあった特徴ペクトルが生成される。
  一例えば、銀行のbankと土手のbankは同じ単語でも特徴の文章によって意味が異なってくる。

## Embeddingモジュール構成図 Embeddingモジュール ・ 文章の長さL分の数字が昇順で BERTではこの位置情報を学習させる。 ならんだもの [0.1,2,3,4,・・・・レ-1] n個 位置情報 position\_ids [バッチサイズ,長さ,次元数 ] (N, L, D) input\_ids [パッチサイズ,長さ,次元数 ] (N, L, D) [バッチサイズ,長さ] (N, L) [バッチサイズ,長さ,次元数 ] (N, L, D) 文章の1 or2文章目を表す情報(1or0) token\_type\_ids [パッチサイズ,長さ] (N, L) [バッチサイズ,長さ,次元数 ] (N, L, D) token\_type\_idsの例 tensor([[0, 0, 1, 1, 1], [0, 1, 1, 1, 1]])

## BertLayerモジュール全体構成図



書籍「つくりながら学ぶ! PvTorchによる発展ディープラーニング | P409~のコードを図式化



