

請求の範囲

- [請求項1] 予測対象および第1の属性を含む第1のテーブルと、第2の属性を含む第2のテーブルとを取得するテーブル取得手段と、
- 前記第1の属性と前記第2の属性との類似度の算出に用いられる類似度関数と、当該類似度に対する条件とを受け付ける受付手段と、
- 前記類似度関数を用いて、前記第1の属性の値と前記第2の属性の値とにより算出される類似度が、前記条件を満たす当該第1の属性の値を含む第1のテーブルのレコードと、当該第2の属性の値を含む第2のテーブルのレコードとを結合するための結合条件と、第2のテーブルにおける複数のレコードに対する集約方法および当該集約の対象になる列により表される集約条件とを用いて、前記第1のテーブルおよび前記第2のテーブルから、前記予測対象に影響を及ぼし得る特徴量の候補を生成する特徴量生成手段と、
- 前記特徴量の候補から、予測に最適な特徴量を選択する特徴量選択手段とを備えた
- ことを特徴とする特徴量生成装置。
- [請求項2] 特徴量生成手段は、複数の結合条件と複数の集約条件とを用いて、当該結合条件と当該集約条件とを組み合わせた数の特徴量を生成する
- 請求項1記載の特徴量生成装置。
- [請求項3] 第1の属性および第2の属性は、地理的属性であり、類似度関数は、距離が近いほど類似度を高く算出する関数として定義される
- 請求項1または請求項2記載の特徴量生成装置。
- [請求項4] 第1の属性および第2の属性は、時間的属性であり、類似度関数は、時間の差異が小さいほど類似度を高く算出する関数として定義される
- 請求項1または請求項2記載の特徴量生成装置。
- [請求項5] 第1の属性および第2の属性は、文字列属性であり、類似度関数は、文字列を示すテキストの一致度が高いほど類似度を高く算出する関

数として定義される

請求項 1 または請求項 2 記載の特徴量生成装置。

[請求項6] 第 1 の属性および第 2 の属性は、構造的属性であり、類似度関数は、共通するノードまでの距離が近いほど類似度を高く算出する関数として定義される

請求項 1 または請求項 2 記載の特徴量生成装置。

[請求項7] 結合条件と、集約条件と、第 2 のテーブルから条件を満たすレコードを抽出するための抽出条件とを用いて特徴量生成関数を生成する関数生成手段を備えた

請求項 1 から請求項 6 のうちのいずれか 1 項に記載の特徴量生成装置。

[請求項8] 予測対象および第 1 の属性を含む第 1 のテーブルと、第 2 の属性を含む第 2 のテーブルとを取得し、

前記第 1 の属性と前記第 2 の属性との類似度の算出に用いられる類似度関数と、当該類似度に対する条件とを受け付け、

前記類似度関数を用いて、前記第 1 の属性の値と前記第 2 の属性の値とにより算出される類似度が、前記条件を満たす当該第 1 の属性の値を含む第 1 のテーブルのレコードと、当該第 2 の属性の値を含む第 2 のテーブルのレコードとを結合するための結合条件と、第 2 のテーブルにおける複数のレコードに対する集約方法および当該集約の対象になる列により表される集約条件とを用いて、前記第 1 のテーブルおよび前記第 2 のテーブルから、前記予測対象に影響を及ぼし得る特徴量の候補を生成し、

前記特徴量の候補から、予測に最適な特徴量を選択する

ことを特徴とする特徴量生成方法。

[請求項9] 複数の結合条件と複数の集約条件とを用いて、当該結合条件と当該集約条件とを組み合わせた数の特徴量を生成する

請求項 8 記載の特徴量生成方法。

[請求項10]

コンピュータに、

予測対象および第1の属性を含む第1のテーブルと、第2の属性を含む第2のテーブルとを取得するテーブル取得処理、

前記第1の属性と前記第2の属性との類似度の算出に用いられる類似度関数と、当該類似度に対する条件とを受け付ける受付処理、

前記類似度関数を用いて、前記第1の属性の値と前記第2の属性の値とにより算出される類似度が、前記条件を満たす当該第1の属性の値を含む第1のテーブルのレコードと、当該第2の属性の値を含む第2のテーブルのレコードとを結合するための結合条件と、第2のテーブルにおける複数のレコードに対する集約方法および当該集約の対象になる列により表される集約条件とを用いて、前記第1のテーブルおよび前記第2のテーブルから、前記予測対象に影響を及ぼし得る特徴量の候補を生成する特徴量生成処理、および、

前記特徴量の候補から、予測に最適な特徴量を選択する特徴量選択処理

を実行させるための特徴量生成プログラム。

[請求項11]

コンピュータに、

特徴量生成処理で、複数の結合条件と複数の集約条件とを用いて、当該結合条件と当該集約条件とを組み合わせた数の特徴量を生成させる

請求項10記載の特徴量生成プログラム。