Package 'msentropy'

July 23, 2025

Type Package

Title Spectral Entropy for Mass Spectrometry Data

Version 0.1.4

Date 2023-08-07

Description Clean the MS/MS spectrum, calculate spectral entropy, unweighted entropy similarity, and entropy similarity for mass spectrometry data. The entropy similarity is a novel similarity measure for MS/MS spectra which outperform the widely used dot product similarity in compound identification. For more details, please refer to the paper: Yuanyue Li et al. (2021) ``Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification" <doi:10.1038/s41592-021-01331-z>.

License Apache License (== 2.0)

Depends R (>= 3.5.0), Rcpp (>= 1.0.10)

Suggests testthat

LinkingTo Rcpp

RoxygenNote 7.2.3

Encoding UTF-8

URL https://github.com/YuanyueLi/MSEntropy

NeedsCompilation yes

Author Yuanyue Li [aut, cre]

Maintainer Yuanyue Li <liyuanyue@gmail.com>

Repository CRAN

Date/Publication 2023-08-07 23:10:02 UTC

Contents

`calculate_entropy_similarity` *Entropy similarity between two spectra*

Description

Calculate the entropy similarity between two spectra

Usage

```
calculate_entropy_similarity(
  peaks_a,
  peaks_b,
  ms2_tolerance_in_da,
  ms2_tolerance_in_ppm,
  clean_spectra,
  min_mz,
  max_mz,
  noise_threshold,
  max_peak_num
)
```

Arguments

`peaks_a`	A matrix of spectral peaks, with two columns: mz and intensity
`peaks_b`	A matrix of spectral peaks, with two columns: mz and intensity
`ms2_tolerance_in_da`	The MS2 tolerance in Da, set to -1 to disable
`ms2_tolerance_in_ppm`	The MS2 tolerance in ppm, set to -1 to disable
`clean_spectra`	Whether to clean the spectra before calculating the entropy similarity, see `clean_spectrum`
`min_mz`	The minimum mz value to keep, set to -1 to disable
`max_mz`	The maximum mz value to keep, set to -1 to disable
`noise_threshold`	The noise threshold, set to -1 to disable, all peaks have intensity < noise_threshold * max_intensity will be removed
`max_peak_num`	The maximum number of peaks to keep, set to -1 to disable

Value

The entropy similarity

Examples

```
mz_a <- c(169.071, 186.066, 186.0769)
intensity_a <- c(7.917962, 1.021589, 100.0)
mz_b <- c(120.212, 169.071, 186.066)
intensity_b <- c(37.16, 66.83, 999.0)
peaks_a <- matrix(c(mz_a, intensity_a), ncol = 2, byrow = FALSE)
peaks_b <- matrix(c(mz_b, intensity_b), ncol = 2, byrow = FALSE)
calculate_entropy_similarity(peaks_a, peaks_b,
                             ms2_tolerance_in_da = 0.02, ms2_tolerance_in_ppm = -1,
                             clean_spectra = TRUE, min_mz = 0, max_mz = 1000,
                             noise_threshold = 0.01,
                             max_peak_num = 100)
```

`calculate_spectral_entropy` *Calculate spectral entropy of a spectrum*

Description

Calculate spectral entropy of a spectrum

Usage

```
calculate_spectral_entropy(peaks)
```

Arguments

`peaks`	A matrix of peaks, with two columns: m/z and intensity.

Value

A double value of spectral entropy.

Examples

```
mz <- c(100.212, 300.321, 535.325)
intensity <- c(37.16, 66.83, 999.0)
peaks <- matrix(c(mz, intensity), ncol = 2, byrow = FALSE)
calculate_spectral_entropy(peaks)
```

`calculate_unweighted_entropy_similarity` *Unweighted entropy similarity between two spectra*

Description

Calculate the unweighted entropy similarity between two spectra

Usage

```
calculate_unweighted_entropy_similarity(
  peaks_a,
  peaks_b,
  ms2_tolerance_in_da,
  ms2_tolerance_in_ppm,
  clean_spectra,
  min_mz,
  max_mz,
  noise_threshold,
  max_peak_num
)
```

Arguments

`peaks_a`	A matrix of spectral peaks, with two columns: mz and intensity
`peaks_b`	A matrix of spectral peaks, with two columns: mz and intensity
`ms2_tolerance_in_da`	The MS2 tolerance in Da, set to -1 to disable
`ms2_tolerance_in_ppm`	The MS2 tolerance in ppm, set to -1 to disable
`clean_spectra`	Whether to clean the spectra before calculating the entropy similarity, see `clean_spectrum`
`min_mz`	The minimum mz value to keep, set to -1 to disable
`max_mz`	The maximum mz value to keep, set to -1 to disable
`noise_threshold`	The noise threshold, set to -1 to disable, all peaks have intensity < noise_threshold * max_intensity will be removed
`max_peak_num`	The maximum number of peaks to keep, set to -1 to disable

Value

The unweighted entropy similarity

Examples

```
mz_a <- c(169.071, 186.066, 186.0769)
intensity_a <- c(7.917962, 1.021589, 100.0)
mz_b <- c(120.212, 169.071, 186.066)
intensity_b <- c(37.16, 66.83, 999.0)
peaks_a <- matrix(c(mz_a, intensity_a), ncol = 2, byrow = FALSE)
peaks_b <- matrix(c(mz_b, intensity_b), ncol = 2, byrow = FALSE)
calculate_unweighted_entropy_similarity(peaks_a, peaks_b,
                                    ms2_tolerance_in_da = 0.02, ms2_tolerance_in_ppm = -1,
                                        clean_spectra = TRUE, min_mz = 0, max_mz = 1000,
                                        noise_threshold = 0.01,
                                        max_peak_num = 100)
```

---

clean_spectrum *Clean a spectrum*

Description

Clean a spectrum

This function will clean the peaks by the following steps: 1. Remove empty peaks (mz <= 0 or intensity <= 0). 2. Remove peaks with mz >= max_mz or mz < min_mz. 3. Centroid the spectrum by merging peaks within min_ms2_difference_in_da or min_ms2_difference_in_ppm. 4. Remove peaks with intensity < noise_threshold * max_intensity. 5. Keep only the top max_peak_num peaks. 6. Normalize the intensity to sum to 1.

Note: The only one of min_ms2_difference_in_da and min_ms2_difference_in_ppm should be positive.

Usage

```
clean_spectrum(
  peaks,
  min_mz,
  max_mz,
  noise_threshold,
  min_ms2_difference_in_da,
  min_ms2_difference_in_ppm,
  max_peak_num,
  normalize_intensity
)
```

Arguments

peaks	A matrix of spectral peaks, with two columns: mz and intensity
min_mz	The minimum mz value to keep, set to -1 to disable
max_mz	The maximum mz value to keep, set to -1 to disable

noise_threshold
: The noise threshold, set to -1 to disable, all peaks have intensity < noise_threshold * max_intensity will be removed

min_ms2_difference_in_da
: The minimum mz difference in Da to merge peaks, set to -1 to disable, any two peaks with mz difference < min_ms2_difference_in_da will be merged

min_ms2_difference_in_ppm
: The minimum mz difference in ppm to merge peaks, set to -1 to disable, any two peaks with mz difference < min_ms2_difference_in_ppm will be merged

max_peak_num
: The maximum number of peaks to keep, set to -1 to disable

normalize_intensity
: Whether to normalize the intensity to sum to 1

Value

A matrix of spectral peaks, with two columns: mz and intensity

Examples

```
mz <- c(100.212, 169.071, 169.078, 300.321)
intensity <- c(0.3716, 7.917962, 100., 66.83)
peaks <- matrix(c(mz, intensity), ncol = 2, byrow = FALSE)
clean_spectrum(peaks, min_mz = 0, max_mz = 1000, noise_threshold = 0.01,
               min_ms2_difference_in_da = 0.02, min_ms2_difference_in_ppm = -1,
               max_peak_num = 100, normalize_intensity = TRUE)
```

---

msentropy_similarity *Calculate spectral entropy similarity between two spectra*

Description

`msentropy_similarity` calculates the spectral entropy between two spectra (Li et al. 2021). It is a wrapper function defining defaults for parameters and calling the `calculate_entropy_similarity()` or `calculate_unweighted_entropy_similarity()` functions to perform the calculation.

Usage

```
msentropy_similarity(
  peaks_a,
  peaks_b,
  ms2_tolerance_in_da = 0.02,
  ms2_tolerance_in_ppm = -1,
  clean_spectra = TRUE,
  min_mz = 0,
  max_mz = 1000,
  noise_threshold = 0.01,
```

```
  max_peak_num = 100,
  weighted = TRUE,
  ...
)
```

Arguments

peaks_a	A two-column numeric matrix with the m/z and intensity values for peaks of one spectrum.
peaks_b	A two-column numeric matrix with the m/z and intensity values for peaks of one spectrum.
ms2_tolerance_in_da	The MS2 tolerance in Da, set to -1 to disable. Defaults to `ms2_tolerance_in_da = 0.02`.
ms2_tolerance_in_ppm	The MS2 tolerance in ppm, set to -1 to disable. Defaults to `ms2_tolerance_in_ppm = -1`.
clean_spectra	Whether to clean the spectra before calculating the entropy similarity, see `clean_spectrum()`.
min_mz	The minimum mz value to keep, set to -1 to disable. Defaults to `min_mz = 0`.
max_mz	The maximum mz value to keep, set to -1 to disable. Defaults to `max_mz = 1000`.
noise_threshold	The noise threshold, set to -1 to disable, all peaks have intensity < noise_threshold * max_intensity will be removed. Defaults to `noise_threshold = 0.01`, thus, by default, all peaks with an intensity less than 1% of the maximum intensity of a spectrum will be removed.
max_peak_num	The maximum number of peaks to keep, set to -1 to disable. Defaults to `max_peak_num = 1000`.
weighted	`logical(1)` whether the weighted or unweighted entropy similarity should be calculated. Defaults to `weighted = TRUE`, thus `calculate_entropy_similarity()` is used for the calculation. For `weighted = FALSE` `calculate_unweighted_entropy_similarity()` is used instead.
...	Optional additional parameters (currently ignored)

Value

The entropy similarity

References

Li, Y., Kind, T., Folz, J. et al. (2021) Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification. Nat Methods 18, 1524-1531. doi: 10.1038/s41592-02101331z.

Examples

```
peaks_a <- cbind(mz = c(169.071, 186.066, 186.0769),
    intensity = c(7.917962, 1.021589, 100.0))
peaks_b <- cbind(mz = c(120.212, 169.071, 186.066),
    intensity <- c(37.16, 66.83, 999.0))
msentropy_similarity(peaks_a, peaks_b, ms2_tolerance_in_da = 0.02)
```

Index