Special note:

The present work developed a new alignment algorithm called SFA-MS (Subwindow Factor Analysis based on Mass Spectral information), applying to nontargeted metabolic profiling analysis.By comparing with several alignment methods, it is demonstrated that the method has the capacity to solve the overlapping phenomenon and preserve the shape of peaks and performs well with nonlinear retention time shifts.

Algorithm operation steps and methods:

1. Collecting sub-matrices of interest from different data matrices of different batches, and “mat” format should be saved.

2. Using “load” to import data.

3. Establishing reference chromatogram and alignment chromatograms. The characteristic ion curve need to be selected manually.

4. Setting parameters, different parameters have meaning:

```
phi=20;                 % SNR for peak detection
l=6;                    % level for DWT , Discrete Wavelet Transform (DWT)
lambda=10e4;            % Smoothness for fitted baseline
valve=0.9;              % Eigenvalue evaluation criteria
max_shift=20;           % Maximum distance allowed to move
AMWFA_wz=3;             % The size of the mass spectrum window.
```

5. airPLS algorithm was utilized to proceed background subtraction for high-throughput chromatographic data within a short time.

```
%baseline correction
[sa,ssa]=airPLS(Data260.TIC, lambda,2,0.05);
[xa,xxa]=airPLS(Data247.TIC, lambda,2,0.05);
```

6. SFAMS operation is automatically performed, in which Haar CWT and FFT cross correlation are used for peak detection and candidate shifts calculation, respectively. Certainly, baseline correction and smoothing are executed beforehand for all chromatograms.

The corresponding function(in function ‘SFAMS.m’) is:

```
% smoothing are executed beforehand for all chromatograms
[xs,xln]=wavelet_denoising(xtic,l,6,'soft');
[ss,sln]=wavelet_denoising(reftic,l,6,'soft');
%peak detection and candidate shifts calculation
peaks=peak_detection(xtic,3,xln,phi);
peaks=peak_clustering(xtic,peaks,1);
```

7. The optimal shifts of each candidate peak were determined by comparing the corresponding mass spectra of the spectrum via the method of subwindow factor analysis (SFA).

   The corresponding function is:

```
[R,h]=SWFA(X,Y,Noise_threshold,AMWFA_wz,valve)
```

```
[xn,peaks,CoCe,shiftvalue]=SFAMS(reftic,refmz,xtic,xmz,phi,max_shift,valve,AMWFA_wz,l)
```

8. After FFT cross correlation calculation and SFA steps, the optimum drifts can be determined. Through linear interpolation in the non-peak parts, the aligned chromatograms remain unchanged peak shape and less defect data. By this means, the peak information can maintain the original state to the maximum extent, and mistakes are greatly reduced.