Fourier Transform, Time Series and Point Process

Zheng Wang

Outline

- Fourier Series & Fourier Transform
- Stochastic Process, Time Series, Point Process
- Spectral Analysis

Fourier Series

1 Suppose that the function $f(x)$ with period 2π is absolutely integrable on $[-\pi, \pi]$ so that the following so-called Dirichlet integral is finite:

$$\int_{-\pi}^{\pi} |f(x)| dx < \infty;$$

2 Suppose also that the function $f(x)$ is a single valued, piecewise continuous (must have a finite number of jump discontinuities), and piecewise monotonic (must have a finite number of maxima and minima).

If the conditions 1 and 2 are satisfied, the Fourier series for the function $f(x)$ exists and converges to the given function (see also the Convergence of Fourier Series page about convergence conditions.)

At a discontinuity x_0, the Fourier Series converges to

$$\lim_{\varepsilon \to 0} \frac{1}{2}[f(x_0 - \varepsilon) - f(x_0 + \varepsilon)].$$

The Fourier series of the function $f(x)$ is given by

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \{a_n \cos nx + b_n \sin nx\},$$

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) dx, \quad a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx, \quad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx dx.$$

Even function:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx,$$

$$a_0 = \frac{2}{\pi} \int_{0}^{\pi} f(x) dx, \quad a_n = \frac{2}{\pi} \int_{0}^{\pi} f(x) \cos nx dx.$$

Odd function:

$$f(x) = \sum_{n=1}^{\infty} b_n \sin nx,$$

$$b_n = \frac{2}{\pi} \int_{0}^{\pi} f(x) \sin nx dx.$$

Fourier Series

Fourier Series Expansion on the Interval $[a, b]$

If the function $f(x)$ is defined on the interval $[a, b]$, then its Fourier series representation is given by the same formula

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right),$$

where $L = \frac{b-a}{2}$ and Fourier coefficients are calculated as follows:

$$a_0 = \frac{1}{L} \int\limits_{-L}^{L} f(x) dx, \quad a_n = \frac{1}{L} \int\limits_{-L}^{L} f(x) \cos \frac{n\pi x}{L} dx, \quad b_n = \frac{1}{L} \int\limits_{-L}^{L} f(x) \sin \frac{n\pi x}{L} dx, \quad n = 1, 2, 3, .$$

Fourier Series

Complex Form:

$$
\begin{aligned}
f(x) &= \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \\
&= \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \frac{e^{inx} + e^{-inx}}{2} + b_n \frac{e^{inx} - e^{-inx}}{2i} \right) \\
&= \frac{a_0}{2} + \sum_{n=1}^{\infty} \frac{a_n - ib_n}{2} e^{inx} + \sum_{n=1}^{\infty} \frac{a_n + ib_n}{2} e^{-inx} \\
&= \sum_{n=-\infty}^{\infty} c_n e^{inx}.
\end{aligned}
$$

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx, \quad n = 0, \pm 1, \pm 2, \ldots$$

Euler's formulas:

$$\cos \varphi = \frac{e^{i\varphi} + e^{-i\varphi}}{2}, \quad \sin \varphi = \frac{e^{i\varphi} - e^{-i\varphi}}{2i},$$

$$\boxed{f(x) = \sum_{n=-\infty}^{\infty} c_n e^{\frac{in\pi x}{L}},}$$

where

$$\boxed{c_n = \frac{1}{2L} \int_{-L}^{L} f(x) e^{-\frac{in\pi x}{L}} dx, \quad n = 0, \pm 1, \pm 2, \ldots}$$

Temporal or spatial part: x

'Frequency': $\frac{n\pi}{L}$

'Amplitude': c_n

Fourier Series

Orthogonal Polynomials

Two polynomials $p(x)$ and $q(x)$ defined on the interval $[a, b]$ are orthogonal if

$$\int_a^b p(x)q(x)w(x)dx = 0,$$

where $w(x)$ is a nonnegative weight function.

A polynomial sequence $p_n(x)$, $n = 0, 1, 2, \ldots$, where n is the degree of $p_n(x)$, is said to be a sequence of orthogonal polynomials if

$$\int_a^b p_m(x)p_n(x)w(x)dx = c_n\delta_{mn},$$

where c_n are given constants and δ_{mn} is the Kronecker delta.

Generalized Fourier Series

A generalized Fourier series is a series expansion of a function based on a system of orthogonal polynomials. By using this orthogonality, a piecewise continuous function $f(x)$ can be expressed in the form of generalized Fourier series expansion:

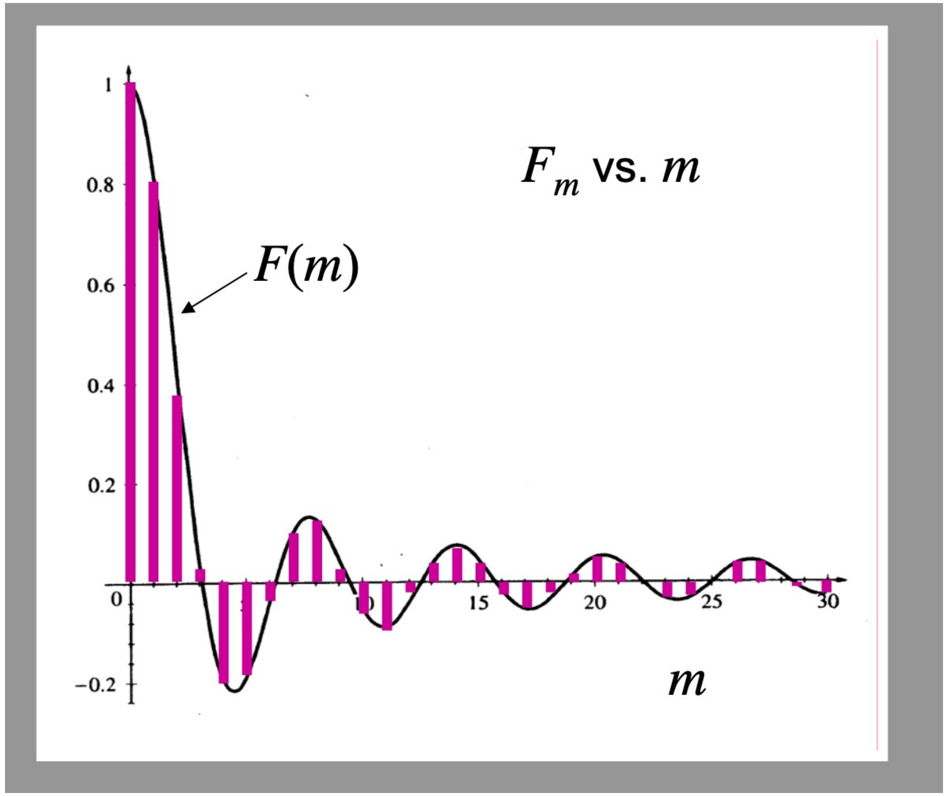
$$\sum_{n=0}^{\infty} c_n p_n(x) = \begin{cases} f(x), & \text{if } f(x) \text{ is continuous} \\ \frac{f(x-0)+f(x+0)}{2}, & \text{at a jump discontinuity} \end{cases}.$$

Fourier Transform

- Let the amplitude be a continuous function w.r.t. the frequency
- Frequency: extend discrete frequency to R;
- Temporal or spatial part: extend limit interval to infinity
- Change summation to integral accordingly

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-i\omega t)\, dt \qquad \textbf{FT}$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \exp(i\omega t)\, d\omega \qquad \textbf{IFT}$$



Fourier Transform

- DTFT

$$\mathcal{X}(\omega) = \sum_{n=-\infty}^{\infty} x[n]\, \mathrm{e}^{-\jmath\omega n}\,, \qquad x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{X}(\omega)\, \mathrm{e}^{\jmath\omega n}\, \mathrm{d}\omega\,.$$

- DFT

$$x(n) = \begin{cases} 0, & n < 0, \\ y(n), & 0 \le n \le (N-1), \\ 0, & n \ge N, \end{cases}$$

$$\begin{aligned} X(k) &= X(k\Delta\omega), \quad \Delta\omega = \frac{2\pi}{N} \quad \Longrightarrow \\ X(k) &= \sum_{n=0}^{N-1} x(n) e^{-j2\pi\frac{kn}{N}} \quad \text{DFT.} \end{aligned}$$

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k) e^{j2\pi\frac{kn}{N}}.$$

$$\begin{aligned} x(n) &= \frac{1}{N}\sum_{k=0}^{N-1}\left\{\sum_{m=0}^{N-1} x(m) e^{-j2\pi\frac{km}{N}}\right\} e^{j2\pi\frac{kn}{N}} \\ &= \sum_{m=0}^{N-1} x(m) \underbrace{\left\{\frac{1}{N}\sum_{k=0}^{N-1} e^{-j2\pi\frac{k(m-n)}{N}}\right\}}_{\delta(m-n)} = x(n). \end{aligned}$$

IDFT

Fourier Transform

- Energy preserving $\int_{\mathbb{R}} |x(t)|^2 \, dt = \int_{\mathbb{R}} |X(f)|^2 \, df$
- Fast Convolution(Cross-correlation)

$$z(t) = x(t) \star y(t) \qquad Z(f) = X(f) \cdot Y(f)$$

- A shifted (delayed) signal in the time domain manifests as a phase change in the frequency domain.
- Derivatives $\dfrac{d^n f(x)}{dx^n}$ | $(2\pi i\xi)^n \hat{f}(\xi)$

Stochastic Process

- Stochastic Process

In other words, for a given probability space $(\Omega, \mathcal{F}, P)$ and a measurable space (S, Σ), a stochastic process is a collection of S-valued random variables, which can be written as:[81]

$$\{X(t) : t \in T\}.$$

- Time Series
  - When t is the timestamp
- Point Process
  - Usually described by a counting process N(t), where for any time t between 0 and T, N(t) is the number of points occurring at or before time t.

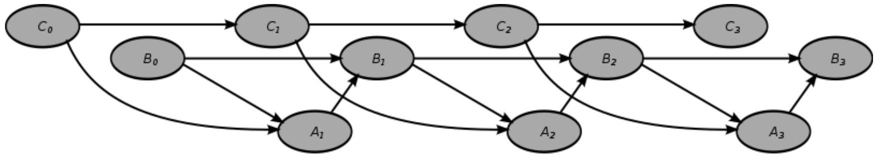
Time Series

- Autoregressive Models (ARIMA)

$$AR(p) = \sum_{i=1}^{p} a_i.X(t-i) + c + \epsilon_t, \quad MA(q) = \sum_{i=1}^{q} b_i.\epsilon_{t-i} + \mu + \epsilon_t,$$

- Dynamic Bayesian Networks: unroll Bayesian Networks along the time axis

$$P(V_t|V_{t-1}) = \prod_{x \in V, \pi_x \in V} P(x_t|\pi_{x_t})$$



- Gaussian Process $\mathbf{y}_n = f(\mathbf{x}_n) + \epsilon_n, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I})$
- Neural Networks

$$P(h_j|x) = \sigma(b_j + \sum_i W_{ij} x_i) + \sum_k \sum_i B_{ijk} x_i(t-k)$$
$$P(x_i|h) = \sigma(c_i + \sum_i W_{ij} h_j) + \sum_k \sum_i A_{ijk} x_i(t-k)$$

Point Process

- Conditional Intensity Function (Happening Rate)

$$\lambda^*(t) = \frac{f(t|\mathcal{H}_{t_n})}{1 - F(t|\mathcal{H}_{t_n})}.$$

$$\begin{aligned}
\lambda^*(t)\mathrm{d}t &= \frac{f(t|\mathcal{H}_{t_n})\mathrm{d}t}{1 - F(t|\mathcal{H}_{t_n})} \\
&= \frac{\mathbb{P}(t_{n+1} \in [t, t+\mathrm{d}t]|\mathcal{H}_{t_n})}{\mathbb{P}(t_{n+1} \notin (t_n, t)|\mathcal{H}_{t_n})} \\
&= \frac{\mathbb{P}(t_{n+1} \in [t, t+\mathrm{d}t], t_{n+1} \notin (t_n, t)|\mathcal{H}_{t_n})}{\mathbb{P}(t_{n+1} \notin (t_n, t)|\mathcal{H}_{t_n})} \\
&= \mathbb{P}(t_{n+1} \in [t, t+\mathrm{d}t]|t_{n+1} \notin (t_n, t), \mathcal{H}_{t_n}) \\
&= \mathbb{P}(t_{n+1} \in [t, t+\mathrm{d}t]|\mathcal{H}_{t-}) \\
&= \mathbb{E}[N([t, t+\mathrm{d}t])|\mathcal{H}_{t-}],
\end{aligned}$$

Example 2.3 (Hawkes process)**.** Define a point process by the conditional intensity function

$$\lambda^*(t) = \mu + \alpha \sum_{t_i < t} \exp(-(t - t_i)), \tag{2}$$

Point Process

Proposition 2.1. *The reverse relation of (1) is given by*

$$f(t|\mathcal{H}_{t_n}) = \lambda^*(t) \exp\left(-\int_{t_n}^{t} \lambda^*(s)\mathrm{d}s\right), \tag{3}$$

or

$$F(t|\mathcal{H}_{t_n}) = 1 - \exp\left(-\int_{t_n}^{t} \lambda^*(s)\mathrm{d}s\right), \tag{4}$$

where t_n is the last point before t.

Proof. By (1), we get that

$$\lambda^*(t) = \frac{f(t|\mathcal{H}_{t_n})}{1 - F(t|\mathcal{H}_{t_n})} = \frac{\frac{\mathrm{d}}{\mathrm{d}t}F(t|\mathcal{H}_{t_n})}{1 - F(t|\mathcal{H}_{t_n})} = -\frac{\mathrm{d}}{\mathrm{d}t}\log(1 - F(t|\mathcal{H}_{t_n})). \tag{5}$$

Integrating both sides, we get by the fundamental theorem of calculus that

$$\int_{t_n}^{t} \lambda^*(s)\mathrm{d}s = -(\log(1-F(t|\mathcal{H}_{t_n}))-\log(1-F(t_n|\mathcal{H}_{t_n}))) = -\log(1-F(t|\mathcal{H}_{t_n})),$$

Point Process

$$\Lambda^*(t) = \int_0^t \lambda^*(s)\mathrm{d}s.$$

Proposition 3.1. *Given an unmarked point pattern* $(t_1, \ldots, t_n)$ *on an observation interval* $[0, T)$, *the likelihood function is given by*

$$L = \left(\prod_{i=1}^{n} \lambda^*(t_i)\right) \exp(-\Lambda^*(T)).$$

Given a marked point pattern $((t_1, \kappa_1), \ldots, (t_n, \kappa_n))$ *on* $[0, T) \times \mathbb{M}$, *the likelihood function is given by*

$$L = \left(\prod_{i=1}^{n} \lambda^*(t_i, \kappa_i)\right) \exp(-\Lambda^*(T)).$$

$$L = f(t_1|\mathcal{H}_0) f(t_2|\mathcal{H}_{t_1}) \cdots f(t_n|\mathcal{H}_{t_{n-1}})(1 - F(T|\mathcal{H}_{t_n})),$$

Point Process

A point process N may be called *self-exciting* if $cov\{N(s,t), N(t,u)\} > 0$ *for* $s < t < u$. N is *self-correcting* if instead this covariance is negative. Thus the occurrence of points in a self-exciting point process causes other points to be more likely to occur, whereas in a self-correcting process, the points have an inhibitory effect. By definition, a Poisson process is neither self-exciting nor self-correcting.

Spectral Analysis (Time Series)

$$\mathrm{K}_{XX}(t_1, t_2) = \mathrm{cov}[X_{t_1}, X_{t_2}] = \mathrm{E}[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})] = \mathrm{E}[X_{t_1} X_{t_2}] - \mu_{t_1}\mu_{t_2}$$

Corollary 20.1. *if $\gamma(h)$ is an autocovariance function, then*

$$\gamma(h) = \int_{-\pi}^{\pi} e^{\mathbf{i}ht} f(t)dt,$$

where f is the spectral density.

How to get the spectral density from the covariance function?

Theorem 20.5.

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$$

$$f(t) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-\mathbf{i}ht} \gamma(h)$$

Corollary 20.2. *$\gamma(h)$ is a covariance function, if and only if*

$$f(t) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-\mathbf{i}ht} \gamma(h) \geq 0 \qquad -\pi \leq t \leq \pi$$

Spectral Analysis (Point Process)

Conditional density function:

$$\Lambda_r(t) = \nu_r + \sum_{s=1}^{k} \int_{-\infty}^{t} g_{rs}(t-u)\, dN_s(u)$$

$$\mathcal{F}\left\{\int_{-\infty}^{t} f(\tau)\, d\tau\right\} = \frac{1}{i\omega} F(\omega) + \pi F(0)\delta(\omega),$$

Complete covariance density matrix:

$$\boldsymbol{\mu}^{(c)}(\tau) = \delta(\tau)\,\mathrm{diag}\,(\boldsymbol{\Lambda}) + \boldsymbol{\mu}(\tau),$$

$$\lambda = E\{dN(t)\}/dt$$

$$\mu(\tau) = E\{dN(t+\tau)\, dN(t)\}/(dt)^2 - \lambda^2$$

Spectral density function:

$$F(\omega) = \frac{1}{2\pi}\left\{\mathrm{diag}\,(\boldsymbol{\Lambda}) + \int_{-\infty}^{\infty} e^{-i\tau\omega} \boldsymbol{\mu}(\tau)\, d\tau\right\}$$

Spectral Analysis

Theorem 1 (Bochner (Rudin, 1962)). *A continuous kernel of the form $\nu(x, x') = \kappa(x - x')$ defined over a locally compact set $\mathcal{X} \subset \mathbb{R}^d$ is positive definite if and only if g is the Fourier transform of a non-negative measure:*

$$\nu(x, x') = \kappa(x - x') = \int_{\Omega} p(\omega) e^{jw^\top (x - x')} d\omega, \qquad (6)$$

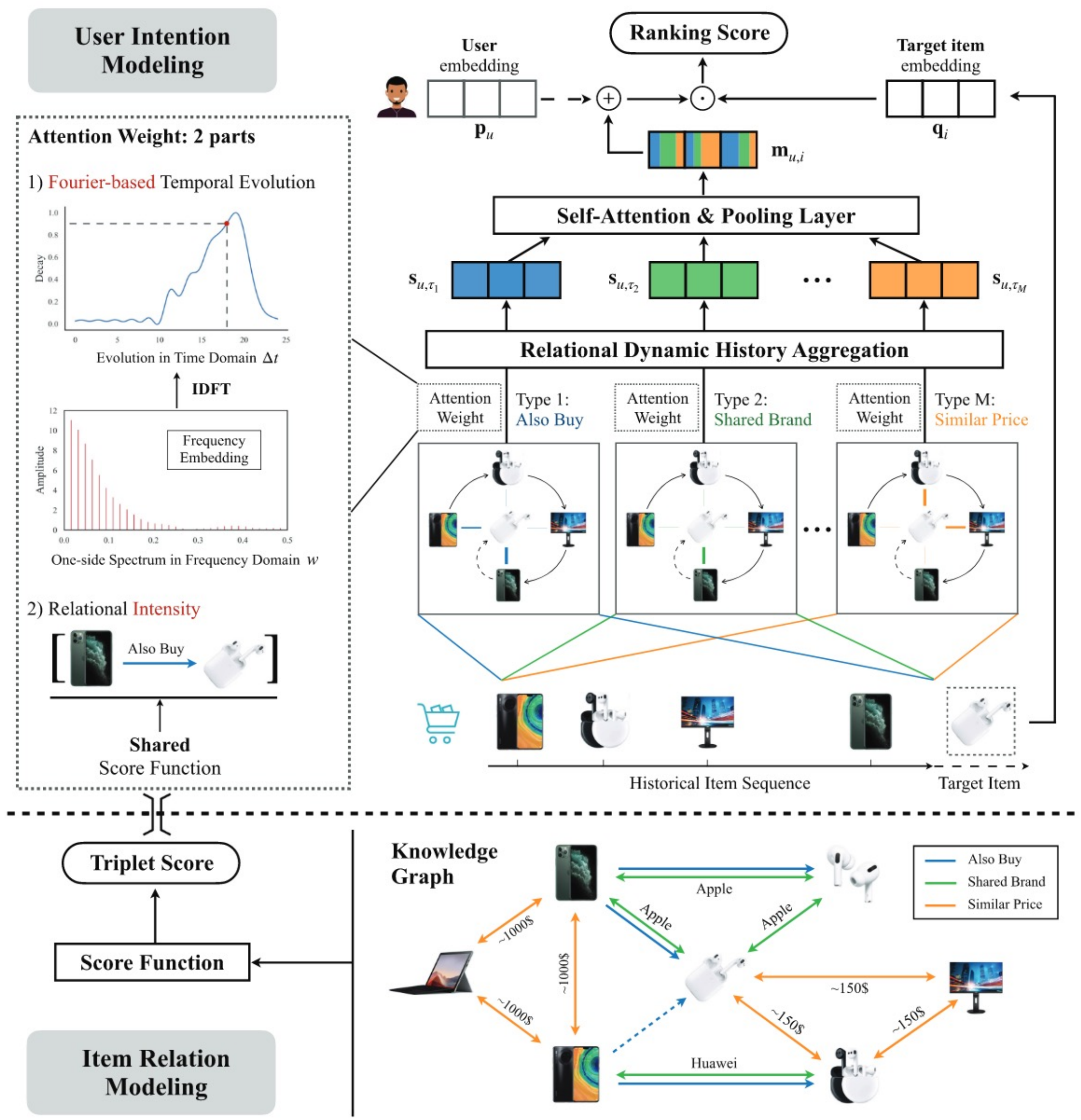
where p is a non-negative measure, Ω is the Fourier feature space, and kernels of the form $\nu(x, x')$ are called shift-invariant kernel.

If a shift-invariant kernel $\kappa(\cdot)$ is properly scaled such that $\kappa(0) = 1$, Bochner's theorem guarantees that its Fourier transform $p(\omega)$ is a proper probability distribution.

$$\lambda(x|\boldsymbol{h}(x)) = \underbrace{\mu(x)}_{\text{base intensity}} + \underbrace{g\left(\boldsymbol{h}(x)^\top W + b\right)}_{\text{triggering effect}},$$

$$h^{(k)}(x) = \sum_{t_i < t} \widetilde{\nu}^{(k)}(x, x_i) \varphi^{(k)}(x_i), \;\; k = 1, \dots, K, \quad (4)$$

$$\nu^{(k)}(x, x') := \mathbb{E}\left[\phi_\omega^{(k)}(x) \cdot \phi_\omega^{(k)}(x')\right],$$


User Intention Modeling
Attention Weight: 2 parts
1) Fourier-based Temporal Evolution
Decay
Evolution in Time Domain Δt
IDFT
Amplitude
Frequency Embedding
One-side Spectrum in Frequency Domain w
2) Relational Intensity
Also Buy
Shared Score Function
Ranking Score
User embedding
p_u
m_{u,i}
Target item embedding
q_i
Self-Attention & Pooling Layer
s_{u,τ1}
s_{u,τ2}
s_{u,τM}
Relational Dynamic History Aggregation
Attention Weight
Type 1: Also Buy
Type 2: Shared Brand
Type M: Similar Price
Historical Item Sequence
Target Item
Triplet Score
Score Function
Item Relation Modeling
Knowledge Graph
Apple
~1000$
~150$
Huawei
Also Buy
Shared Brand
Similar Price

Thanks!