1 Features

Let z denote a keyword, $\mathcal{Z}$ a set of keywords and $\mathcal{Z}([t_0;t])$ this set of keywords observed during the time interval $[t_0;t]$. Our goal is to learn a ranking function f that produces, for a given subset of keywords, a ranking of their activity during the next time periods:

$$f : \mathcal{Z}([t_0;t]) \rightarrow R(\mathcal{Z}, t+\delta)$$

where $R(\mathcal{Z}, t+\delta)$ denotes a ranking over the keywords in $\mathcal{Z}$ for the time interval $[t; t+\delta]$.

1.1 Content centric approach

The information one can extract from the user network can be useful; however, (a) it is not always possible to retrieve it, and (b), if retrievable, it remains costly to keep it up to date. We thus propose here a framework that does not rely on it.

Définition 1.1 *(Atomic container $\langle z, u, \tau \rangle$) Each content publication within a social media is an atomic container $\langle z, u, \tau \rangle$ for a set of keywords $z \subseteq \mathcal{Z}$ produced by a user $u \in \mathcal{U}$ at the time stamp $\tau \in \mathcal{T}$. We define $\mathcal{C}$ to be the set of any existing atomic containers.*

Définition 1.2 *(Discussion d_t) A discussion d_t is defined as a sequence of temporally ordered atomic containers:*

$$d_t = \left\{ \langle z^1, u^1, \tau^1 \rangle, \ldots, \langle z^{l_{d_t}}, u^{l_{d_t}}, \tau^{l_{d_t}} \rangle \right\}, \text{ with } \tau^{l_{d_t}} \leq t$$

$\mathcal{D}$ denotes the set of all discussions. A discussion generally encompasses several keywords, it is conceivable to define topics on top of them. The function $\texttt{pair} : \mathcal{D} \times \mathcal{C} \mapsto \mathcal{D}$ *is used to increment a discussion with new atomic container.*

Définition 1.3 *(Users and Activity functions) We define two functions that provide information used to define features:*

1. $\texttt{users} : \mathcal{D} \mapsto \mathcal{U}^n$ *that provides the set of users involved in d_t:* $\texttt{users}(d_t) = \{u \in \mathcal{U} \mid \langle z, u, \tau \rangle \in d_t\}$*;*
2. $\texttt{activity} : \mathcal{Z} \times \tau \mapsto \mathbb{N}^+$ *that provides keyword's observed activity at a given time:* $\texttt{activity}(z,t) = |\{\langle z, u, t \rangle \in \mathcal{C}\}|$. *We abbreviate it by* $\texttt{A}(z,t)$*;*

The above functions furthermore allow one to obtain the set of users interacting in discussions $\mathcal{D}_{t,z}$ related to a keyword z until time t:

$$\mathcal{U}_{t,z} \quad = \quad \{\texttt{users}(d_t) \mid d_t \in \mathcal{D}_{t,z}\}$$

As an illustration of the above framework, consider a social media as Twitter, in which users exchange size-bounded text messages called "tweets". The users network is directed as the "follow" relation is asymmetric, when user u follows user v, u receives each publications emitted by v implying nothing for v. In a such case a tweet equals to an atomic container. A discussion equals to a series of tweets for which the `pair` function is either "to reply" or "to re-tweet" indistinctly. Finally functions `users` and `activity` are simply implemented as enumerations applied on discussions.

1.2 Feature set

In order to predict keyword activity we use a simple feature set containing the objective feature itself (`activity`, defined above) and five other features. Among them three are shared between pairwise and point-wises approaches, two are specific to the pairewise approach.
Shared features.

1. **Number of Users** (NU). Denoted by $\text{NU}(t, z) = |\mathcal{U}_{t,z}|$, it corresponds to the number of users interacting on a keyword z at time t;

2. **User balance** (UB). This feature corresponds to the number of users interacting for the first time on a keyword z at time t: $\text{UB}(t, z) = |\mathcal{U}_{t,z} \setminus \mathcal{U}_{t-1,z}|$;

3. **Attention Level** (AL). $\rho = \text{NU}(t, z)$ or $\rho = \mathtt{A}(t, z)$ are surrogate estimators of the attention payed by users to keyword z at time t. We normalize them with the attention payed to any other keyword at time t, therefore it should cope better with external events. $\text{AL}(t, z) = \rho(t, z) \ / \sum\limits_{z' \in \mathcal{Z}} \rho(t, z')$.

Pairwise features. The following two features are not available in the dataset but can be easily computed for a keyword pair (z_1, z_2) when considering a pairwise approach. Here we considered that z_1 has a greater activity during the evaluation period than z_2 and t_f is the last observation time-step.

1. **Activity difference** (AD). This feature corresponds to the difference of activities at the end of the observation period. It is defined as $\text{AD}(z_1, z_2) = \mathtt{A}(z_1, t_f) - \mathtt{A}(z_2, t_f)$;

2. **Activity order** (AO). This feature counts the number of time steps for which z_1 has a higher activity than z_2 during the observation period. It is defined as $\text{AO} = \sum_{t=0}^{t_f} \mathbb{1}(\mathtt{A}(z_1, t) > \mathtt{A}(z_2, t))$ where $\mathbb{1}$ is the standard indicator function.