1) *Predicting the underlying topics and the underlying entities of a document* based on the observed information and the global knowledge. We call such a task the ***prediction task***;

2) *Estimating the global knowledge from data*. Notice that the topic knowledge, the entity name knowledge and the entity context knowledge are all not previously given, thus we need to estimate them from data. We call such a task the ***knowledge discovery task***.

Because the accurate inference of the above two tasks is intractable in our entity-topic model, this paper also develops an approximate inference algorithm – the Gibbs sampling algorithm to solve them.

Contributions. The main contributions of this paper are summarized below:

- We propose a generative probabilistic model, the *entity-topic model*, which can *jointly* model and exploit the context compatibility, the topic coherence and the correlation between them for better EL performance;
- We develop a *Gibbs sampling algorithm* to solve the two inference tasks of our model: 1) Discovering the global knowledge from data; and 2) Collectively making accurate EL decisions.

This paper is organized as follows. Section 2 describes the proposed entity-topic model. Section 3 demonstrates the Gibbs sampling algorithm. The experimental results are presented and discussed in Section 4. The related work is reviewed in Section 5. Finally we conclude this paper in Section 6.

2 The Entity-Topic Model for Entity Linking

In this section, we describe the proposed entity-topic model. In following we first demonstrate how to capture the context compatibility, the topic coherence and the correlation between them in the document generative process, then we incorporate the global knowledge generation into our model for knowledge estimation from data.

2.1 Document Generative Process

As shown in Section 1, we jointly model the context compatibility and the topic coherence as the statistical dependencies in the entity-topic model by assuming that all documents are generated in a topical coherent and context compatible way. In following we describe the document generative process.

In our model, each document d is assumed composed of two types of information, i.e., the *mentions* and the *words*. Formally, we represent a document as:

*A **document** is a collection of M mentions and N words, denoted as* $\mathbf{d} = \{m_1, \dots, m_M; w_1, \dots, w_N\}$*, with* m_i *the* i^{th} *mention and* w_j *the* j^{th} *word.*

For example, the document in Figure 1 is represented as $\mathbf{d}$ = {***WWDC***, ***Apple***, ***Lion***; *at*, *the*, *conference*, …}, where *WWDC*, *Apple*, *Lion* are the three mentions and the other are the words.

To generate a document, our model relies on three types of global knowledge, including:

- **Topic Knowledge** ϕ (*The entity distribution of topics*): In our model, all entities in a document are generated based on its underlying topics, with each topic is a group of semantically related entities. Statistically, we model each topic as a multinomial distribution of entities, with the probability indicating the likelihood an entity to be extracted from this topic. For example, we may have a topic $\phi_{Apple\ Inc.}$= {*Steve Jobs*$^{0.12}$, *iPhone*$^{0.07}$, *iPod*$^{0.08}$, *...*}, indicating the likelihood of the entity *Steve Jobs* be extracted from this topic is *0.12*, etc.
- **Entity Name Knowledge** ψ (*The name distribution of entities*): In our model, all name mentions are generated using the name knowledge of its referent entity. Specifically, we model the name knowledge of an entity as a multinomial distribution of its names, with the probability indicating the likelihood this entity is mentioned by the name. For example, the name knowledge of the entity *Apple Inc.* may be $\psi_{Apple\ Inc.}$ = {*Apple*$^{0.51}$, *Apple Computer Inc.*$^{0.10}$, *Apple Inc.*$^{0.07}$, *...*}, indicating that the entity *Apple Inc.* is mentioned by the name *Apple* with probability *0.51*, etc.
- **Entity Context Knowledge** ξ (*The context word distribution of entities*): In our model, all context words of an entity's mention are generated using its context knowledge. Concretely, we model the context knowledge of an entity as a multinomial distribution of words, with the probability indicating the likelihood a word appearing in this entity's context. For example, we may have $\xi_{Apple\ Inc.}$= {*phone*$^{0.07}$, *computer*$^{0.10}$, *IT*$^{0.06}$, *phone*$^{0.002}$, *...*}, indicating that the word *computer* appearing in the context of the entity *Apple Inc.* with probability *0.1*, etc.