can provide a *document-specific entity prior* for EL. Concretely, using the topic knowledge and the topic distribution of documents, the prior for an entity appearing in a document *d* is highly related to the document's topics:

$$P(e|d) = \sum_z P(z|d)P(e|z)$$

This prior is obviously more reasonable than the "information less prior" (i.e., all entities have equal prior) or "a global entity popularity prior" (Han & Sun, 2011). To demonstrate, Table 2-3 show the 3 topics where the *Apple Inc.* and the fruit *Apple* have the largest generation probability *P(e|z)* from these topics. We can see that the topic knowledge can provide a reasonable prior for entities appearing in a document: the *Apple Inc.* has a large prior in documents about *Computer*, *Video* and *Software*, and the fruit *Apple* has a large prior in documents about *Wine*, *Food* and *Plant*.

Topic(**Computer**)	Topic(**Video**)	Topic(**Software**)
Computer	*Video*	*Computer software*
CPU	*Mobile phone*	*Microsoft Windows*
Hardware	*Mass media*	*Linux*
Personal computer	*Music*	*Web browser*
Computer memory	*Television*	*Operating system*

Table 2. The 3 topics where the *Apple Inc.* has the largest P(e|z)

Topic(**Wine**)	Topic(**Food**)	Topic(**Plant**)
Wine	*Food*	*Plant*
Grape	*Restaurant*	*Flower*
Vineyard	*Meat*	*Leaf*
Winery	*Cheese*	*Tree*
Apple	*Vegetable*	*Fruit*

Table 3. The 3 topics where the fruit *Apple* has the largest P(e|z)

2) **The effects of a fine-tuned context model**. The second advantage of our model is that it provides a statistical framework for fine-tuning the context model from data. To demonstrate such an effect, Table 4 compares the EL performance of ① the entity-topic model with no context model is used (*No Context*), i.e., we determine the referent entity of a mention by deleting the 3rd term of the formula $P(e_i = e|\mathbf{z}, \mathbf{e}_{-i}, \mathbf{a}, \mathbf{D})$ in Section 3; ② with the context model estimated using the entity's Wikipedia page (*Article Content*), ③ with the context model estimated using the 50 word window of all its mentions in Wikipedia (*Mention Context*) and; ④ with the context model in the original entity-topic model (*Entity-Topic Model*). From Table 4 we can see that a fine-tuned context model will result in a 2~7% F1 improvement.

Context Model	**F1**
No Context	*0.73*
Article Content	*0.75*
Mention Context	*0.78*
Entity-Topic Model	***0.80***

Table 4. The *F1* using different context models

3) **The effects of joint model**. The third advantage of our model is that it *jointly* model the context compatibility and the topic coherence, which bring two benefits: ① the mutual reinforcement between the two directions can be captured in our model; ② the context compatibility and the topic coherence are uniformly modeled and jointly estimated, which makes the model more accurate for EL.

4.5.4 *EL Accuracies on TAC 2009 dataset*

We also compare our method with the top 5 EL systems in TAC 2009 and the two state-of-the-art systems (*EM-Model* and *EL-Graph*) on TAC 2009 data set in Figure 5 (For *EL-Graph* and our method, a NIL threshold is used to detect whether the referent entity is contained in the knowledge base, if the knowledge base not contains the referent entity, we assign the mention to a NIL entity). From Figure 5, we can see that our method is competitive: 1) Our method can achieve a 3.4% accuracy improvement over the best system in TAC 2009; 2) *Our method*, *EM-Model* and *EL-Graph* get very close accuracies (0.854, 0.86 and 0.838 correspondingly), we believe this is because: ① The mentions to be linked in TAC data set are mostly salient mentions; ② The influence of the NIL referent entity problem, i.e., the referent entity is not contained in the given knowledge base: Most referent entities (67.5%) on TAC 2009 are NIL entity and our method has no special handling on this problem, rather than other methods such as the *EM-Model*, which affects the overall performance of our method.

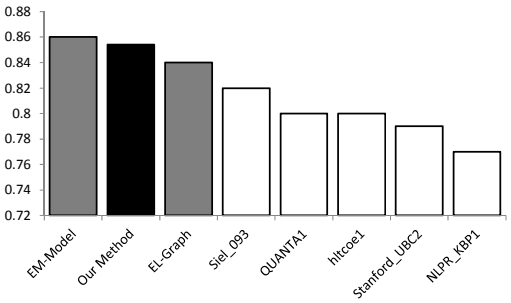

Figure 5. The EL accuracies on TAC 2009 dataset