4.5 Experimental Results

4.5.1 *Overall Performance*

We compared our method with all the above five baselines. For our method, we estimate the global knowledge using all the articles in the Jan. 30, 2010 English version of Wikipedia, and totally there were 3,083,158 articles. For each article, the mentions within it are detected using the methods described in Medelyan et al.(2008) and all terms in an article are used as context words, so a term may both be a mention and a context word. The topic number of our model is $T = 300$ (will be empirically set in Sect 4.5.2). To train the entity-topic model, we run 500 iterations of our Gibbs sampling algorithm to converge. The training time of our model is nearly one week on our server using 20 GB RAM and one core of 3.2 GHz CPU. Since the training can be done offline, we believe that the training time is not critical to the real-world usage as the online inference on new document is very quick. Using the above settings, the overall results are shown in Table 1.

	Precision	Recall	F1
Wikify!	*0.55*	*0.28*	*0.37*
EM-Model	*0.82*	*0.48*	*0.61*
M&W	*0.80*	*0.38*	*0.52*
CSAW	*0.65*	*0.73*	*0.69*
EL-Graph	*0.69*	*0.76*	*0.73*
Our Method	***0.81***	***0.80***	***0.80***

Table 1. The overall results on IITB data set

From the overall results in Table 1, we can see that:

1) By *jointly* modeling and exploiting the context compatibility and the topic coherence, our method can achieve competitive performance: ① compared with the context compatibility baselines *Wikify!* and *EM-Model,* our method correspondingly gets 43% and 19% F1 improvement; ② compared with the topic coherence baselines *M&W*, our method achieves 28% F1 improvement; ③ compared with the hybrid baselines *CSAW* and *EL-Graph*, our method correspondingly achieves 11% and 7% F1 improvement.

2) Compared with the *context compatibility only* and the *topic coherence only* methods, the main advantage of our method is that, rather than only achieved high entity linking precision on *salient* mentions, it can also effectively link the *non-salient* mentions in a document: this is demonstrated in our method's significant *Recall* improvement: a 32~52% Recall improvement over baselines *Wikify!*, *EM-Model* and *M&W*. We believe this is because a document usually contains little evidence for EL decisions on non-salient mentions, so with either only context compatibility or only topic coherence the evidence is not enough for EL decisions on these non-salient mentions, and bring these two directions together is critical for the accurate EL on these mentions.

3) Compared with the *hybrid* methods, the main advantage of our method is the improvement of EL precision (a 11~16% improvement over baselines *CSAW* and *EL-Graph*), we believe this is because: ① Our method can further capture the mutual reinforcement effect between the context compatibility and the topic coherence; ② The traditional hybrid methods usually determine the topic coherence of an entity to a document using *all entities* in the document, in comparison our method uses only *the entities in the same topic*, we believe this is more reasonable for EL decisions.

4.5.2 *Parameter Tuning*

One still parameter of our method is the topic number *T*. An appropriate *T* will distribute entities into well-organized topics, in turn it will capture the co-occurrence information of entities. Figure 4 plots the *F1* at different *T* values. We can see that the *F1* is not very sensitive to the topic number and with $T = 300$ our method achieves its best *F1* performance.

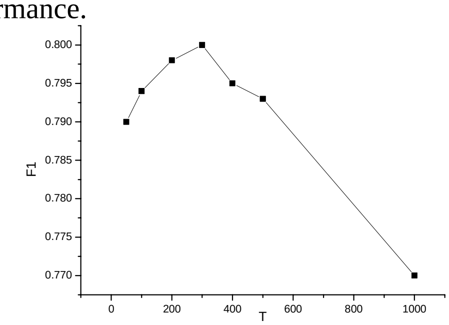

Figure 4. The *F1* vs. the topic number *T*

4.5.3 *Detailed Analysis*

In this section we analyze why and how our method works well in detail. Generally, we believe the main advantages of our method are:

1) **The effects of topic knowledge**. One main advantage of our model is that the topic knowledge