Thank you for your comments and suggestions.

Review #1

* In our ranking formulation of Task 2, it is more LIKELY that a higher ranked pair is physically near to each other than a lower ranked pair. In general KB like Freebase, the relation between entities is factual and thus can be treated as a classification problem. In real world, nearly any two physical objects can be co-located, but the second setting focuses on the TYPICALITY of such commonsense relation. (Mention ConceptNet weight?)
* Yes. The two tasks are independent because they are different problems in different settings. Task 1’s sentential relation classifier can be used as one of the solutions for Task 2.
* We trained our human annotators about the criterion of LocatedNear relation: whether the scene depicted by the sentence involves two physical objects that both locates in the same sight as if someone were observing it.
* Yes it is. We will clarify it more explicitly in camera-ready version, thank you!
* Yes and we will add the following results to the camera-ready version, 55.1% instances in our dataset are positive(LocatedNear):

Model Acc Pre Rec F1

Random 0.500 0.551 0.500 0.524

Majority 0.551 0.551 1.000 0.710

* ConceptNet only contains 49 pairs with LocatedNear relations in total, while our dataset contains 500 pairs. No matter what formulation we choose, ConceptNet cannot be compared with other results against our dataset due to lack of LocatedNear pairs in ConceptNet.

Review #2

* Although our topic is rather specific, it can benefit Computer Vision and NLP research, such as object detection in image, machine comprehension involving visible scene, improving machine translation quality (e.g. helps disambiguates the word “bank” through context word “river”) and other intelligent systems.
* This is actually exactly how we split the data. It is our writing issue, and we will clarify it more explicitly in the camera-ready version.
* Thank you for pointing out. We now make the SVM/LSTM comparison more fair by adding bag-of-word features of the whole sentence (BW) as well as bag-of-word features along shortest dependency path between two physical objects (BPW) to the SVM model, and thus providing the SVM sentence pattern. The following is the current result comparison:

Model Acc Pre Rec F1

SVM(all) 0.584 0.606 0.702 0.650

SVM(w/o. BW) 0.577 0.579 0.675 0.623

SVM(w/o. BPW) 0.556 0.567 0.681 0.619

SVM(w/o. BAP) 0.563 0.573 0.811 0.672

SVM(w/o. GF) 0.605 0.616 0.751 0.677

SVM(w/o. SDP) 0.579 0.597 0.728 0.656

SVM(w/o. SS) 0.584 0.605 0.708 0.652

DRNN 0.635 0.658 0.702 0.679

LSTM+Word 0.637 0.635 0.800 0.708

LSTM+POS 0.641 0.650 0.751 0.697

LSTM+Norm 0.653 0.654 0.784 0.713

The ablation tests also show that word-based features are less effective than

structure-based features, such as SDP, BPW. We also found that the bag-of-words features (BW) are the most important one.

* The baseline using basic summation (f0) is compared to ours as follows:

f MAP P@50 P@100 P@200 P@300

f0 0.42 0.40 0.44 0.42 0.38

f1 0.58 0.70 0.60 0.53 0.44

f2 0.48 0.56 0.52 0.49 0.42

f3 0.59 0.68 0.63 0.55 0.44

f4 0.56 0.40 0.48 0.50 0.42

One can see that our accumulative score outperforms the baseline in all aspects, justifying our intuition.

Review #3

* (HOWTO Rebut this? However, the purpose of the experiments is not entirely clear. I was struggling to identify the research questions underlying the experiments in section 4)
* The motivation and intuition inside our proposed sentence normalization method is to reduce the effect of unwanted semantics from some words irrelevant to LocatedNear relation in the sentence. For example, given two sentences in the paper “The king led the dog into his nice garden.” and “A criminal led the dog into a poor garden.”. In fact, whether the person is a king or criminal, and whether the garden is nice or poor doesn’t change the fact that the dog is being led by someone into a garden, and thus both sentences depict the scene that dog and garden are LocatedNear each other. So we normalized the words like “king” and “criminal”, “nice” and “poor” into something else that shares the semantics, which contributes to noise reduction of the representation learning process, making the representation learned by LSTM representing more on aspects of LocatedNear relation, rather than general semantics. This step also reduces vocabulary size which makes model work better with limited labeled data.
* Besides our sentence normalization processing leveraging irrelevant word canonicalization, the use of POS, lemma, dependency and position information is following the success of using such additional information in state-of-the-art general relation classification model (Xu et al., 2016, Zeng et al., 2014)
* We did do empirical analysis apart from the mere intuition. We showed in Table 5 that by comparing our though-more-complex normalization process (LSTM-Norm) with simple word-based models (LSTM-Word), our intuition worked as expected. The readers should be convinced by such empirical comparison between all benchmarks.