

Spoiler Detection in Book Reviews

Jue Wang

NetID: jwr0983

Description

This project aims to detect sentences in user reviews that have high probability of containing spoilers. Within one user review only 1-2 sentences are actual spoilers, and book/movie websites can give viewer the option to mask the detected spoiler reviews partially or in full.

Dataset Description

The Goodreads Book Review Dataset

Link: <https://sites.google.com/eng.ucsd.edu/ucsdbookgraph/home>

As the following snippet illustrates, this dataset contains sentence-level reviews, as well as whether the sentence contains a spoiler (0/1). The dataset contains more than 1.3 million book reviews, where each book/user has at least one associated spoiler review. I might randomly choose a subset of the data to reduce training time or a specific genre such as fiction where spoiler review detection are more relevant.

```
{ 'user_id': '01ecla320ffded6b2dd47833f2c8e4fb',  
  'timestamp': '2016-01-14',  
  'review_sentences': [[0, '3.5 - 4 Stars'],  
    [0,  
      'This is the very sexy and very sweet sequel to the short story, Hearts in Darkness.'],  
    [0,  
      'This book picks up shortly after Caden and Makenna first met in the darkened elevator.'],  
    [0,  
      'They have developed an almost-perfect relationship - and this book is verra verra steamy - but Caden  
has severe PTSD due to a very shocking incident in his past that has completely traumatized him.'],  
    [0,  
      "This is one of those stories where the hero doesn't think he is good enough for the heroine - she is  
smart, kind, beautiful, has a wonderful family, etc."],  
    [0, '- so the hero must distance himself from her "for her own good."'],  
    [0,  
      'Makenna loves Caden but realizes that he is gun-shy and she is scared to make her true feelings  
known.'],  
    [0, 'When she does, all hell breaks loose.'],  
    [0,  
      'This is a lovely, steamy story with lots of angst and a beautiful HEA.'],  
    [0,  
      "If you are a fan of Laura Kaye's writing, then you will love this book."],  
    [0, '{ARC provided in return for an honest review.}']],  
  'rating': 4,  
  'has_spoiler': False,  
  'book_id': '25396316',  
  'review_id': '27f0566f9f72429e4d25bbbf584434b4' }
```