<table>
<tr><th rowspan="2">cases</th><th>doc_1</th><th>doc_2</th><th>decision</th><th>id</th></tr>
<tr><td>
<table>
<tr><th>authors</th><td>• Hao Wang
• Yangguang Li
• Zhen Huang
• Yong Dou
• Lingpeng Kong
• Jing Shao</td></tr>
<tr><th>title</th><td>SNCSE: Contrastive Learning for Unsupervised Sentence Embedding with Soft Negative Samples</td></tr>
<tr><th>publication_date</th><td>2022-01-16 06:15:43+00:00</td></tr>
<tr><th>source</th><td>SupportedSources.ARXIV</td></tr>
<tr><th>journal</th><td>None</td></tr>
<tr><th>volume</th><td></td></tr>
<tr><th>doi</th><td></td></tr>
<tr><th>urls</th><td>• http://arxiv.org/pdf/2201.05979v3
• http://arxiv.org/abs/2201.05979v3
• http://arxiv.org/pdf/2201.05979v3</td></tr>
<tr><th>id</th><td>id-2406434478053389508</td></tr>
<tr><th>abstract</th><td>Unsupervised sentence embedding aims to obtain the most appropriate embedding for a sentence to reflect its semantic. Contrastive learning has been attracting developing attention. For a sentence, current models utilize diverse data augmentation methods to generate positive samples, while consider other independent sentences as negative samples. Then they adopt InfoNCE loss to pull the embeddings of positive pairs gathered, and push those of negative pairs scattered. Although these models have made great progress on sentence embedding, we argue that they may suffer from feature suppression. The models fail to distinguish and decouple textual similarity and semantic similarity. And they may overestimate the semantic similarity of any pairs with similar textual regardless of the actual semantic difference between them. This is because positive pairs in unsupervised contrastive learning come with similar and even the same textual through data augmentation. To alleviate feature suppression, we propose contrastive learning for unsupervised sentence embedding with soft negative samples (SNCSE). Soft negative samples share highly similar textual but have surely and apparently different semantic with the original samples. Specifically, we take the negation of original sentences as soft negative samples, and propose Bidirectional Margin Loss (BML) to introduce them into traditional contrastive learning framework, which merely involves positive and negative samples. Our experimental results show that SNCSE can obtain state-of-the-art performance on semantic textual similarity (STS) task with average Spearman's correlation coefficient of 78.97% on BERTbase and 79.23% on RoBERTabase. Besides, we adopt rank-based error analysis method to detect the weakness of SNCSE for future study.</td></tr>
<tr><th>versions</th><td></td></tr>
</table>
</td><td>
<table>
<tr><th>authors</th><td>• Hao Wang
• Yangguang Li
• Zhen Huang
• Yong Dou
• Lingpeng Kong
• J. Shao</td></tr>
<tr><th>title</th><td>SNCSE: Contrastive Learning for Unsupervised Sentence Embedding with Soft Negative Samples</td></tr>
<tr><th>publication_date</th><td>2022-01-16 00:00:00</td></tr>
<tr><th>source</th><td>SupportedSources.SEMANTIC_SCHOLAR</td></tr>
<tr><th>journal</th><td>ArXiv</td></tr>
<tr><th>volume</th><td>abs/2201.05979</td></tr>
<tr><th>doi</th><td></td></tr>
<tr><th>urls</th><td>• https://www.semanticscholar.org/paper/995ee983a7dd66bf30408a5d1c1a8c3684431998</td></tr>
<tr><th>id</th><td>id244352396605885208</td></tr>
<tr><th>abstract</th><td>Unsupervised sentence embedding aims to obtain the most appropriate embedding for a sentence to reflect its semantic. Contrastive learning has been attracting developing attention. For a sentence, current models utilize diverse data augmentation methods to generate positive samples, while consider other independent sentences as negative samples. Then they adopt InfoNCE loss to pull the embeddings of positive pairs gathered, and push those of negative pairs scattered. Although these models have made great progress on sentence embedding, we argue that they may suffer from feature suppression. The models fail to distinguish and decouple textual similarity and semantic similarity. And they may overestimate the semantic similarity of any pairs with similar textual regardless of the actual semantic difference between them. This is because positive pairs in unsupervised contrastive learning come with similar and even the same textual through data augmentation. To alleviate feature suppression, we propose contrastive learning for unsupervised sentence embedding with soft negative samples (SNCSE). Soft negative samples share highly similar textual but have surely and apparently different semantic with the original samples. Specifically, we take the negation of original sentences as soft negative samples, and propose Bidirectional Margin Loss (BML) to introduce them into traditional contrastive learning framework, which merely involves positive and negative samples. Our experimental results show that SNCSE can obtain state-of-the-art performance on semantic textual similarity (STS) task with average Spearmanâ€™s correlation coefficient of 78.97% on BERTbase and 79.23% on RoBERTabase. Besides, we adopt rank-based error analysis method to detect the weakness of SNCSE for future study. âˆ—Work was done as an internship in Sensetime. Communication E-mail: wanghao2@sensetime.com</td></tr>
<tr><th>versions</th><td></td></tr>
</table>
</td><td>DUPLICATES</td><td>160</td></tr>
</table>