

V této práci představujeme metodu strojového učení modelů emedující dokumenty, která není náročná na výpočetní zdroje ani nevyžaduje anotovaná trénovací data. S přístupem učitele a studenta, distilujeme kapacitu SBERTa zaznamenat strukturu textu a schopnost Paragraph Vektoru zpracovat dlouhé dokumenty do našeho výsledného embedovacího modelu. Naši metodu testujeme na Longformeru, Transformeru s řídkou attention vrstvou, který je schopný zpracovat dokumenty dlouhé až 4096 tokenů. Prozkoumáme několik ztrátových funkcí, které nutí studenta (Longformera) napodobovat výstupy obou učitelů (SBERTa a Paragraph Vektoru). V experimentech ukazujeme, že i přes omezený kontext SBERTa, je distilace jeho výstupů pro výkon studenta zásadnější. Nicméně student dokáže získat prospěch z obou učitelů. Naše metoda vylepšuje výsledek Longformera na osmi úlohách, které zahrnují predikci citace, detekci plagiarismu i vyhledávání na základě podobnosti dokumentů. Naše metoda se navíc ukazuje jako obzvláště účinná v situacích s málo dotrénovávacími daty, kde námi natrénovaný student překoná i oba učitele. Podobným výkonem odlišně natrénovaných studentů ukazujeme, že naše metoda je robustní vůči různým změnám, a navrhuje možné oblasti budoucího výzkumu.