

各位，

我利用汇联提供的 8794 句（主要包含五言（1264）和七言（4036）的对联，其它有长有短）对联用统计机器翻译模型(SMT)进行字级别的训练，以下是我在网上找的一些测试集，我随便挑的，好像效果都不错，看来在句子较少的情况下，字级别可以 handle 更多的 state 转移，给大家简单展示一下，明天我再想一想收集更多的语料再训练一个更好点 model 给大家展示下效果。

（一）五言春联（预测句只选择含五个字的）：

-----  
原句：一声爆竹响，万里春光明

预测：

一声爆竹响，万盏梅花响。

一声爆竹响，梅花响万盏。

原句：人间多美景，大地足春光。

预测：

人间多美景，天上尽春光。

人间多美景，世上尽春光。

人间多美景，天下尽春光。

（后面是各种天下，世上，尽，天上的次序颠倒的组合）

原句：春光迎盛世，旭日耀新春。

预测：

春光迎盛世，雄舞庆新春。

春光迎盛世，福舞庆新春。

春光迎盛世，马舞庆新春。

（二）七言春联（预测句只选择含七个字的）：

-----  
原句：人逢盛世千家乐，户沐春阳万事兴。

预测：

人逢盛世千家乐，岁值华年万户欢。

人逢盛世千家乐，岁值华年万众心。

人逢盛世千家乐，岁值华年万里春。

原句：大地寒尽春光好，普天日丽前景佳。

预测：

大地寒尽春光好，神州尽喜气象新。

大地寒尽春光好，神州尽喜气象新。

大地寒尽春光好，人间尽喜气象新。

大地寒尽春光好，神州喜尽事业兴。

（后面是各种人间，神州，尽，喜，事业的次序颠倒的组合）

(三) 其它春联：（九言的只有 199 句，也能训练到以下程度）

-----  
原句：国泰民安一岁胜一岁，人寿年丰一年强一年。

预测：

国泰民安一岁胜一岁，长天日丽万程万年春。

国泰民安一岁胜一岁，长天日丽万年程万年。

国泰民安一岁胜一岁，长天日丽万年美万年。

原句：勤奋学习只是天天长，努力工作幸福年年增。

预测：

勤奋学习知识天天长，马德智优大文明日日。

勤奋学习知识天天长，马德智优大地地文明。

勤奋学习知识天天长，优马德文明日日大地。