目录结构：

ChSegSuite

```
GenTrain/               // 分别用于生成专有格式训练语料格式转换程序
PerTrain/               // 训练程序
PerSeg/                 // 分词程序
train_seg.txt           // 标准格式训练语料
test.txt                // 测试集
```

数据格式：

train_seg.txt 和 test.txt 遵循相同的数据格式，每行一句，词与词之间以空格隔开

生成专有格式训练语料（目录 GenTrain）：

```
g++ GenTrain.cpp -o gentrain  // 编译
./gentrain train_seg.txt train_seg_bmes.txt n   // 格式转换，n 指明不标注只切分
```

train_seg_bmes.txt 即为专有格式的训练语料

训练（目录 PerTrain）：

```
make clean; make              // 编译
./src/pertrain -f train_seg_bmes.txt -1 7       // 训练，-1 指明训练迭代 7 轮
```

work 中保存了各轮迭代后的感知机分词模型。前缀 avg 指明平均参数模型

分词（目录 PerSeg）：

首先将训练步骤得到的某一模型拷贝到 data 目录下，并重命名为 model。一般取第 7 轮迭代后的平均参数模型

```
make clean; make              // 编译
./src/perseg -s test.txt -t result.txt          // 切分
```

result.txt 即为 test.txt 的切分结果