

Domain	# Sent	# Asp	# non-asp	words
Computer	556	1173	7675	
Camera	609	1646	9849	
Router	509	1239	7264	
Phone	497	986	7178	
Speaker	516	1299	7546	
DVD Player	506	928	7552	
Mp3 Player	503	1186	7607	

Table 2: Annotation details of the datasets

a particular time, it is assumed M has been applied to n past domains, and is now faced with the $n + 1$ domain. L-CRF uses M and reliable aspects (denoted K_{n+1}) mined from S and K^t ($K = K^t \cup K_{n+1}$) to extract from D_{n+1} . Note that aspects K_t from the training data are considered always reliable as they are manually labeled, thus a subset of K . We cannot use all extracted aspects from past domains as reliable aspects due to many extraction errors. But those aspects that appear in multiple past domains are more likely to be correct. Thus K contains those frequent aspects in S . The lifelong extraction phase is in Algorithm 1.

Lifelong Extraction Phase: Algorithm 1 performs extraction on D_{n+1} iteratively.

1. It generates features (F) on the data D_{n+1} (line 3), and applies the CRF model M on F to produce a set of aspects A_{n+1} (line 4).
2. A_{n+1} is added to S , the past aspect store. From S , we mine a set of frequent aspects K_{n+1} . The frequency threshold is λ .
3. If K_{n+1} is the same as K_p from the previous iteration, the algorithm exits as no new aspects can be found. We use an iterative process because each extraction gives new results, which may increase the size of K , the reliable past aspects or past knowledge. The increased K may produce more dependency patterns, which can enable more extractions.
4. Else: some additional reliable aspects are found. M may extract additional aspects in the next iteration. Lines 10 and 11 update the two sets for the next iteration.

5 Experiments

We now evaluate the proposed L-CRF method and compare with baselines.

5.1 Evaluation Datasets

We use two types of data for our experiments. The first type consists of seven (7) annotated benchmark review datasets from 7 domains (types of products). Since they are annotated, they are used in training and testing. The first 4 datasets are from (Hu and Liu, 2004), which actually has 5 datasets from 4 domains. Since we are mainly interested in results at the domain level, we did not use one of the domain-repeated datasets. The last 3 datasets of three domains (products) are from (Liu et al., 2016). These datasets are used to make up our CRF training data D^t and test data D_{n+1} . The annotation details are given in Table 2.

The second type has 50 unlabeled review datasets from 50 domains or types of products (Chen and Liu, 2014). Each dataset has 1000 reviews. They are used as the past domain data, i.e., D_1, \dots, D_n ($n = 50$). Since they are not labeled, they cannot be used for training or testing.

5.2 Baseline Methods

We compare L-CRF with CRF. We will not compare with unsupervised methods, which have been shown improvable by lifelong learning (Chen et al., 2014; Liu et al., 2016). The frequency threshold λ in Algorithm 1 used in our experiment to judge which extracted aspects are considered reliable is empirically set to 2.

CRF: We use the linear chain CRF from ². Note that CRF uses all features including dependency features as the proposed L-CRF but does not employ the 50 domains unlabeled data used for lifelong learning

CRF+R: It treats the reliable aspect set K as a dictionary. It adds those reliable aspects in K that are not extracted by CRF but are in the test data to the final results. We want to see whether incorporating K into the CRF extraction through dependency patterns in L-CRF is actually needed.

We do not compare with domain adaptation or transfer learning because domain adaption basically uses the source domain labeled data to help learning in the target domain with few or no labeled data. Our 50 domains used in lifelong learning have no labels. So they cannot help in transfer learning. Although in transfer learning, the target domain usually has a large quantity of unlabeled data, but the 50 domains are not used as the target domains in our experiments.

²<https://github.com/huangzhengsjtu/pcrf/>