

### Multimedia Appendix 3. Effects of Features on Performance of Adapted Distant Supervision

Table A3-1. Performance of different ADS-fsa<sup>a</sup> systems implemented by using all types of features or by respectively dropping each individual type of features, under 4 conditions using 100, 200, 500, and 1000 target-domain training examples<sup>b</sup>.

ADS-fsa system	AUC-ROC <sup>c</sup>				Average Precision			
	100	200	500	1000	100	200	500	1000
ADS-fsa-ALL <sup>d</sup>	0.746	0.756	0.776	0.790	0.815	0.823	0.839	0.850
ADS-fsa-woWE <sup>e</sup>	0.702 <i>t</i> =30.50 <i>P</i> <.001	0.711 <i>t</i> =33.01 <i>P</i> <.001	0.721 <i>t</i> =64.88 <i>P</i> <.001	0.730 <i>t</i> =125.9 8, <i>P</i> <.001	0.772 <i>t</i> =25.66 <i>P</i> <.001	0.778 <i>t</i> =33.36 <i>P</i> <.001	0.788 <i>t</i> =59.58 <i>P</i> <.001	0.795 <i>t</i> =105.4 2, <i>P</i> <.001
ADS-fsa-woSem <sup>f</sup>	0.749 <i>t</i> =2.69 <i>P</i> =.01	0.756	0.774 <i>t</i> =3.32 <i>P</i> =.001	0.785 <i>t</i> =13.82 <i>P</i> <.001	0.821 <i>t</i> =4.98 <i>P</i> <.001	0.826 <i>t</i> =2.97 <i>P</i> =.004	0.840	0.848 <i>t</i> =5.24 <i>P</i> <.001
ADS-fsa-woATR <sup>g</sup>	0.746	0.756	0.776	0.790	0.816	0.823	0.839	0.850
ADS-fsa-woTF <sup>h</sup>	0.735 <i>t</i> =10.72 <i>P</i> <.001	0.746 <i>t</i> =11.98 <i>P</i> <.001	0.767 <i>t</i> =15.59 <i>P</i> <.001	0.781 <i>t</i> =27.86 <i>P</i> <.001	0.809 <i>t</i> =5.69 <i>P</i> <.001	0.818 <i>t</i> =6.69 <i>P</i> <.001	0.835 <i>t</i> =8.06 <i>P</i> <.001	0.845 <i>t</i> =14.68 <i>P</i> <.001
ADS-fsa-woTL <sup>i</sup>	0.736 <i>t</i> =10.67 <i>P</i> <.001	0.748 <i>t</i> =8.61 <i>P</i> <.001	0.768 <i>t</i> =17.79 <i>P</i> <.001	0.782 <i>t</i> =40.83 <i>P</i> <.001	0.803 <i>t</i> =15.54 <i>P</i> <.001	0.813 <i>t</i> =12.42 <i>P</i> <.001	0.830 <i>t</i> =19.97 <i>P</i> <.001	0.842 <i>t</i> =43.03 <i>P</i> <.001

<sup>a</sup>ADS-fsa: adapted distant supervision-feature space augmentation.

<sup>b</sup>We report the *P* values (if the *P* value  $\leq .05$ ) and the corresponding *t*<sub>99</sub> values for differences between each implementation and ADS-fsa-ALL.

<sup>c</sup>AUC-ROC: area under the receiver operating characteristic curve.

<sup>d</sup>ADS-fsa-ALL: ADS-fsa with all types of features.

<sup>e</sup>ADS-fsa-woWE: ADS-fsa without word embedding.

<sup>f</sup>ADS-fsa-woSem: ADS-fsa without semantic features.

<sup>g</sup>ADS-fsa-woATR: ADS-fsa without features derived from automatic term recognition.

<sup>h</sup>ADS-fsa-woGTF: ADS-fsa without general-domain term frequency.

<sup>i</sup>ADS-fsa-woTL: ADS-fsa without term length.

Table A3-2. Performances of different ADS<sup>a</sup> systems implemented by using baseline features derived from automatic term recognition, general-domain term frequency, and term length or by adding semantic features to the baseline features, under 4 conditions using 100, 200, 500, and 1000 target-domain training examples<sup>b</sup>.

ADS system	AUC-ROC <sup>c</sup>				Average Precision			
	100	200	500	1000	100	200	500	1000
ADS-fsa-base <sup>d</sup>	0.642	0.650	0.664	0.669	0.660	0.685	0.730	0.744
ADS-fsa-base+Sem <sup>e</sup>	0.702	0.711	0.721	0.730	0.772	0.778	0.788	0.795
<i>t</i> <sub>99</sub> and <i>P</i> values (ADS-fsa-base+Sem vs. ADS-fsa-base)	<i>t</i> =32.26 <i>P</i> <.001	<i>t</i> =35.11 <i>P</i> <.001	<i>t</i> =41.38 <i>P</i> <.001	<i>t</i> =75.90 <i>P</i> <.001	<i>t</i> =31.86 <i>P</i> <.001	<i>t</i> =20.07 <i>P</i> <.001	<i>t</i> =14.58 <i>P</i> <.001	<i>t</i> =25.56 <i>P</i> <.001
ADS-sds-base <sup>f</sup>	0.657	0.665	0.669	0.671	0.713	0.734	0.743	0.747
ADS-sds-base+Sem <sup>g</sup>	0.711	0.718	0.726	0.733	0.780	0.785	0.793	0.799
<i>t</i> <sub>99</sub> and <i>P</i> values (ADS-sds-base+Sem vs. ADS-sds-base)	<i>t</i> =21.38 <i>P</i> <.001	<i>t</i> =33.92 <i>P</i> <.001	<i>t</i> =58.62 <i>P</i> <.001	<i>t</i> =128.11 <i>P</i> <.001	<i>t</i> =12.74 <i>P</i> <.001	<i>t</i> =13.06 <i>P</i> <.001	<i>t</i> =20.33 <i>P</i> <.001	<i>t</i> =83.51 <i>P</i> <.001

<sup>a</sup>ADS: adapted distant supervision.

<sup>b</sup>We report the *P* values and the corresponding *t*<sub>99</sub> values between ADS-fsa-base and ADS-fsa-base+Sem and between ADS-sds-base and ADS-sds-base+Sem.

<sup>c</sup>AUC-ROC: area under the receiver operating characteristic curve.

<sup>d</sup>ADS-fsa-base: ADS-feature space augmentation using baseline features derived from automatic term recognition, general-domain term frequency, and term length

<sup>e</sup>ADS-fsa-base+Sem: ADS-feature space augmentation using baseline features plus semantic features.

<sup>f</sup>ADS-sds-base: ADS-supervised distant supervision using baseline features derived from automatic term recognition, general-domain term frequency, and term length.

<sup>g</sup>ADS-sds-base+Sem: ADS-supervised distant supervision using baseline features plus semantic features.