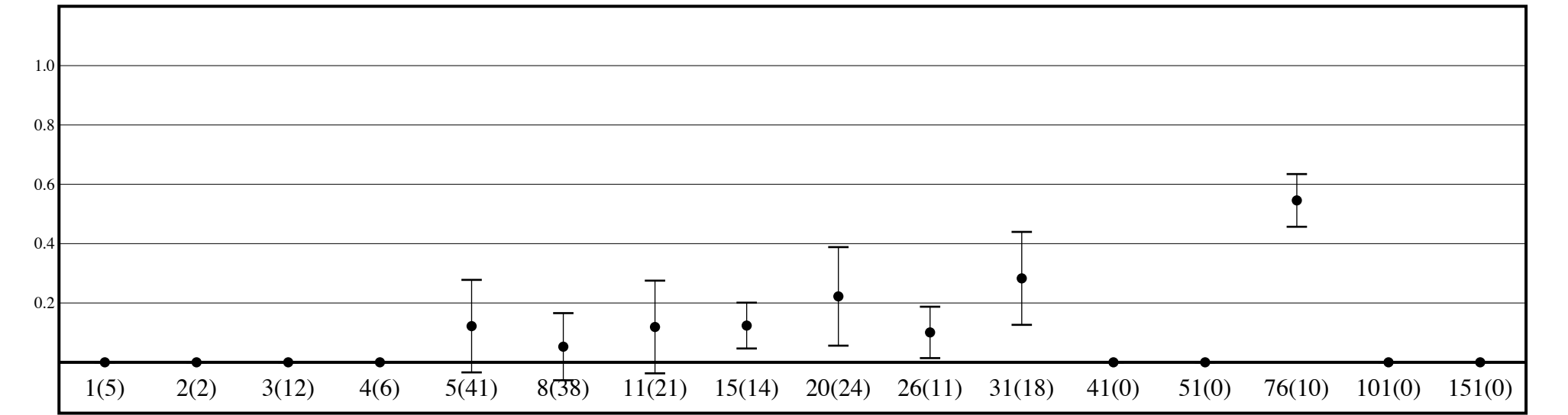


Test Results Categorized by Training Count

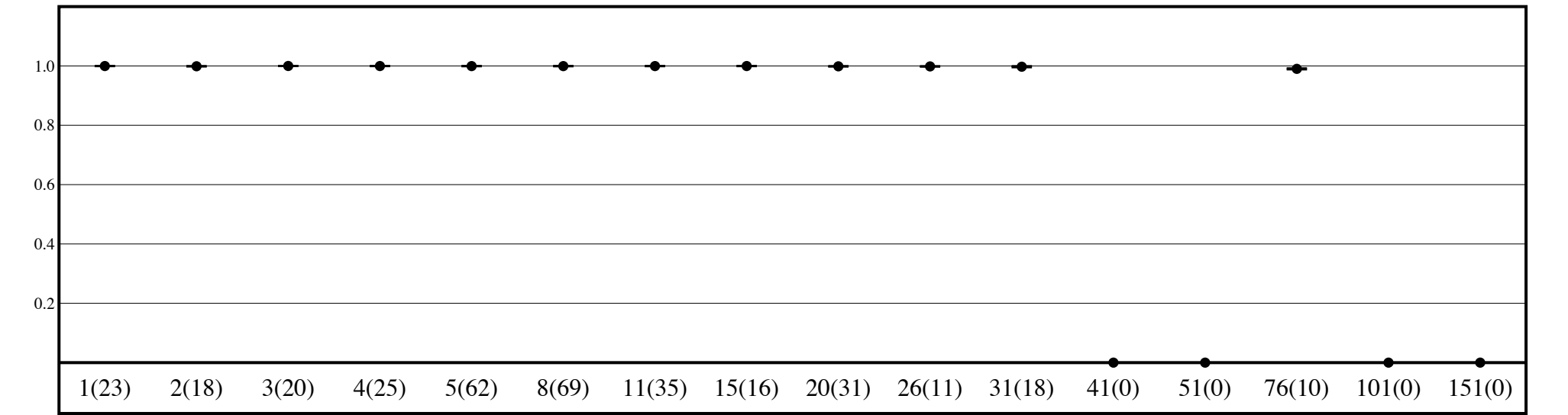
29257050: Ratio for n is determined using number of distinct tokens from each pool, using 'not tag' as background.

These graphs display the cross validation test results for a classifier arranged by the number of items a classifier was trained on for a given tag. The X axis displays the number of training items, where there is a gap, the label defines the lower bound of a range that extends to the next label. The number in brackets is the number of tags that fall into that 'bin'. The Y Axis is the mean true positive or true negative rate for the classifiers that fall in the 'bin'. Standard Deviation is shown as error bars on the chart.

True Positive Results



True Negative Results



Tags in bins

Bin	Tags
1..1	ande-nonprofit, ande-spam, ande-birds, ande-astronomy, ande-cats,
2..2	ande-machine learning, ande-spam, ande-birds, ande-astronomy,
3..3	ande-machine learning, ande-spam, ande-earth, ande-zoology, ande-music,
4..4	ande-machine learning, ande-earth, ande-zoology, ande-racial diversity, ande-music,
5..7	ande-gay rights, ande-censorship, ande-racial diversity, ande-music, ande-anthropology, ande-vista, ande-medicine, ande-Macs, ande-photography,

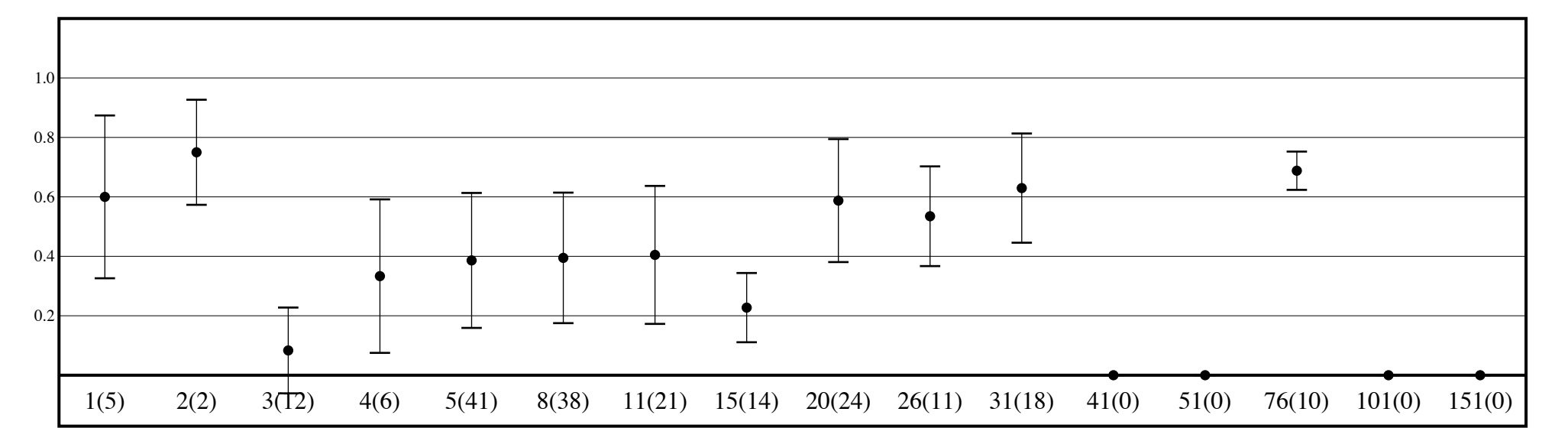
Bin	Tags
	ande-women in politics, ande-global warming,
8..10	ande-education, ande-gay rights, ande-censorship, ande-collaboration, ande-vista, ande-medicine, ande-Macs, ande-food, ande-cultural diversity, ande-women in politics, ande-privacy, ande-global warming,
11..14	ande-education, ande-language, ande-collaboration, ande-food, ande-cultural diversity, ande-literature,
15..19	ande-the media, ande-humor, ande-language, ande-evolution, ande-search engine,
20..25	ande-the media, ande-humor, ande-evolution, ande-search engine, ande-environment,
26..30	ande-social networking, ande-environment,
31..40	ande-social networking, ande-green,
41..50	
51..75	
76..100	ande-web,
101..150	
151..1000	

Test Results Categorized by Training Count

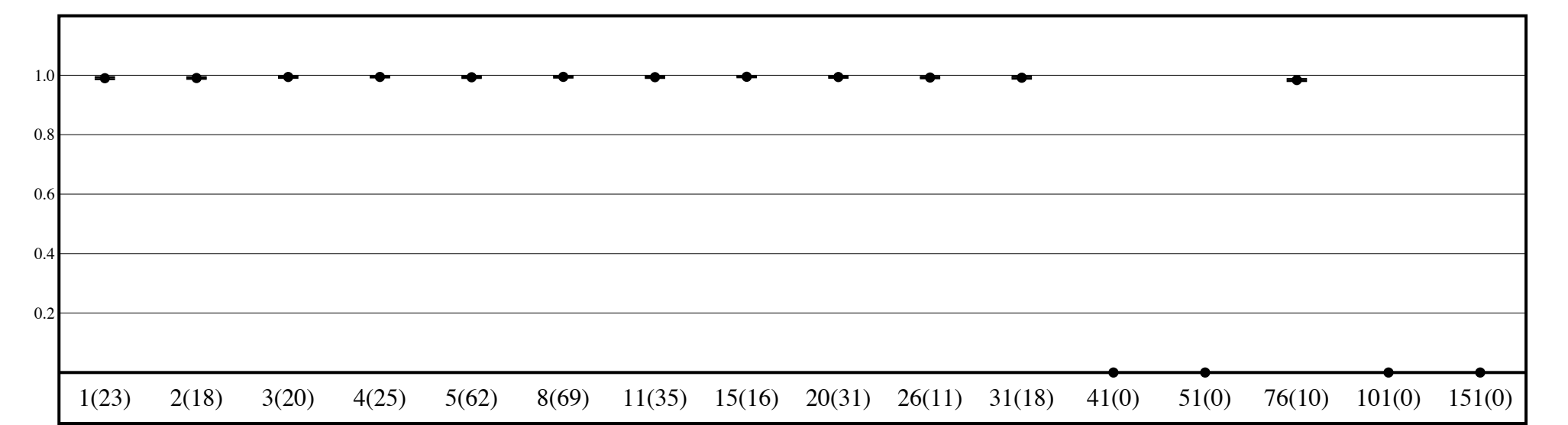
29257090: Ratio for n is determined using total number of tokens from each pool, using 'not tag' as background.

These graphs display the cross validation test results for a classifier arranged by the number of items a classifier was trained on for a given tag. The X axis displays the number of training items, where there is a gap, the label defines the lower bound of a range that extends to the next label. The number in brackets is the number of tags that fall into that 'bin'. The Y Axis is the mean true positive or true negative rate for the classifiers that fall in the 'bin'. Standard Deviation is shown as error bars on the chart.

True Positive Results



True Negative Results



Tags in bins

Bin	Tags
1..1	ande-nonprofit, ande-spam, ande-birds, ande-astronomy, ande-cats,
2..2	ande-machine learning, ande-spam, ande-birds, ande-astronomy,
3..3	ande-machine learning, ande-spam, ande-earth, ande-zoology, ande-music,
4..4	ande-machine learning, ande-earth, ande-zoology, ande-racial diversity, ande-music,
5..7	ande-gay rights, ande-censorship, ande-racial diversity, ande-music, ande-anthropology, ande-vista, ande-medicine, ande-Macs, ande-photography,

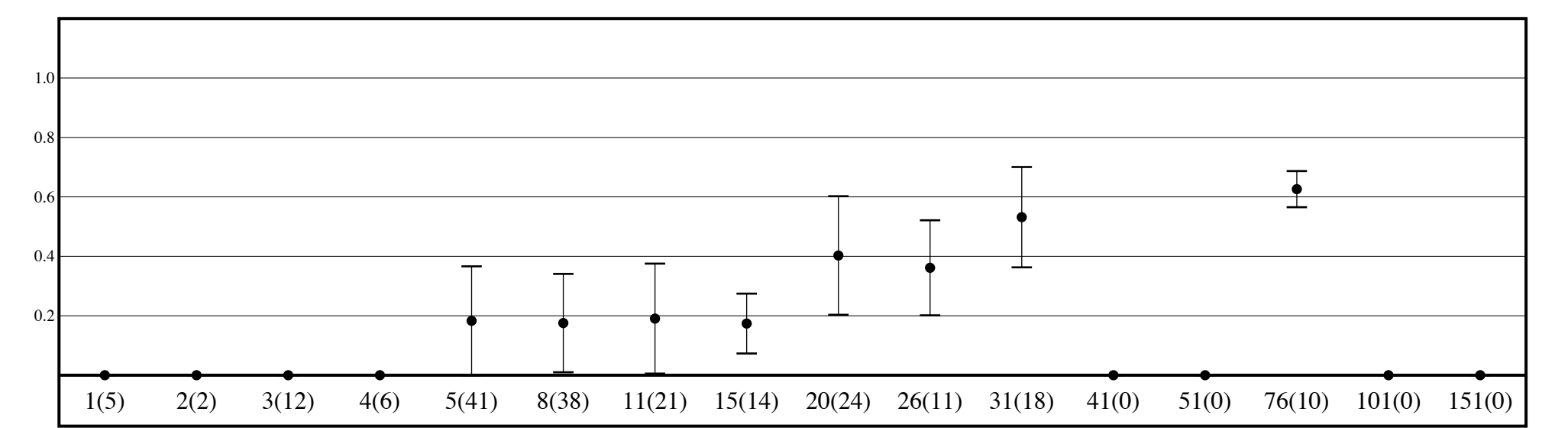
Bin	Tags
	ande-women in politics, ande-global warming,
8..10	ande-education, ande-gay rights, ande-censorship, ande-collaboration, ande-vista, ande-medicine, ande-Macs, ande-food, ande-cultural diversity, ande-women in politics, ande-privacy, ande-global warming,
11..14	ande-education, ande-language, ande-collaboration, ande-food, ande-cultural diversity, ande-literature,
15..19	ande-the media, ande-humor, ande-language, ande-evolution, ande-search engine,
20..25	ande-the media, ande-humor, ande-evolution, ande-search engine, ande-environment,
26..30	ande-social networking, ande-environment,
31..40	ande-social networking, ande-green,
41..50	
51..75	
76..100	ande-web,
101..150	
151..1000	

Test Results Categorized by Training Count

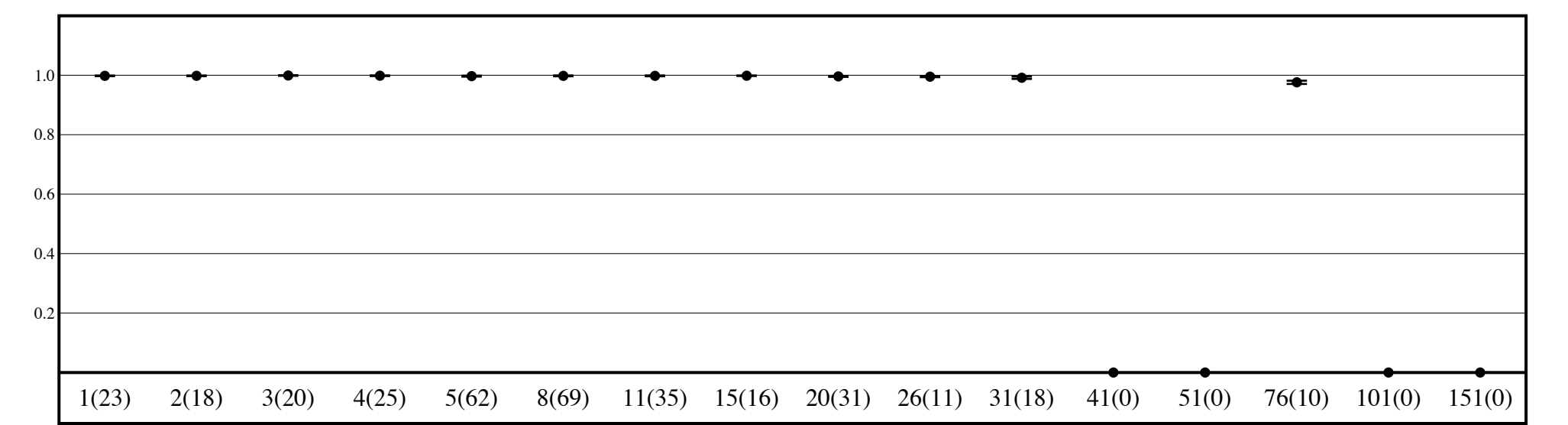
29257050: Ratio for n is determined using number of distinct tokens from each pool, using 'seen' as background.

These graphs display the cross validation test results for a classifier arranged by the number of items a classifier was trained on for a given tag. The X axis displays the number of training items, where there is a gap, the label defines the lower bound of a range that extends to the next label. The number in brackets is the number of tags that fall into that 'bin'. The Y Axis is the mean true positive or true negative rate for the classifiers that fall in the 'bin'. Standard Deviation is shown as error bars on the chart.

True Positive Results



True Negative Results



Tags in bins

Bin	Tags
1..1	ande-nonprofit, ande-spam, ande-birds, ande-astronomy, ande-cats,
2..2	ande-machine learning, ande-spam, ande-birds, ande-astronomy,
3..3	ande-machine learning, ande-spam, ande-earth, ande-zoology, ande-music,
4..4	ande-machine learning, ande-earth, ande-zoology, ande-racial diversity, ande-music,
5..7	ande-gay rights, ande-censorship, ande-racial diversity, ande-music, ande-anthropology, ande-vista, ande-medicine, ande-Macs, ande-photography,

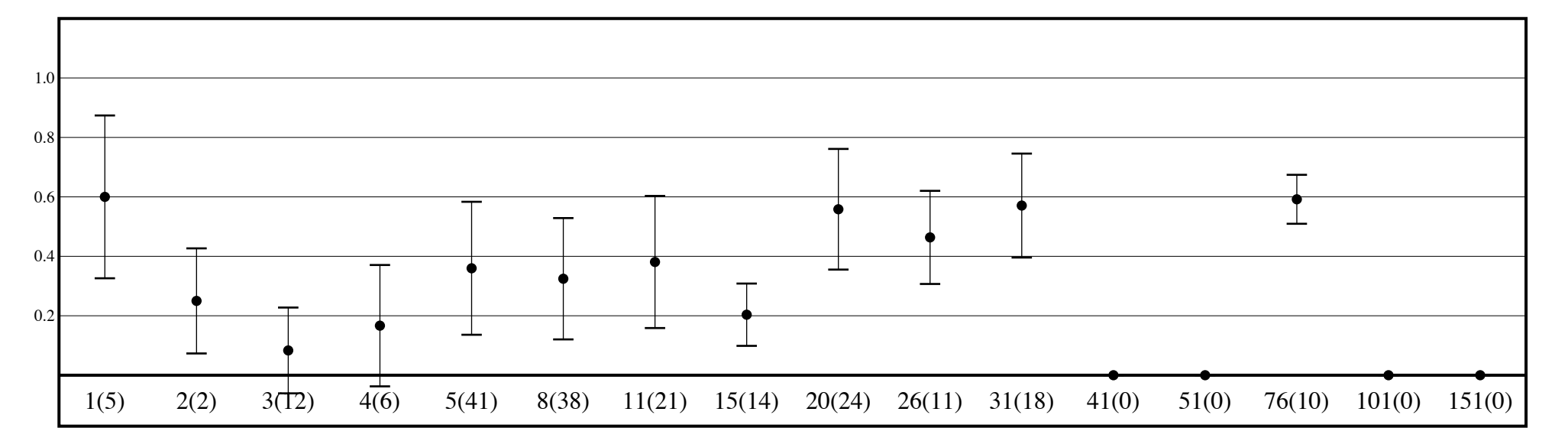
Bin	Tags
	ande-women in politics, ande-global warming,
8..10	ande-education, ande-gay rights, ande-censorship, ande-collaboration, ande-vista, ande-medicine, ande-Macs, ande-food, ande-cultural diversity, ande-women in politics, ande-privacy, ande-global warming,
11..14	ande-education, ande-language, ande-collaboration, ande-food, ande-cultural diversity, ande-literature,
15..19	ande-the media, ande-humor, ande-language, ande-evolution, ande-search engine,
20..25	ande-the media, ande-humor, ande-evolution, ande-search engine, ande-environment,
26..30	ande-social networking, ande-environment,
31..40	ande-social networking, ande-green,
41..50	
51..75	
76..100	ande-web,
101..150	
151..1000	

Test Results Categorized by Training Count

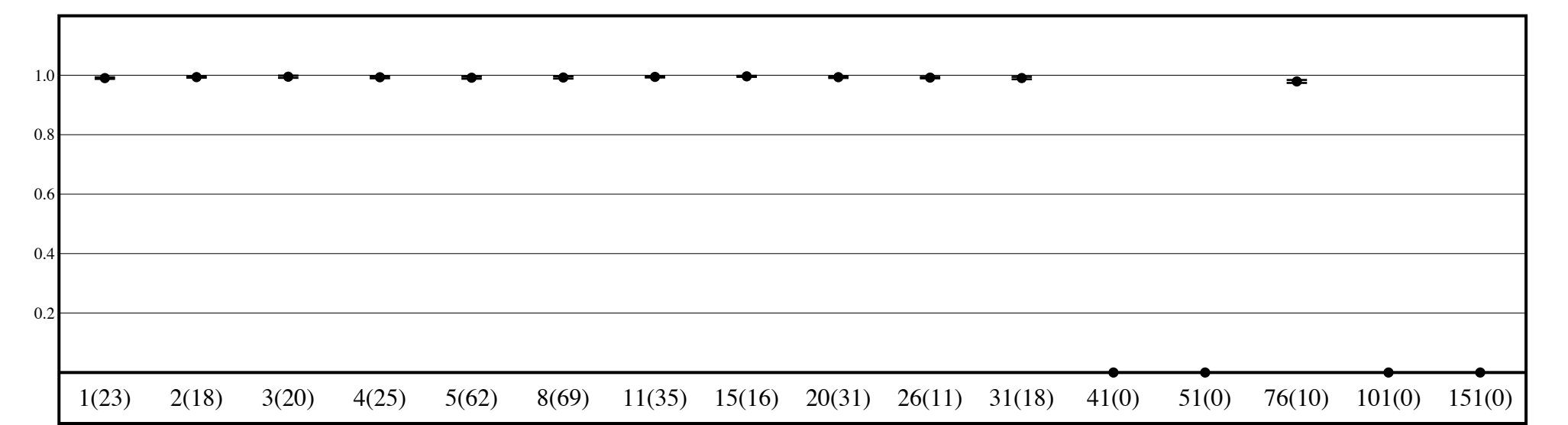
29257090: Ratio for n is determined using total number of tokens from each pool, using 'seen' as background.

These graphs display the cross validation test results for a classifier arranged by the number of items a classifier was trained on for a given tag. The X axis displays the number of training items, where there is a gap, the label defines the lower bound of a range that extends to the next label. The number in brackets is the number of tags that fall into that 'bin'. The Y Axis is the mean true positive or true negative rate for the classifiers that fall in the 'bin'. Standard Deviation is shown as error bars on the chart.

True Positive Results



True Negative Results



Tags in bins

Bin	Tags
1..1	ande-nonprofit, ande-spam, ande-birds, ande-astronomy, ande-cats,
2..2	ande-machine learning, ande-spam, ande-birds, ande-astronomy,
3..3	ande-machine learning, ande-spam, ande-earth, ande-zoology, ande-music,
4..4	ande-machine learning, ande-earth, ande-zoology, ande-racial diversity, ande-music,
5..7	ande-gay rights, ande-censorship, ande-racial diversity, ande-music, ande-anthropology, ande-vista, ande-medicine, ande-Macs, ande-photography,

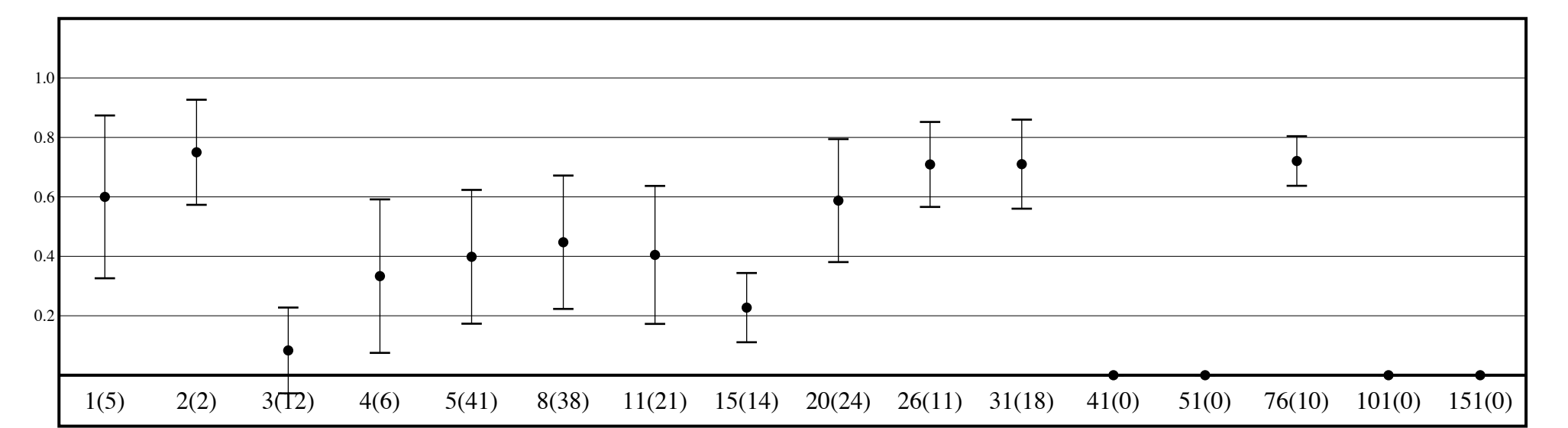
Bin	Tags
	ande-women in politics, ande-global warming,
8..10	ande-education, ande-gay rights, ande-censorship, ande-collaboration, ande-vista, ande-medicine, ande-Macs, ande-food, ande-cultural diversity, ande-women in politics, ande-privacy, ande-global warming,
11..14	ande-education, ande-language, ande-collaboration, ande-food, ande-cultural diversity, ande-literature,
15..19	ande-the media, ande-humor, ande-language, ande-evolution, ande-search engine,
20..25	ande-the media, ande-humor, ande-evolution, ande-search engine, ande-environment,
26..30	ande-social networking, ande-environment,
31..40	ande-social networking, ande-green,
41..50	
51..75	
76..100	ande-web,
101..150	
151..1000	

Test Results Categorized by Training Count

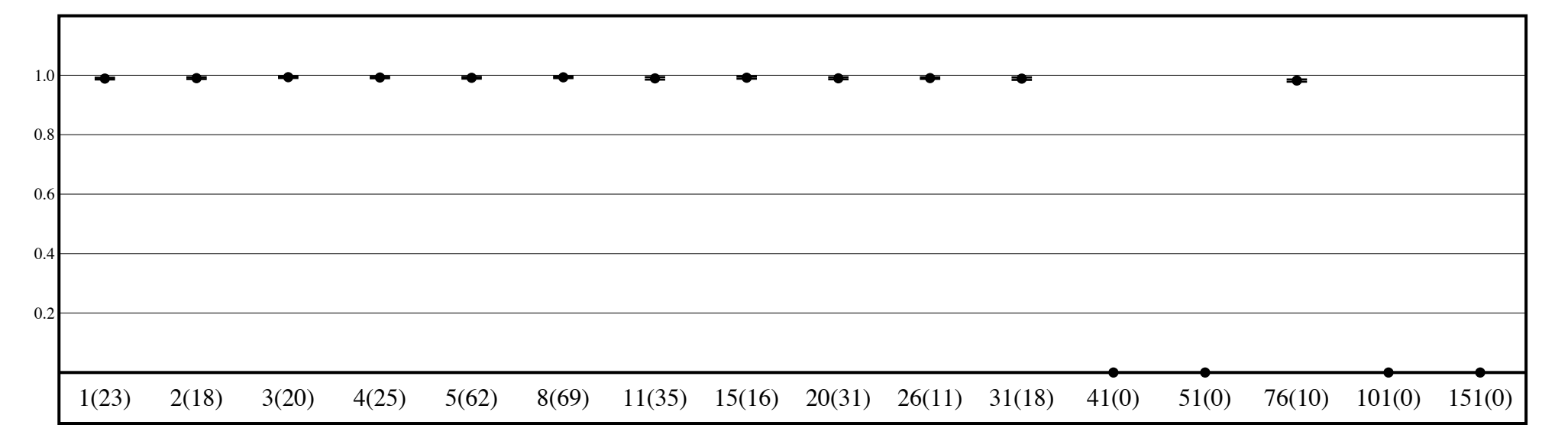
29257090: Ratio for n is determined using total number of tokens from each pool, using 'unwanted' as background.

These graphs display the cross validation test results for a classifier arranged by the number of items a classifier was trained on for a given tag. The X axis displays the number of training items, where there is a gap, the label defines the lower bound of a range that extends to the next label. The number in brackets is the number of tags that fall into that 'bin'. The Y Axis is the mean true positive or true negative rate for the classifiers that fall in the 'bin'. Standard Deviation is shown as error bars on the chart.

True Positive Results



True Negative Results



Tags in bins

Bin	Tags
1..1	ande-nonprofit, ande-spam, ande-birds, ande-astronomy, ande-cats,
2..2	ande-machine learning, ande-spam, ande-birds, ande-astronomy,
3..3	ande-machine learning, ande-spam, ande-earth, ande-zoology, ande-music,
4..4	ande-machine learning, ande-earth, ande-zoology, ande-racial diversity, ande-music,
5..7	ande-gay rights, ande-censorship, ande-racial diversity, ande-music, ande-anthropology, ande-vista, ande-medicine, ande-Macs, ande-photography,

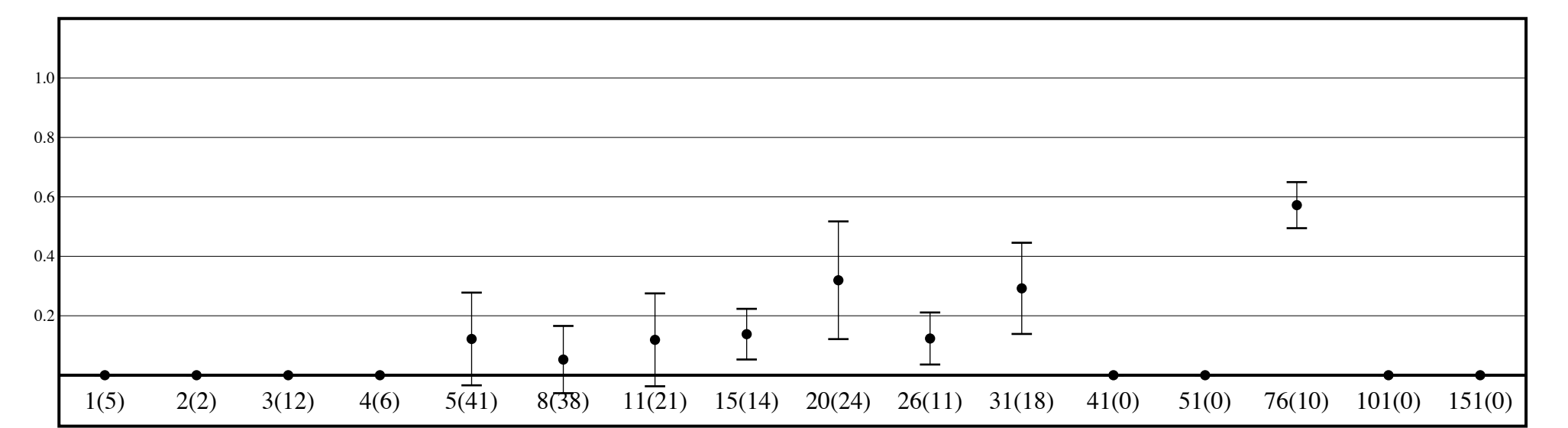
Bin	Tags
	ande-women in politics, ande-global warming,
8..10	ande-education, ande-gay rights, ande-censorship, ande-collaboration, ande-vista, ande-medicine, ande-Macs, ande-food, ande-cultural diversity, ande-women in politics, ande-privacy, ande-global warming,
11..14	ande-education, ande-language, ande-collaboration, ande-food, ande-cultural diversity, ande-literature,
15..19	ande-the media, ande-humor, ande-language, ande-evolution, ande-search engine,
20..25	ande-the media, ande-humor, ande-evolution, ande-search engine, ande-environment,
26..30	ande-social networking, ande-environment,
31..40	ande-social networking, ande-green,
41..50	
51..75	
76..100	ande-web,
101..150	
151..1000	

Test Results Categorized by Training Count

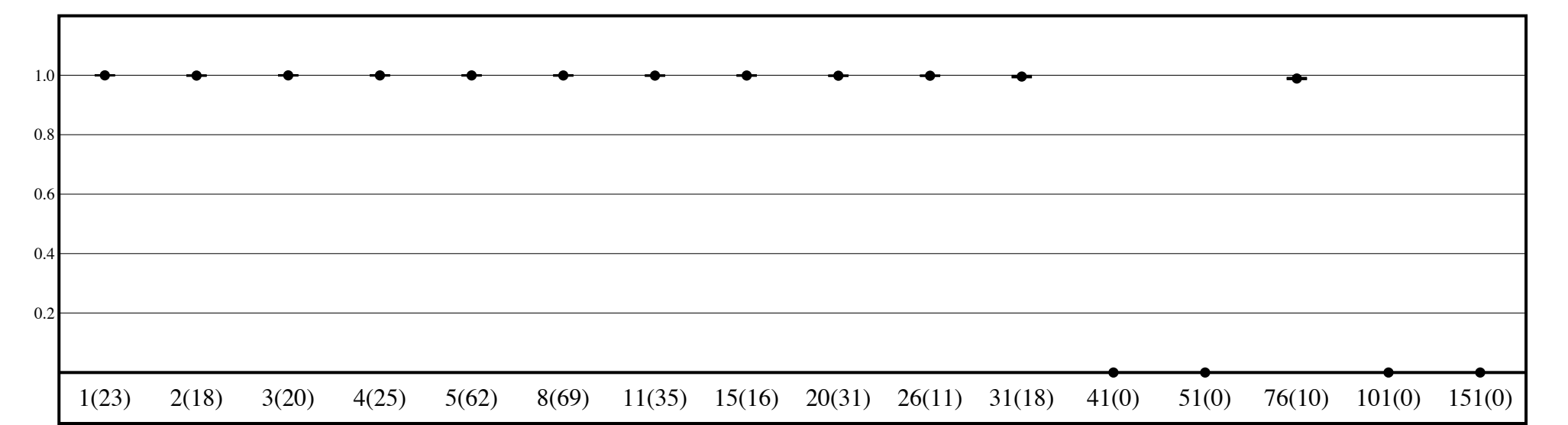
29257090: Ratio for n is determined using number of distinct tokens from each pool, using 'unwanted' as background.

These graphs display the cross validation test results for a classifier arranged by the number of items a classifier was trained on for a given tag. The X axis displays the number of training items, where there is a gap, the label defines the lower bound of a range that extends to the next label. The number in brackets is the number of tags that fall into that 'bin'. The Y Axis is the mean true positive or true negative rate for the classifiers that fall in the 'bin'. Standard Deviation is shown as error bars on the chart.

True Positive Results



True Negative Results



Tags in bins

Bin	Tags
1..1	ande-nonprofit, ande-spam, ande-birds, ande-astronomy, ande-cats,
2..2	ande-machine learning, ande-spam, ande-birds, ande-astronomy,
3..3	ande-machine learning, ande-spam, ande-earth, ande-zoology, ande-music,
4..4	ande-machine learning, ande-earth, ande-zoology, ande-racial diversity, ande-music,
5..7	ande-gay rights, ande-censorship, ande-racial diversity, ande-music, ande-anthropology, ande-vista, ande-medicine, ande-Macs, ande-photography,

Bin	Tags
	ande-women in politics, ande-global warming,
8..10	ande-education, ande-gay rights, ande-censorship, ande-collaboration, ande-vista, ande-medicine, ande-Macs, ande-food, ande-cultural diversity, ande-women in politics, ande-privacy, ande-global warming,
11..14	ande-education, ande-language, ande-collaboration, ande-food, ande-cultural diversity, ande-literature,
15..19	ande-the media, ande-humor, ande-language, ande-evolution, ande-search engine,
20..25	ande-the media, ande-humor, ande-evolution, ande-search engine, ande-environment,
26..30	ande-social networking, ande-environment,
31..40	ande-social networking, ande-green,
41..50	
51..75	
76..100	ande-web,
101..150	
151..1000	