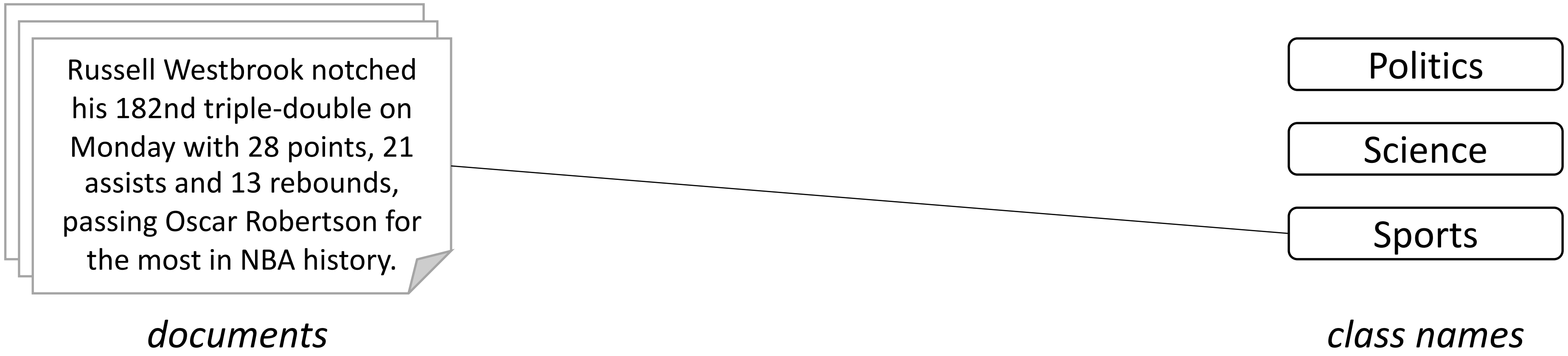
X-Class: Text Classification with Extremely Weak Supervision

Zihan Wang, Dheeraj Mekala, Jingbo Shang

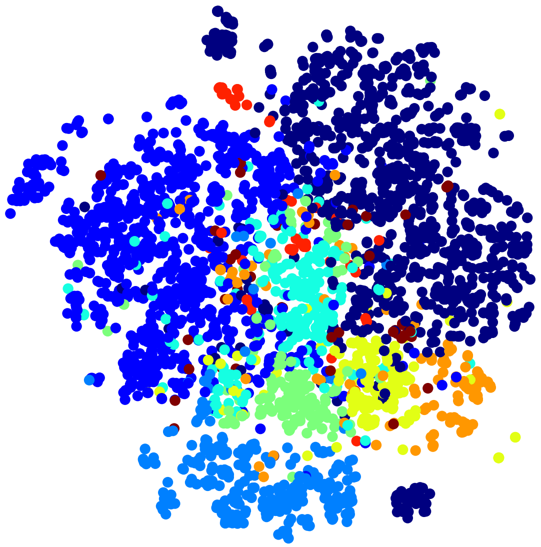
University of California San Diego

Extremely Weak Supervision



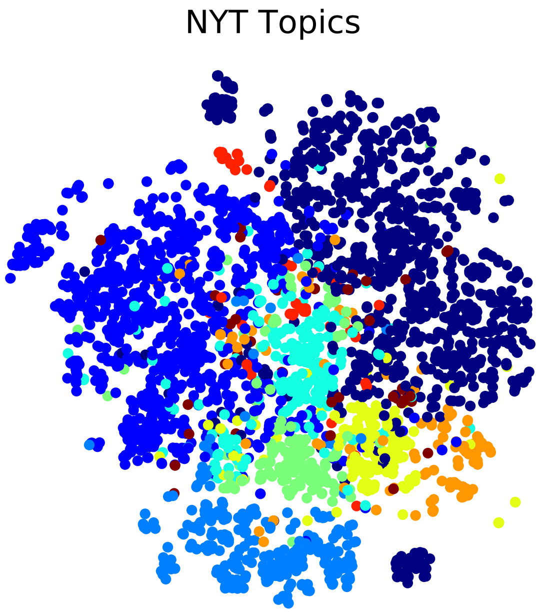
- Align documents with classes
- The only input are the documents and class names --- no alignment between them are given.

Averaging representations “works”



- Embed documents with pre-trained language models and average their last layer representations.
- Colors represent different classes.
- Colors are fairly well separated.

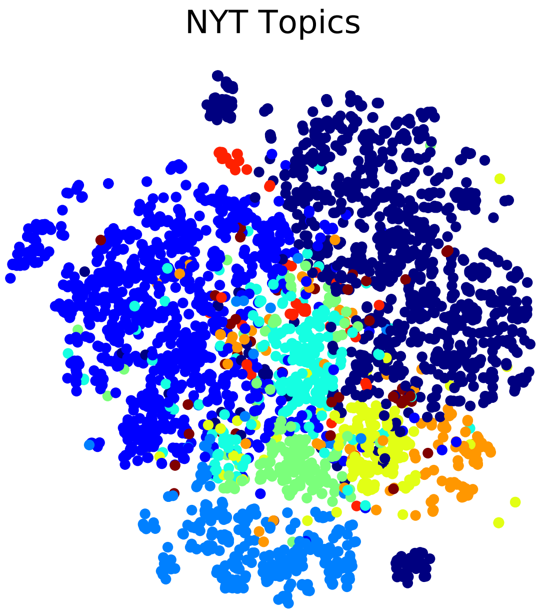
Averaging representations "works"

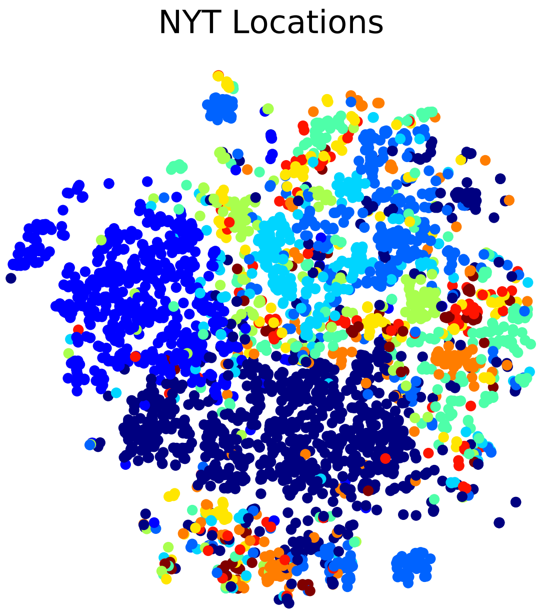


- A document can be related to different aspects, e.g. Topics (News, Sports), Location (U.S., France), Sentiment (Happy, Sad).
- The representations are fixed while the class distribution can change.

- Embed documents with pre-trained language models and average their last layer representations.
- Colors represent different classes.
- Colors are fairly well separated.

Averaging representations "works"



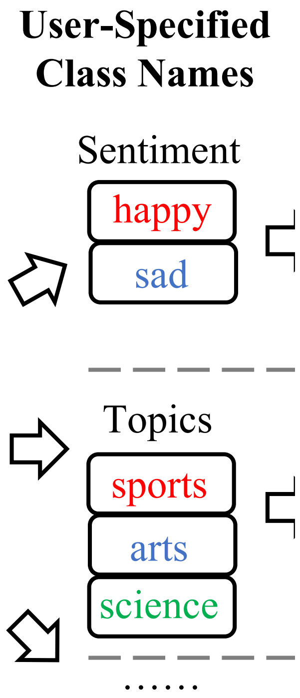


- Colors are mixed up for the same corpus, under locations class type.
- Directly averaging representations without considering the class name fail for different classes

X-Class

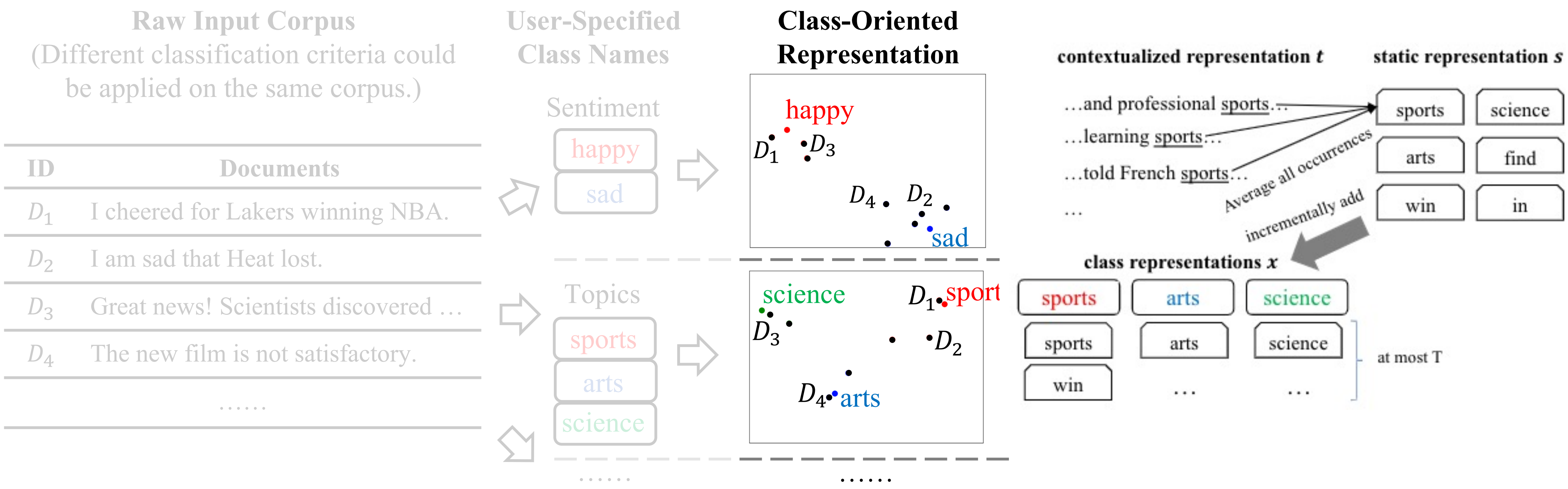
Raw Input Corpus
(Different classification criteria could be applied on the same corpus.)

ID	Documents
D_1	I cheered for Lakers winning NBA.
D_2	I am sad that Heat lost.
D_3	Great news! Scientists discovered …
D_4	The new film is not satisfactory.
	……



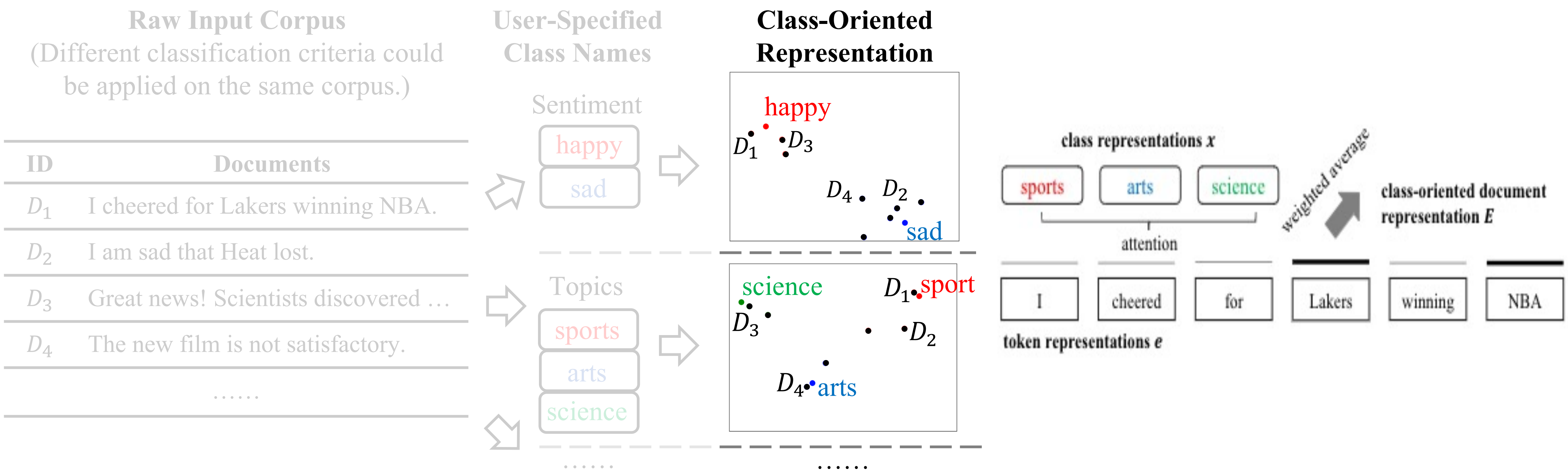
Given a corpus and a set of class names, X-Class aligns them through a representation focused approach

X-Class



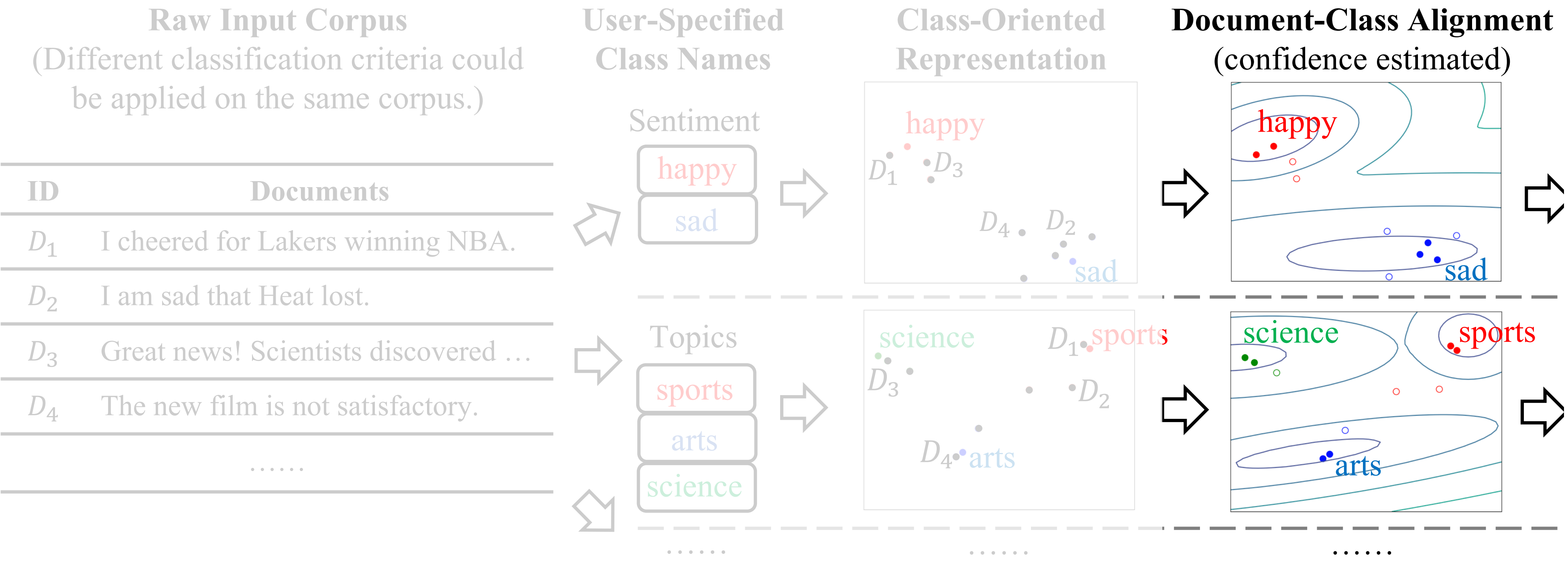
X-Class first finds class representations through iteratively expanding the class semantics, then attend on documents with the class representations to obtain class-oriented document representations.

X-Class



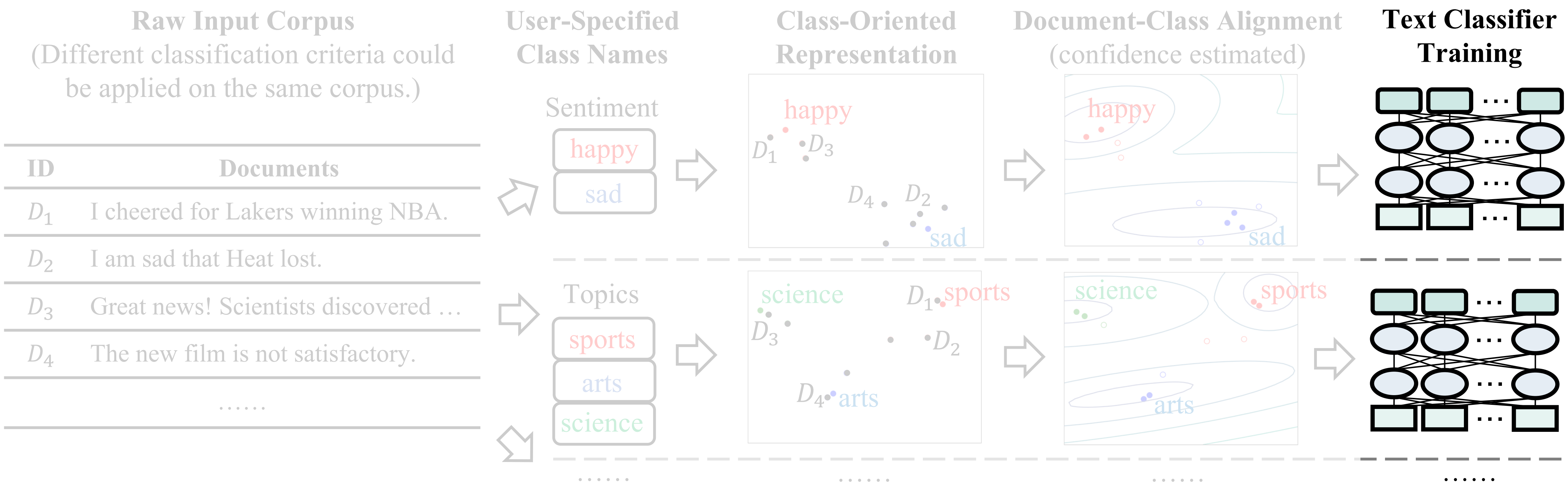
X-Class first finds class representations through iteratively expanding the class semantics, **then attend on documents with the class representations to obtain class-oriented document representations.**

X-Class



Use gaussian mixture model to cluster the document representations.

X-Class



Train a supervised classifier on the confident pseudo document-labels pairs obtained from clustering

Experiments

	AGNews	20News	NYT-Small	NYT-Topic	NYT-Location	Yelp	DBpedia
Corpus Domain	News	News	News	News	News	Reviews	Wikipedia
Class Criterion	Topics	Topics	Topics	Topics	Locations	Sentiment	Ontology
# of Classes	4	5	5	9	10	2	14
# of Documents	120,000	17,871	13,081	31,997	31,997	38,000	560,000
Imbalance	1.0	2.02	16.65	27.09	15.84	1.0	1.0

- Seven datasets covering different domains.
- Imbalance: size of the largest class / size of the smallest class

Experiments

	Model	AGNews	20News	NYT-Small	NYT-Topic	NYT-Location	Yelp	DBpedia
	Supervised	93.99/93.99	96.45/96.42	97.95/95.46	94.29/89.90	95.99/94.99	95.7/95.7	98.96/98.96
Expert-find seed words	WeSTClass†	82.3/82.1§	71.28/69.90	91.2/83.7§	68.26/57.02	63.15/53.22	81.6/81.6§	81.42/81.19
	ConWea†	74.6/74.2	75.73/73.26	95.23/90.79	**81.67/71.54**	85.31/83.81	71.4/71.2	N/A
Only class names	LOTClass‡	**86.89/86.82**	73.78/72.53	78.12/56.05	67.11/43.58	58.49/58.96	87.75/87.68	86.66/85.98
	X-Class ‡	85.74/85.66	**78.62/77.76**	**97.18/94.02**	79.02/68.55	**91.8/91.98**	**90.0/90.0**	**91.32/91.17**

Example expert-find seed words

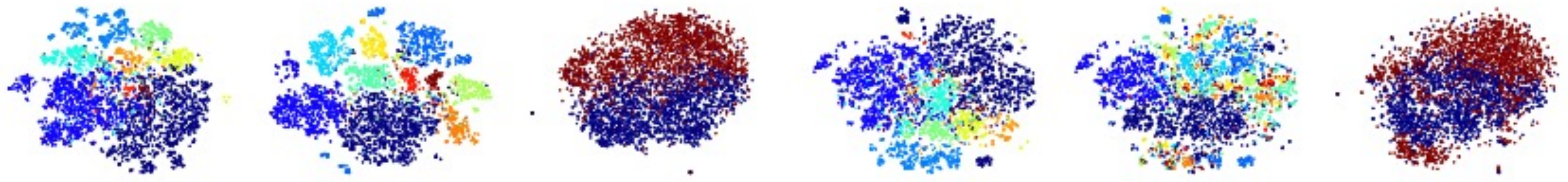
class	Seed words	Class name for X-Class
arts	dance,art, ballet,museum	arts
business	shares,stocks, markets,trading	business

Experiments

	Model	AGNews	20News	NYT-Small	NYT-Topic	NYT-Location	Yelp	DBpedia
	Supervised	93.99/93.99	96.45/96.42	97.95/95.46	94.29/89.90	95.99/94.99	95.7/95.7	98.96/98.96
Expert-find seed words	WeSTClass†	82.3/82.1§	71.28/69.90	91.2/83.7§	68.26/57.02	63.15/53.22	81.6/81.6§	81.42/81.19
	ConWea†	74.6/74.2	75.73/73.26	95.23/90.79	**81.67/71.54**	85.31/83.81	71.4/71.2	N/A
Only class names	LOTClass‡	**86.89/86.82**	73.78/72.53	78.12/56.05	67.11/43.58	58.49/58.96	87.75/87.68	86.66/85.98
	X-Class ‡	85.74/85.66	**78.62/77.76**	**97.18/94.02**	79.02/68.55	**91.8/91.98**	**90.0/90.0**	**91.32/91.17**
	Ablations							
	X-Class-Rep‡	77.86/76.84	75.37/73.7	92.13/83.69	77.06/65.05	86.36/88.1	78.0/77.19	74.05/71.74
	X-Class-Align‡	83.32/83.28	79.19/78.46	96.42/92.32	79.12/67.76	90.09/90.63	87.19/87.13	87.36/87.27

- X-Class-Rep: After obtaining class-oriented document representations, use the nearest class representation as its class
- X-Class-Align: After clustering, use the clusters as the assignment

Experiments

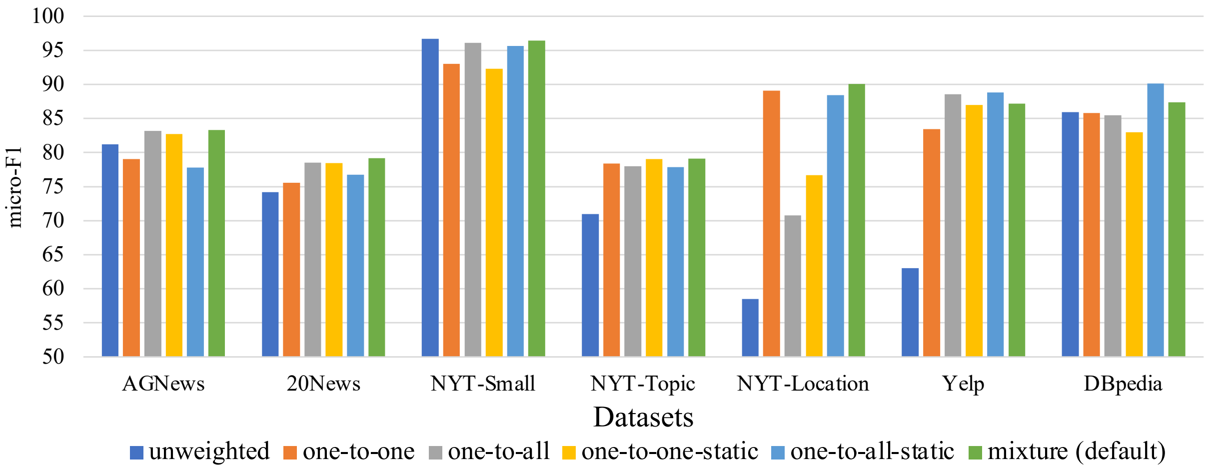


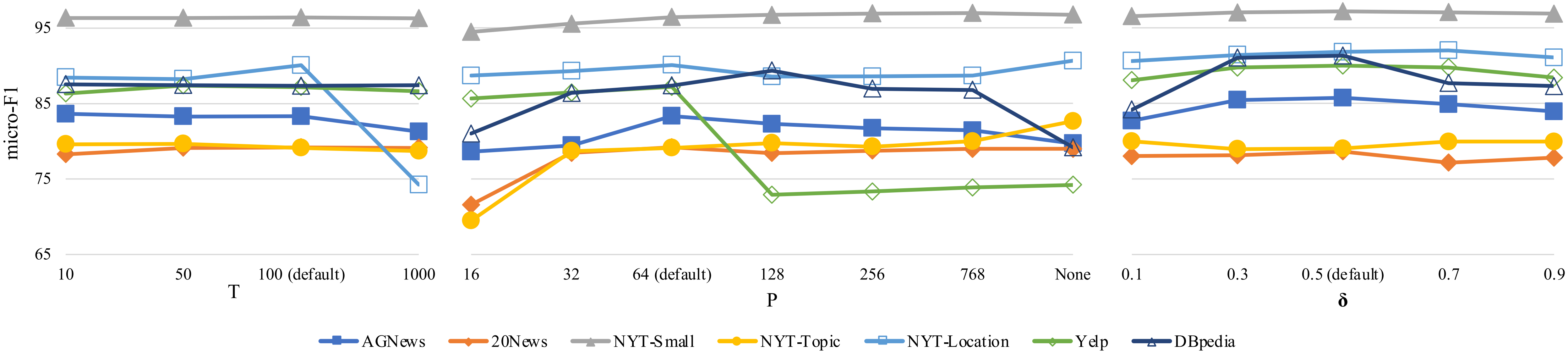
(a) Our Class-Oriented Document Representations (b) Simple Average of BERT Representations

- t-SNE plots on NYT-Topics, NYT-Locations, and Yelp (Sentiment)
- The classes are better separated using our class oriented representations than a simple average

Experiments

- Our method is not hyper parameter sensitive





(a) T in Class Rep. Estimation (b) P in Document-Class Alignment (c) δ in Text Classifier Training

Good properties

Natural extension to coarse->fine grained hierarchical text classification (NYT-Small).

Model	Coarse (5 classes)	Fine (26 classes)
WeSTClass	91/84[§]	50/36[§]
WeSHClass		87.4/63.2[§]
ConWea	95.23/90.79	91/79[§]
X-Class-End	96.67/92.98	86.07/75.30
X-Class-Hier		92.66/80.92

Weak requirement for class names to exist in corpus.

Model	20News		NYT-Small	
	Original	Removed	Original	Removed
X-Class	77.76	74.48	94.02	93.29
LOTClass	72.53	8.82	56.05	29.53

Performance only decrease a bit when removing all but one occurrence of a class name in the dataset.

Conclusions & Future work

- Extending the idea of extremely weak supervision to other tasks, e.g., Named entity recognition.
- A possible *unsupervised text classification* where machines automatically suggest different sets of class names, and then assign documents to each class.

Code & data: https://github.com/ZihanWangKi/XClass