MS, Computer Science - Thesis Defense

Sneha Nagpaul
snagpaul@gmu.edu

16th January 2018

From Language to Location using Multiple Instance Neural Networks

Sneha Nagpaul
snagpaul@gmu.edu

16th January 2018

Outline

- Problem
- Datasets
- Background - MIL, GIR, and NN
- Proposed Method
    - 01: Model Architecture
    - 02: Instance Level Model
    - 03: Aggregation Scheme
    - 04: Training and Loss
- Results
- Conclusion & Future Directions

Problem

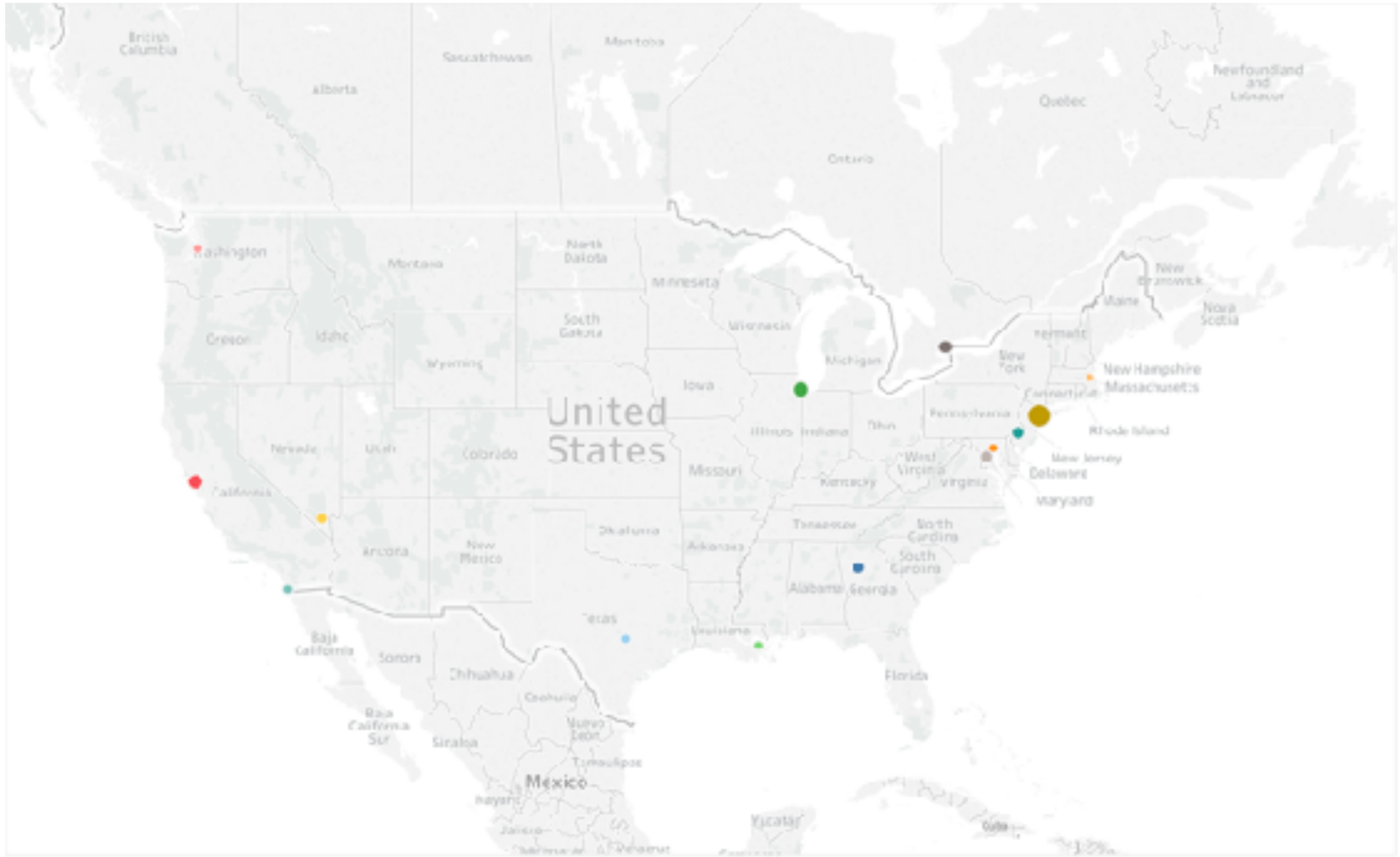
Predict the location of a tweet when its generator's location is given.

Why?

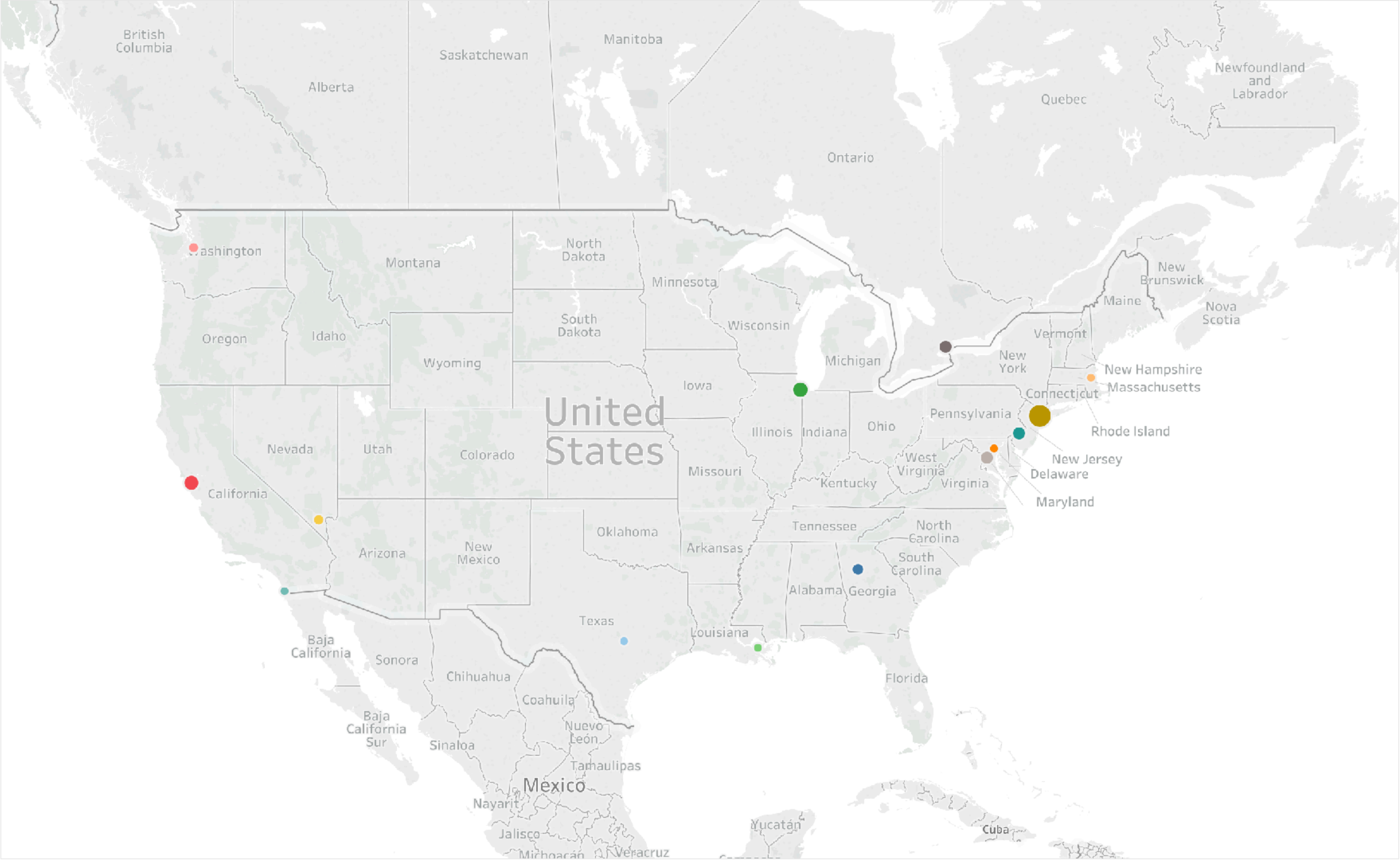
American Red Cross

METROPOLITAN POLICE
WASHINGTON D.C.

Dataset




British Columbia
Alberta
Saskatchewan
Manitoba
Ontario
Quebec
Newfoundland and Labrador
New Brunswick
Nova Scotia
Washington
Oregon
California
Nevada
Idaho
Montana
Wyoming
Utah
Colorado
Arizona
New Mexico
North Dakota
South Dakota
United States
Oklahoma
Texas
Minnesota
Iowa
Missouri
Arkansas
Louisiana
Wisconsin
Illinois
Michigan
Indiana
Ohio
Kentucky
Tennessee
Alabama
Georgia
Florida
South Carolina
North Carolina
Virginia
West Virginia
Pennsylvania
New York
Vermont
Maine
New Hampshire
Massachusetts
Connecticut
Rhode Island
New Jersey
Delaware
Maryland
Baja California
Sonora
Chihuahua
Coahuila
Nuevo León
Baja California Sur
Sinaloa
Tamaulipas
Mexico
Nayarit
Jalisco
Michoacán
Veracruz
Yucatán
Cuba


British Columbia
Alberta
Saskatchewan
Manitoba
Quebec
Newfoundland and Labrador
Ontario
Washington
Montana
North Dakota
Minnesota
New Brunswick
Nova Scotia
Maine
South Dakota
Wisconsin
Vermont
Oregon
Idaho
Wyoming
Michigan
New York
New Hampshire
Massachusetts
Iowa
Connecticut
United States
Pennsylvania
Illinois
Indiana
Ohio
Rhode Island
Nevada
Utah
Colorado
New Jersey
West Virginia
Delaware
Missouri
Kentucky
Virginia
California
Maryland
Tennessee
North Carolina
Oklahoma
Arizona
New Mexico
Arkansas
South Carolina
Alabama
Georgia
Texas
Louisiana
Baja California
Sonora
Chihuahua
Florida
Coahuila
Baja California Sur
Nuevo León
Sinaloa
Tamaulipas
Mexico
Nayarit
Yucatán
Cuba
Jalisco
Michoacán
Veracruz

Datasets

City	Train	Test	Total
Atlanta	3,531	883	4,414
Austin	2,332	583	2,915
Baltimore	2,159	541	2,700
Boston	1,911	478	2,389
Chicago	6,628	1,658	8,286
New Orleans	2,072	520	2,592
New York City	15,200	3,800	19,000
Paradise	2,475	620	3,095
Philadelphia	4,633	1,159	5,792
San Diego	1,960	492	2,452
San Francisco	6,168	1,542	7,710
Seattle	2,680	670	3,350
Toronto	4,029	1,008	5,037
Washington,D.C.	4,584	1,148	5,732
Weehawken	1,756	440	2,196

Background

01

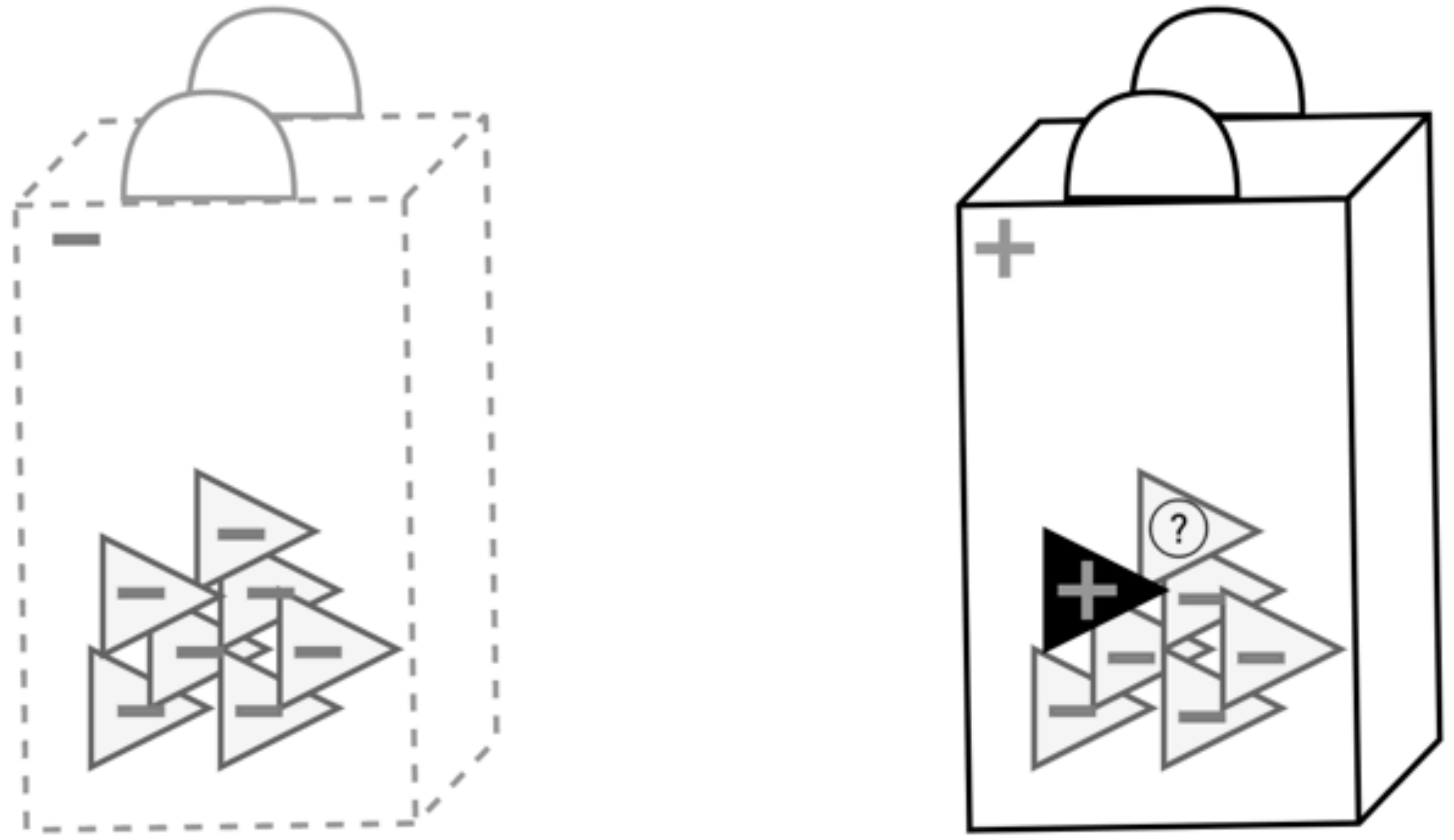
Background: Multiple Instance Learning

01

Background: Multiple Instance Learning

Background: Multiple Instance Learning

01

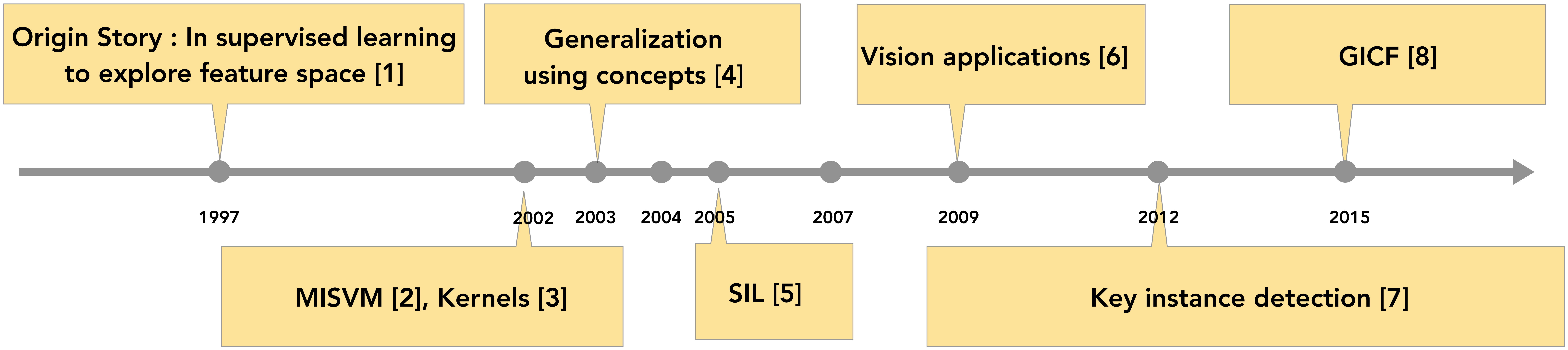


The main components:

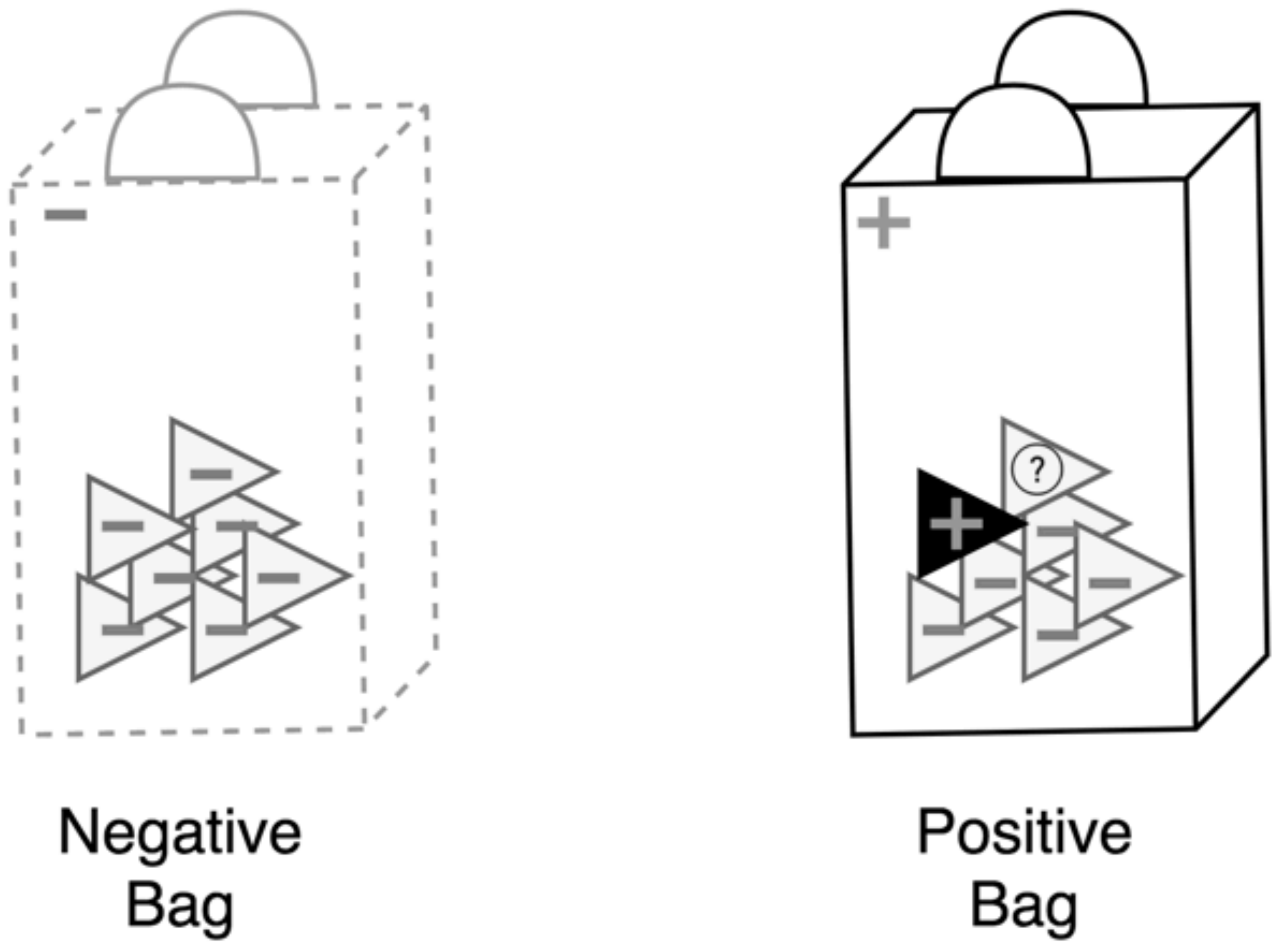
1. Membership Assumption
2. Aggregation from instance to bag level label

Background: Multiple Instance Learning

01

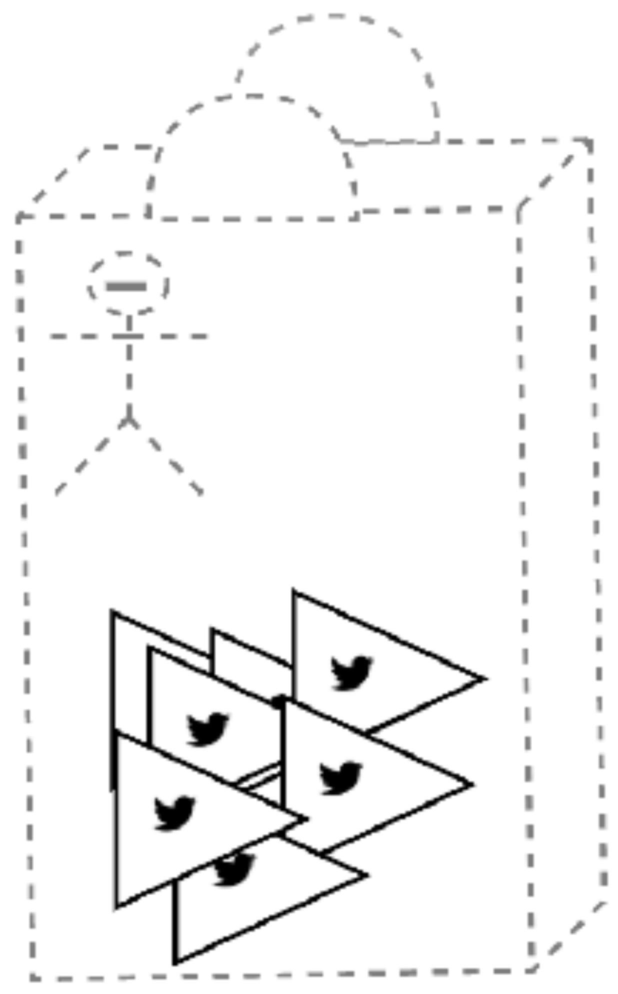


Background: Multiple Instance Learning

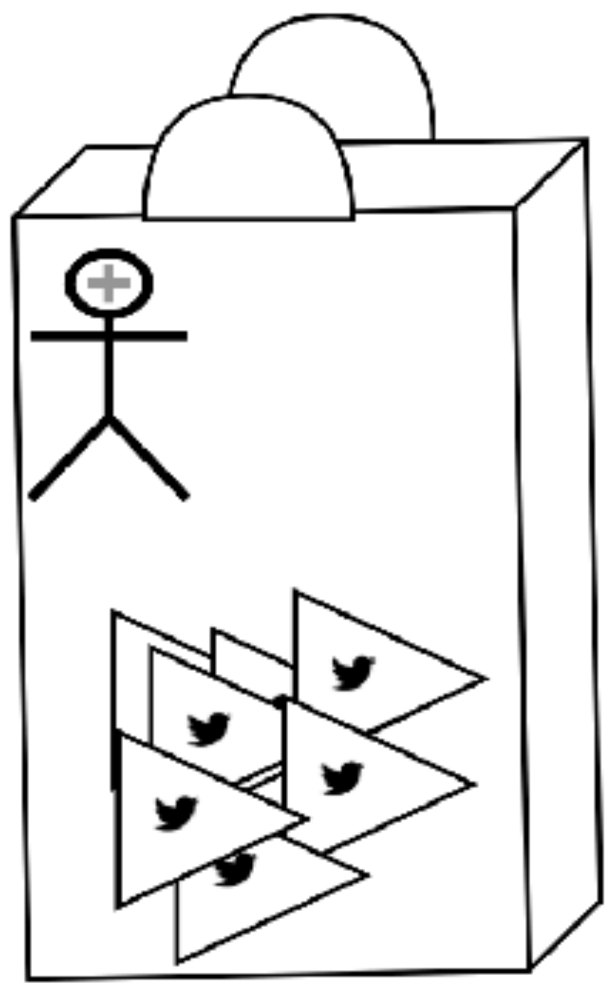


01

Background: Multiple Instance Learning



User not from city X

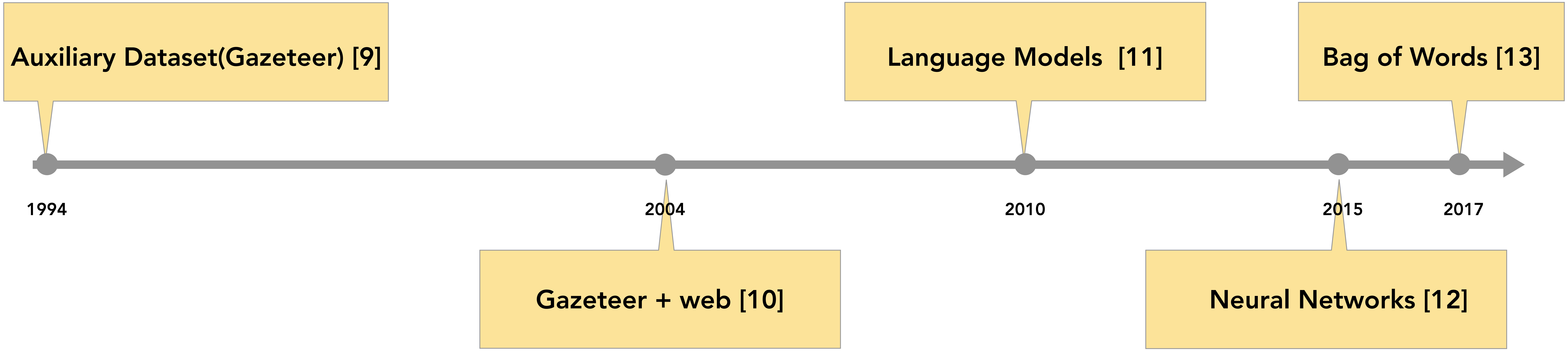


User from city X

02

Background: Geographic Information Retrieval

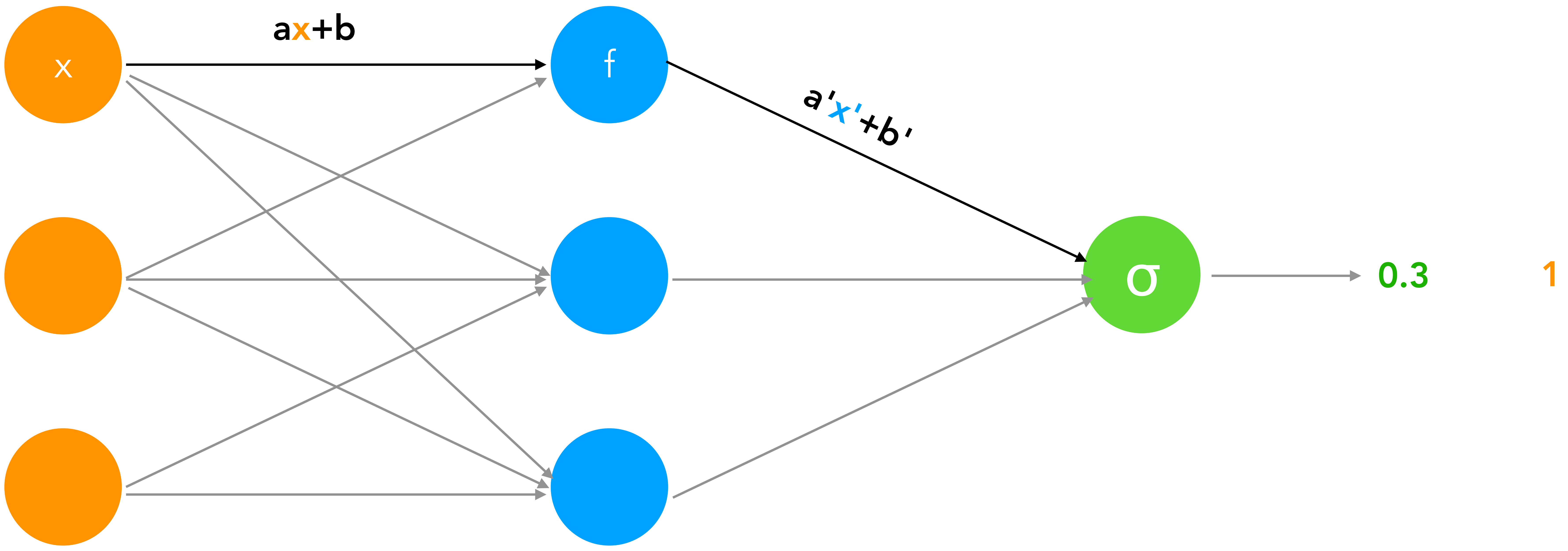
Background: Geographic Information Retrieval



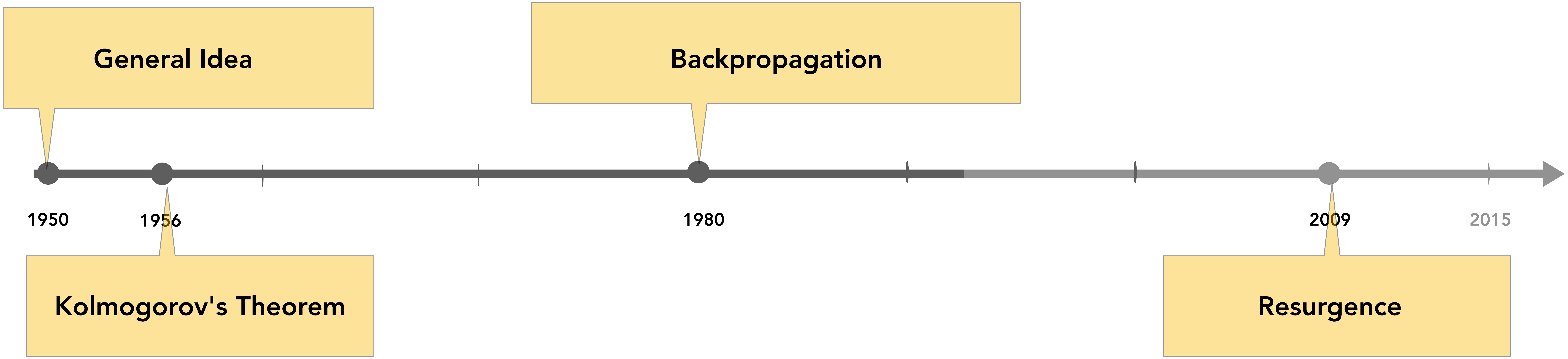
03

Background: Neural Networks

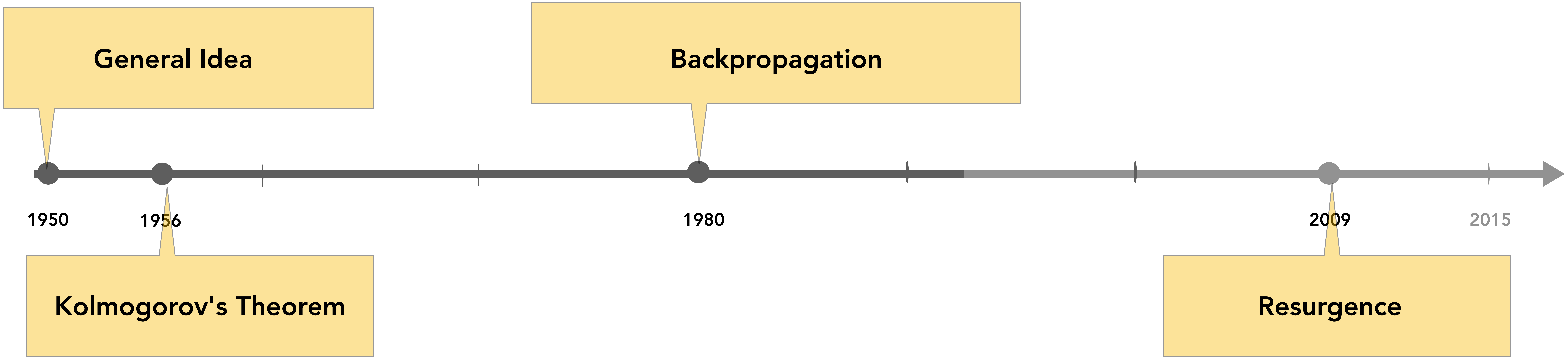
Background: Neural Networks



Background: Neural Networks



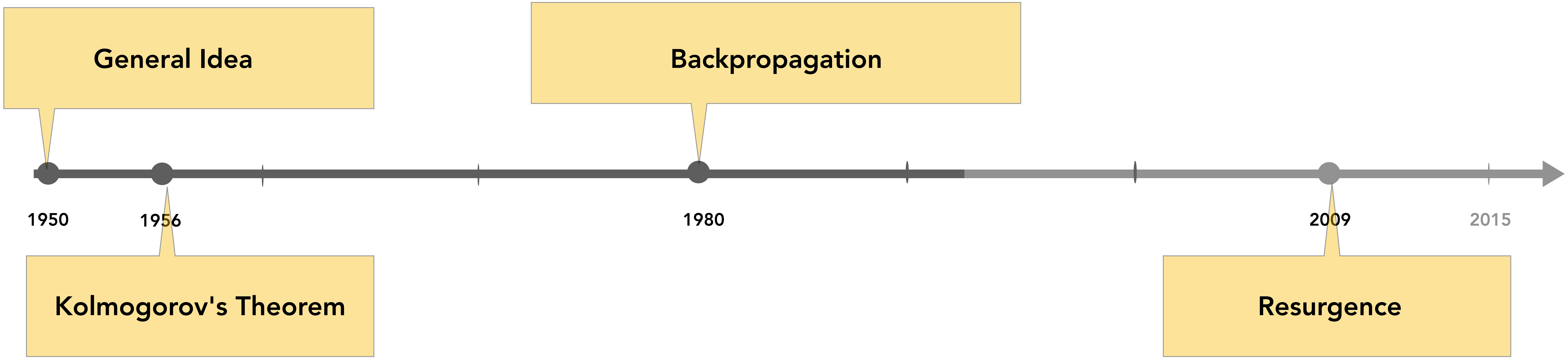
Background: Neural Networks



Affine Transformation : ax+b

Background: Neural Networks

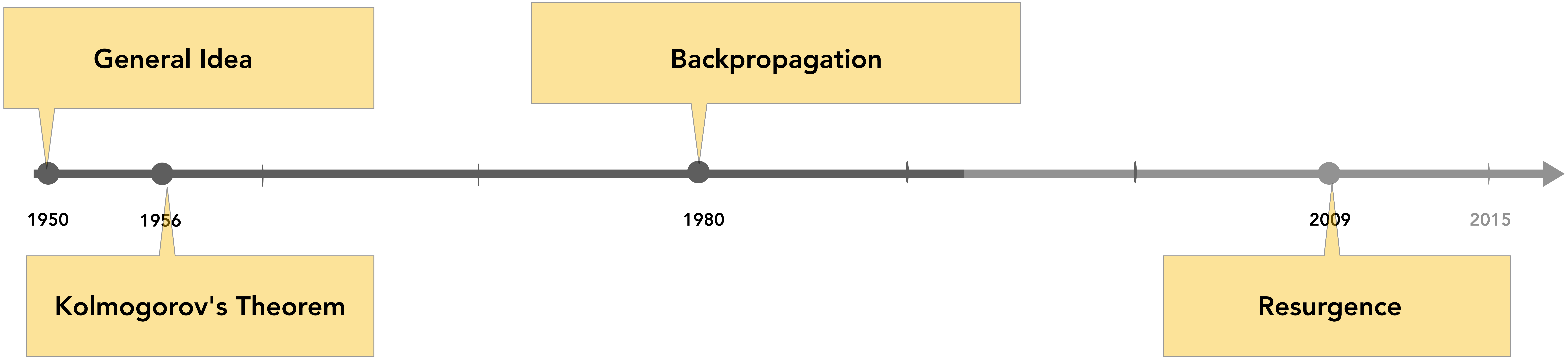
03



Affine Transformation : ax+b
Non linear activation: f(ax+b)

03

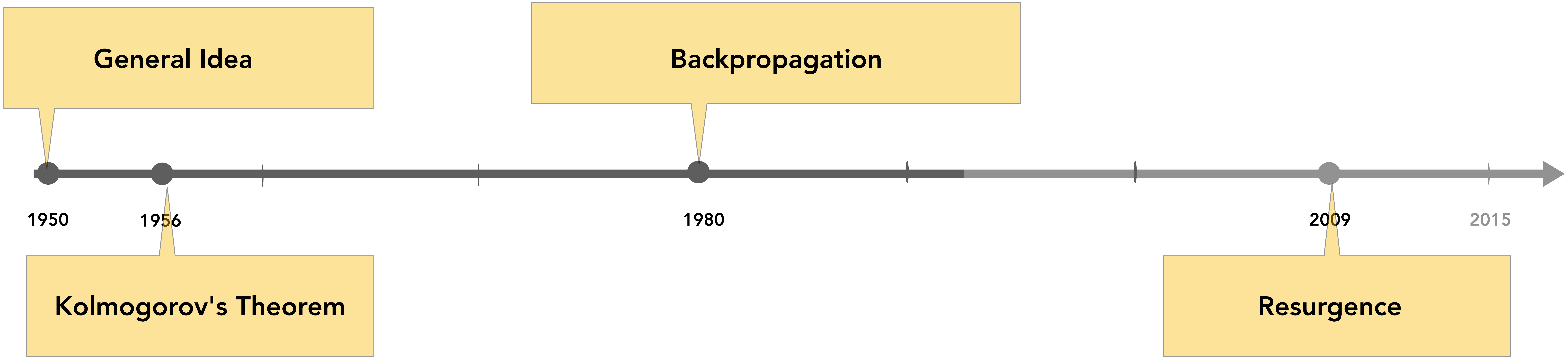
Background: Neural Networks



Affine Transformation : ax+b
Non linear activation: f(ax+b)
Composition

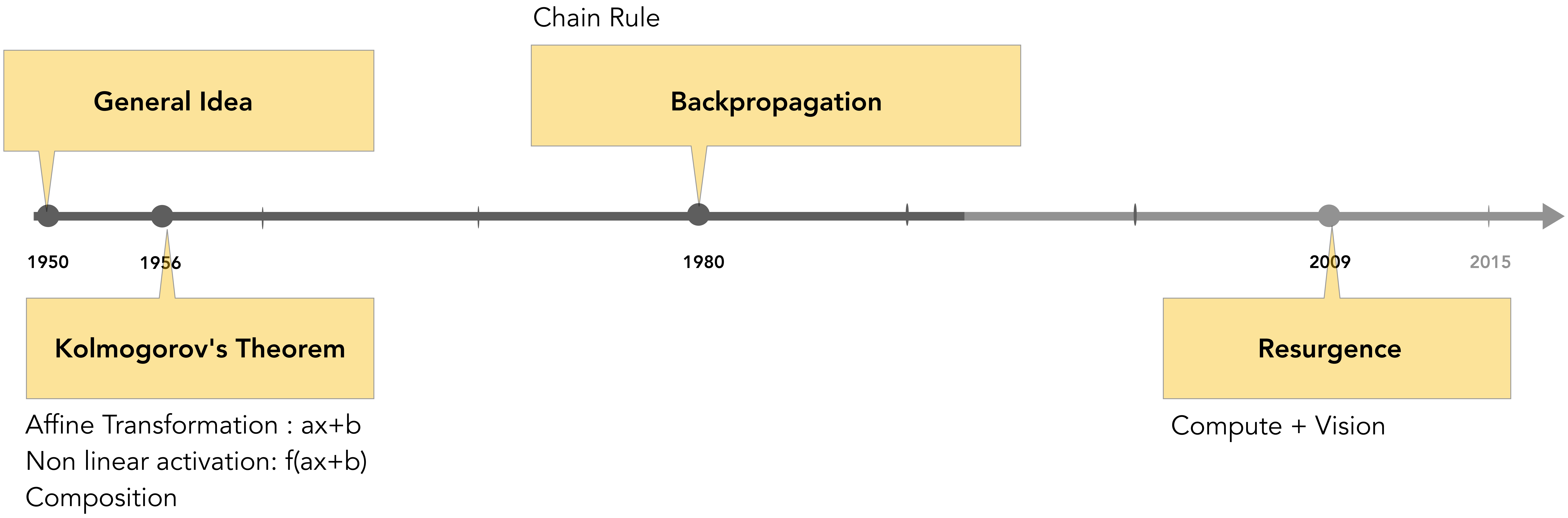
Background: Neural Networks

Chain Rule

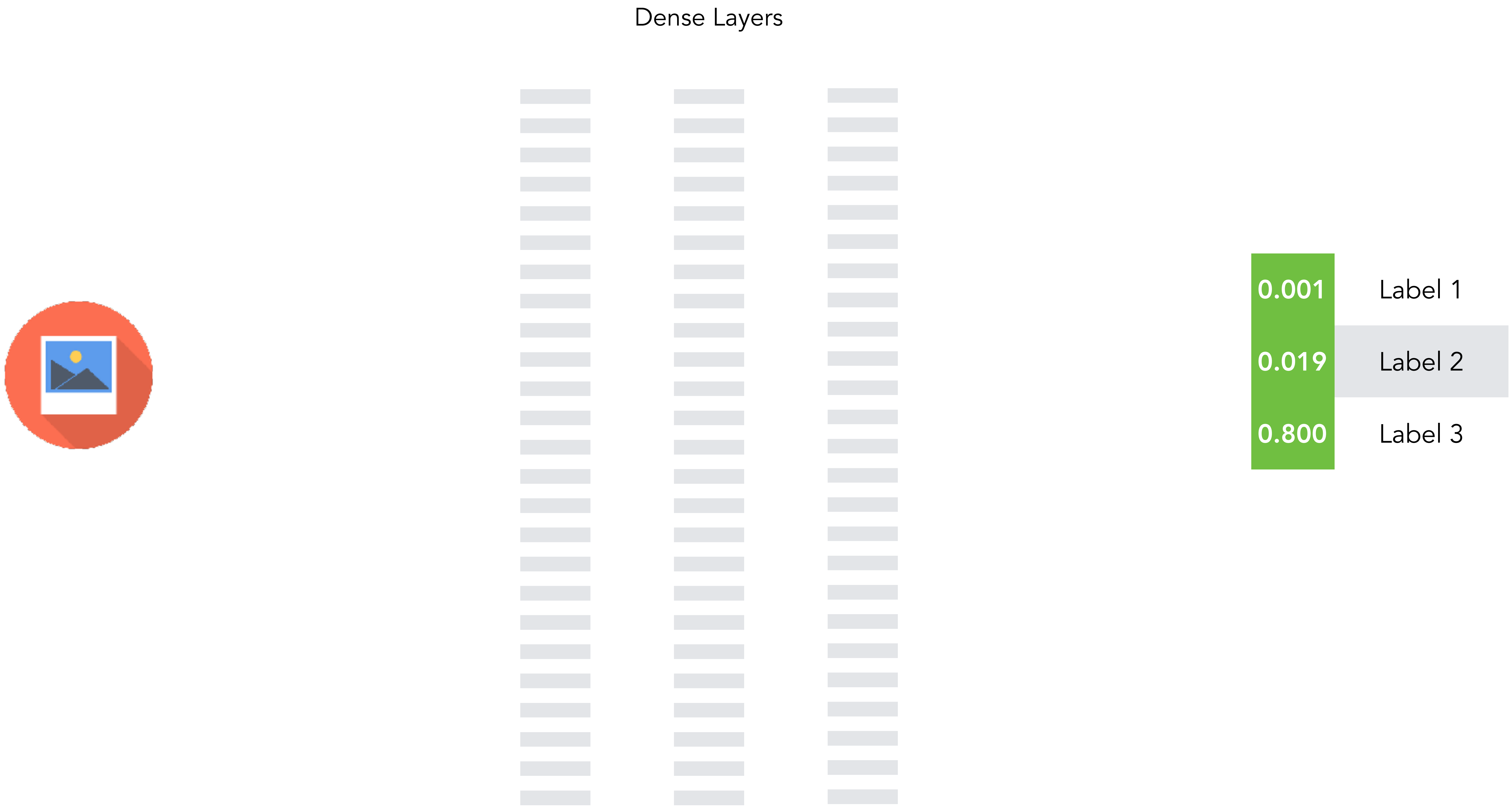


Affine Transformation : ax+b
Non linear activation: f(ax+b)
Composition

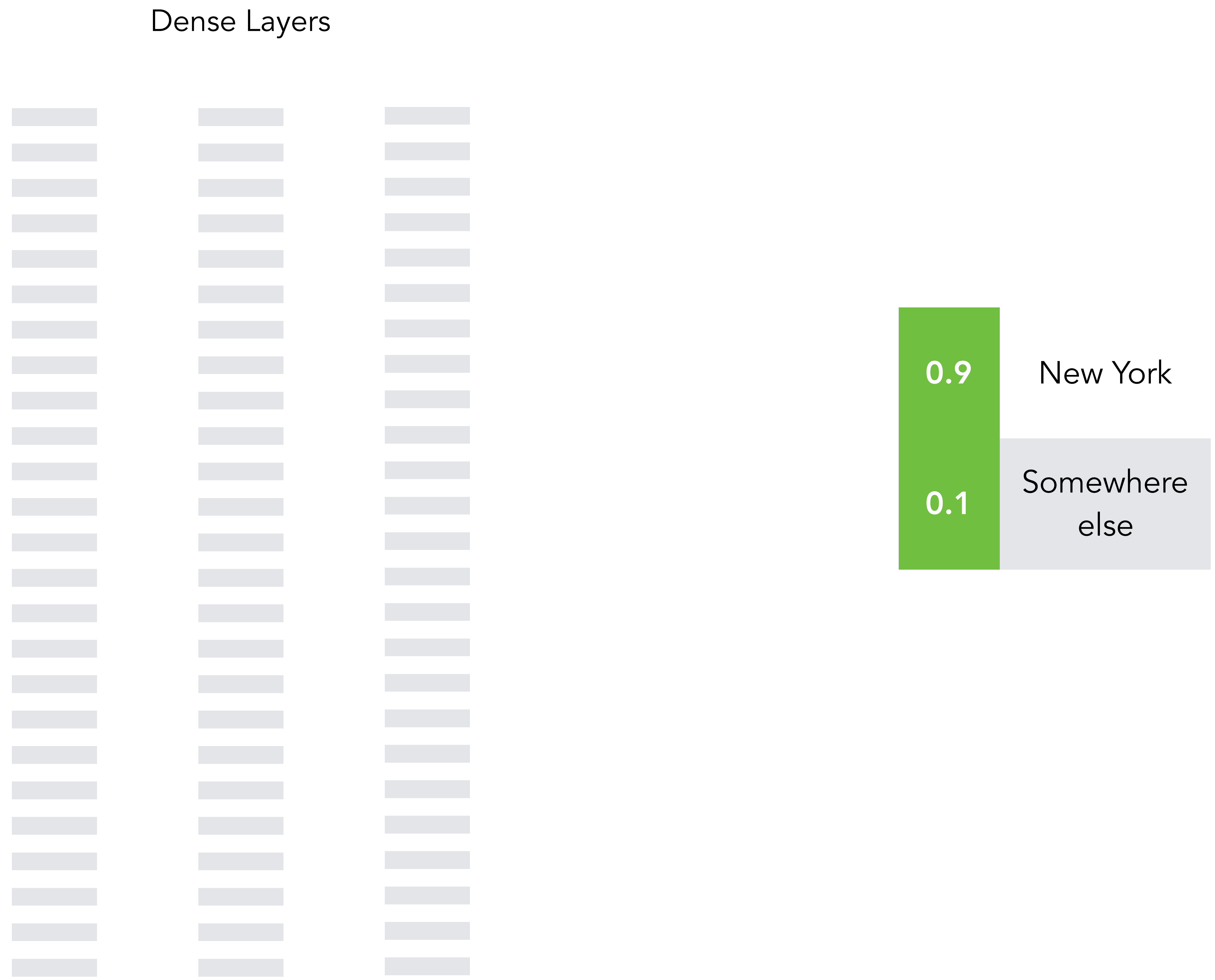
Background: Neural Networks



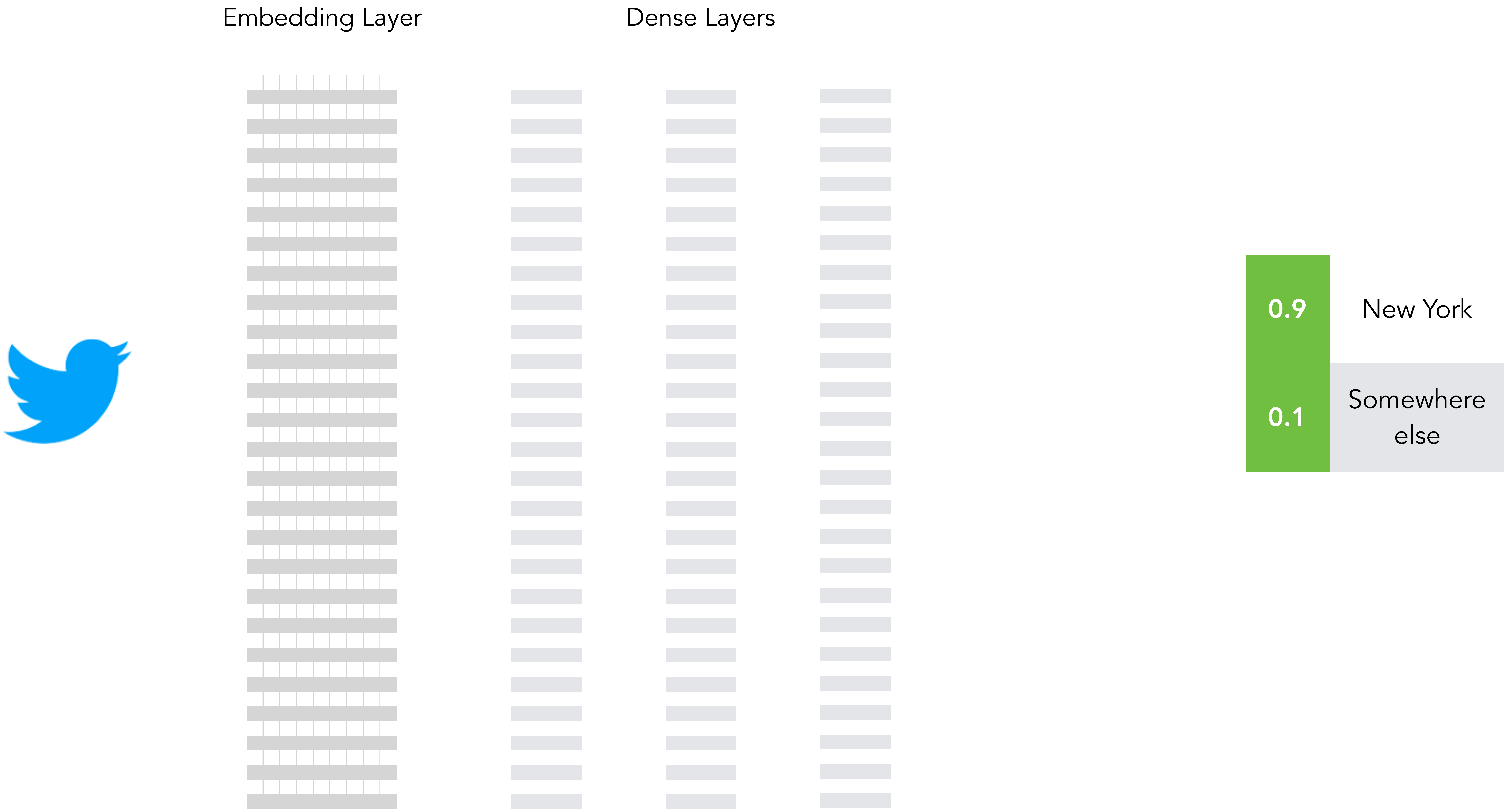
Background: Neural Networks



Background: Neural Networks



Background: Neural Networks



Background : Gaps

MIL

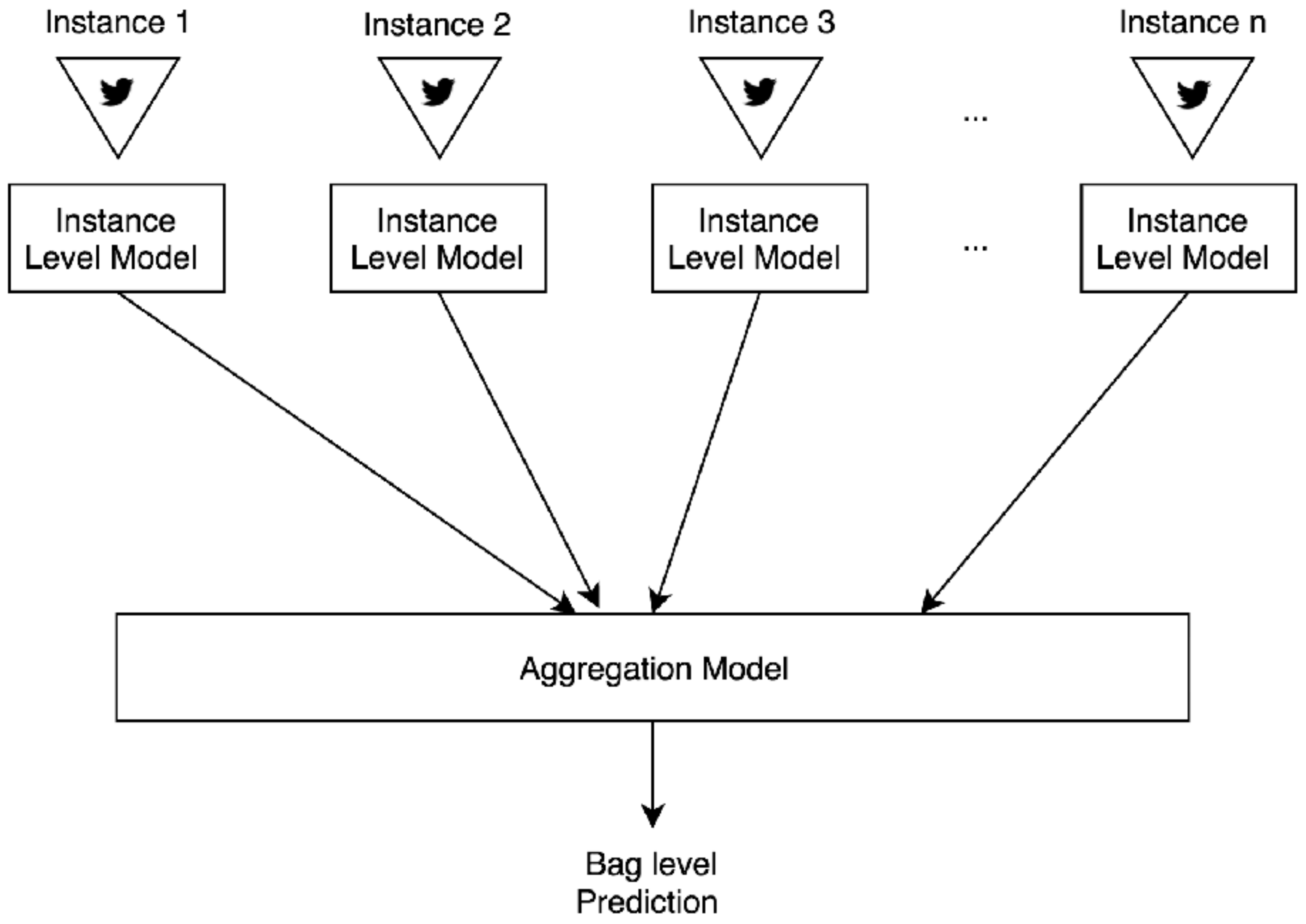
- Kernels
- Running Time - Combinatorial Complexity
- Feature Engineering
- Rigid Assumptions

GIR

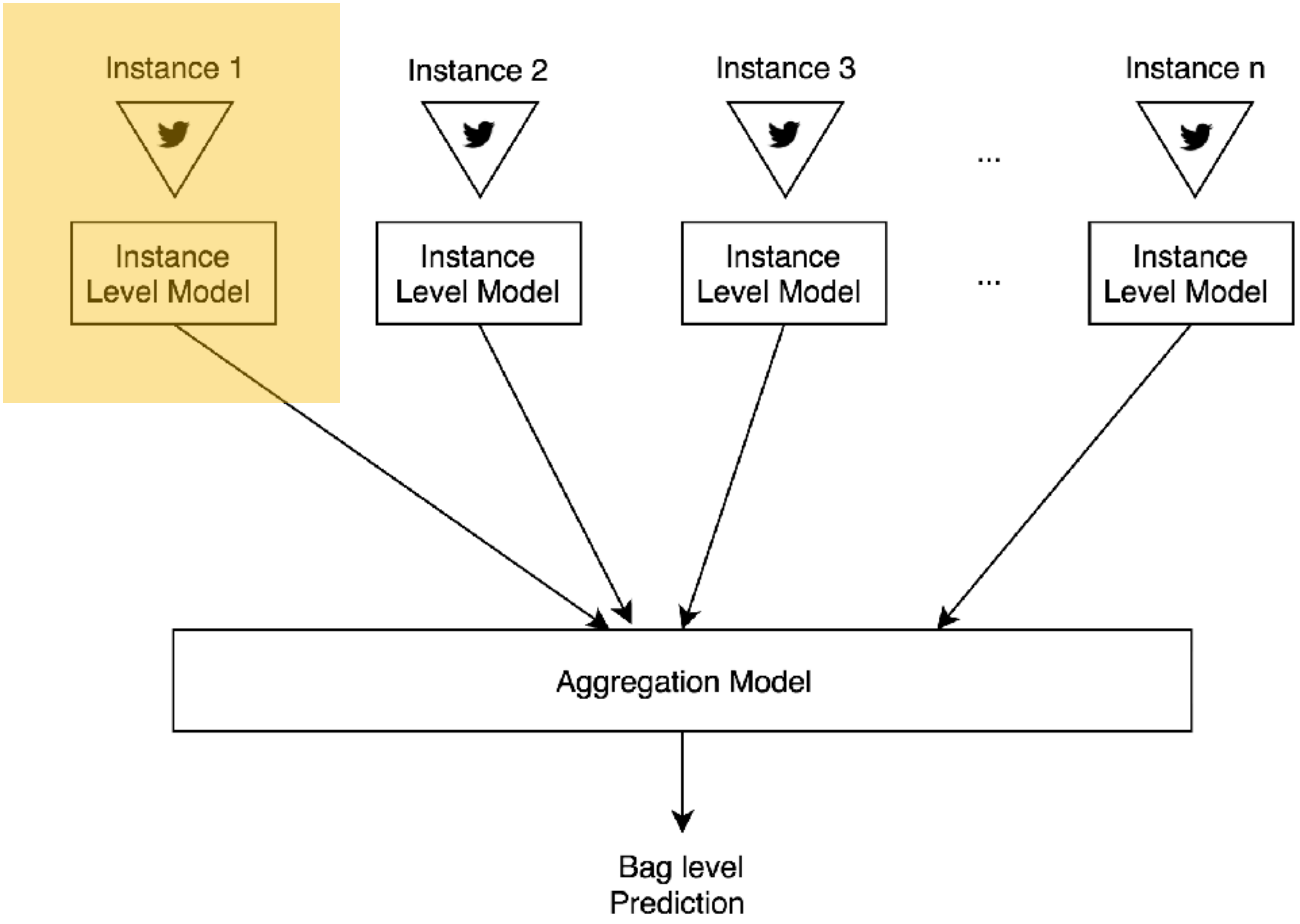
- Bag of Words Model

Proposed Method

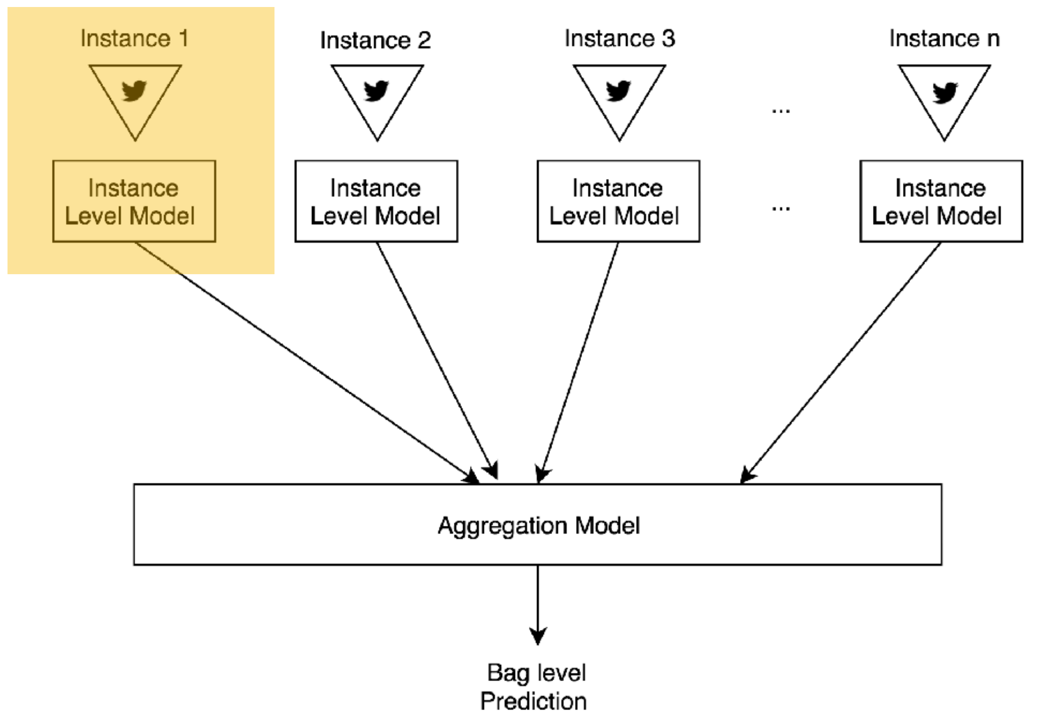
Model Architecture

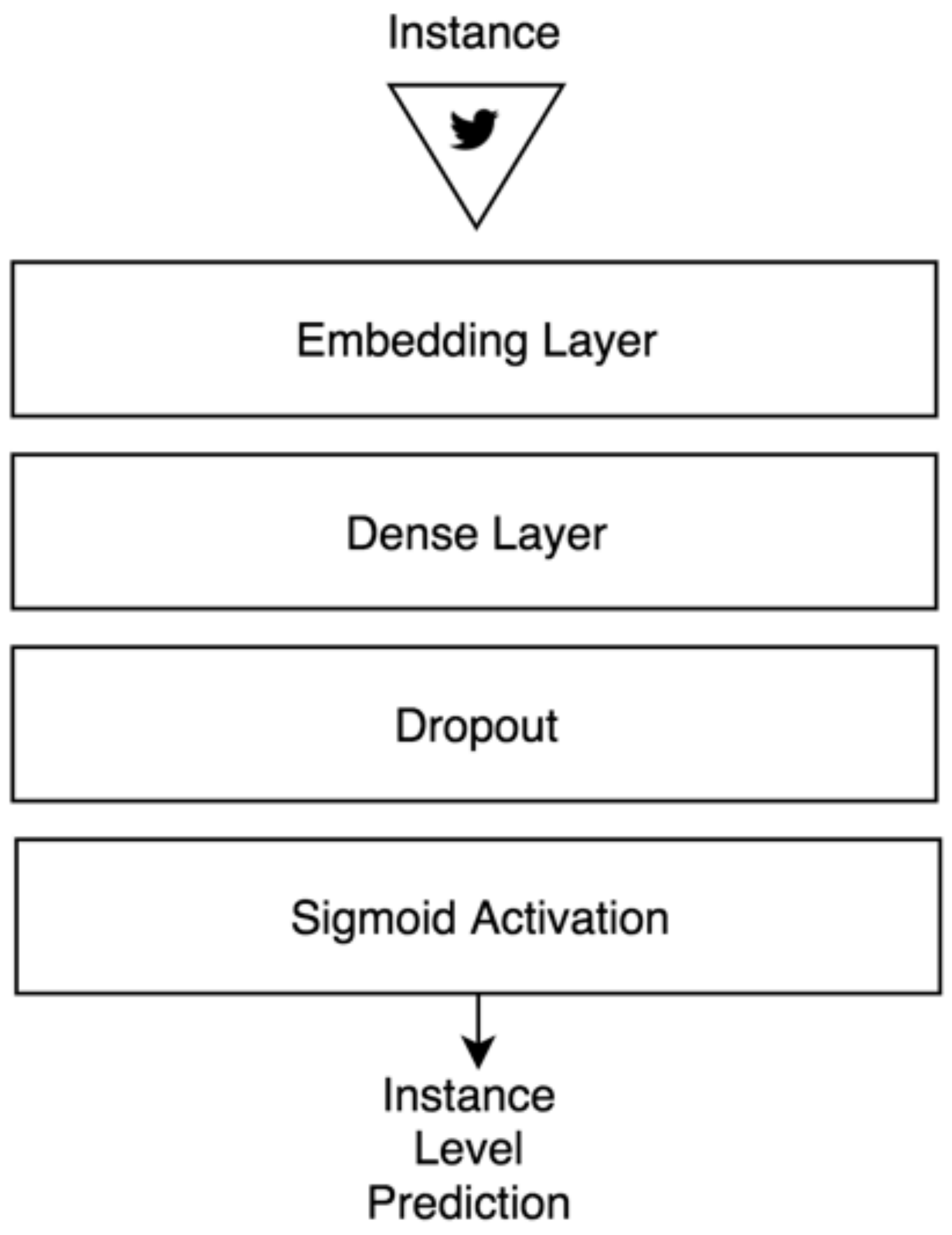


Model Architecture

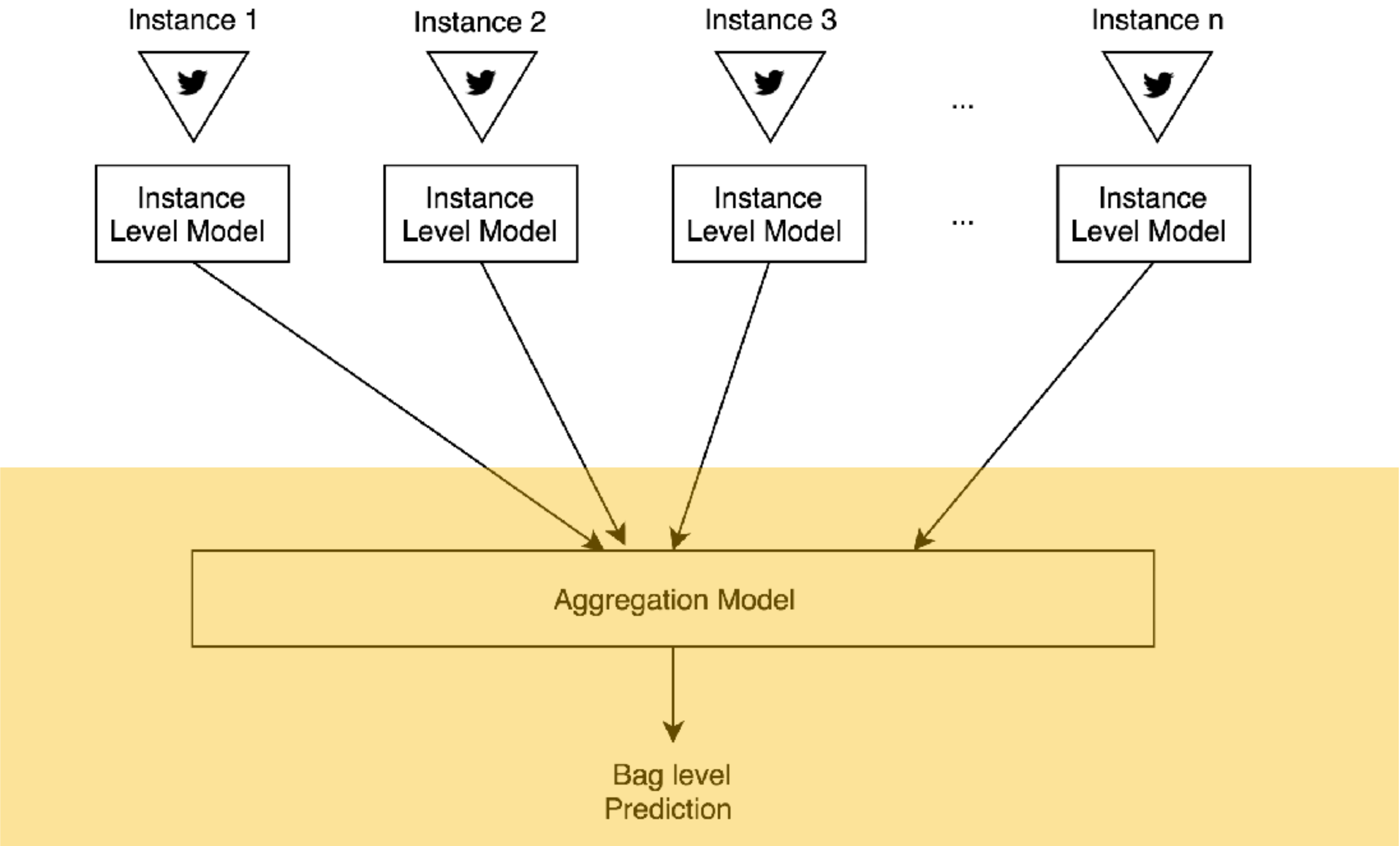


Instance Level Model

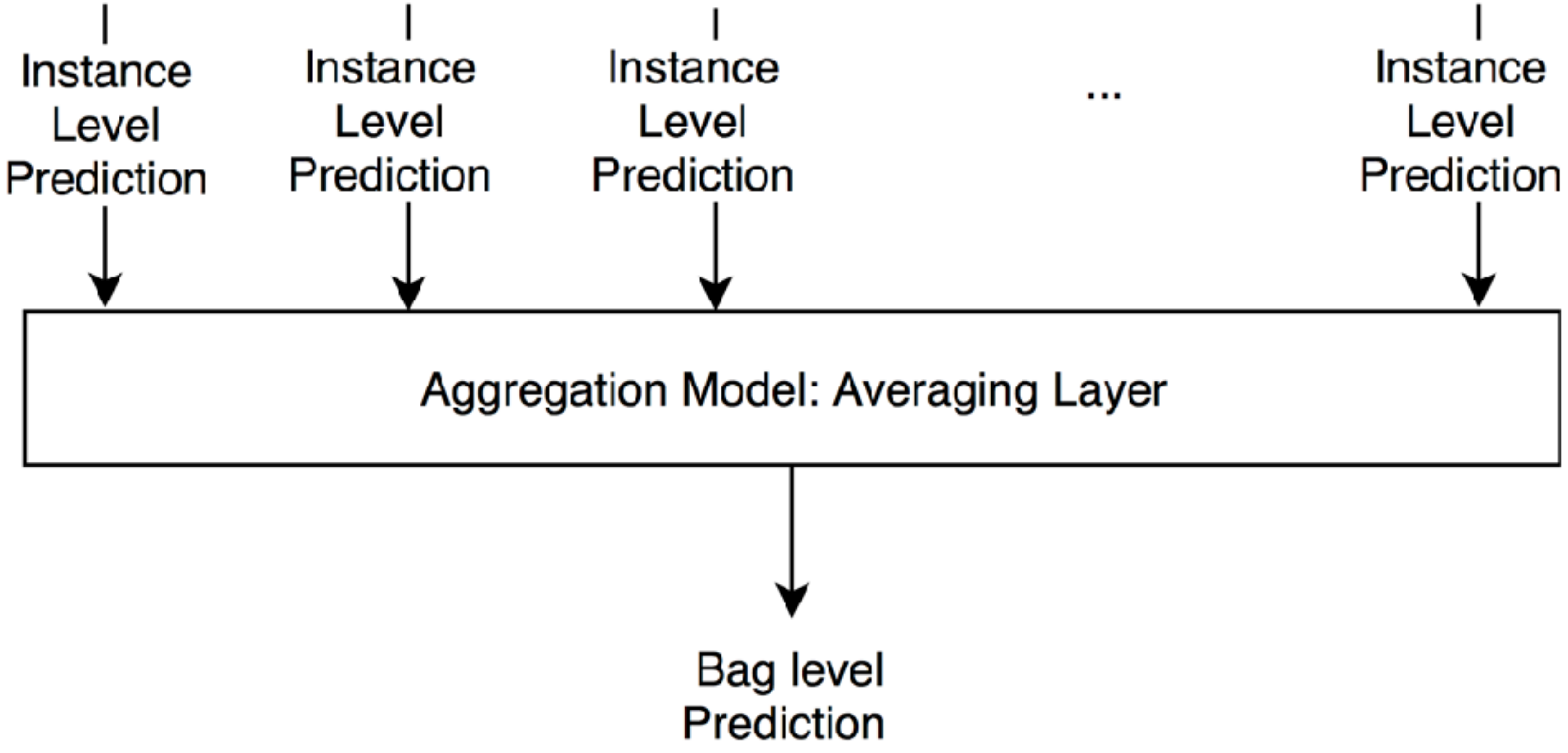


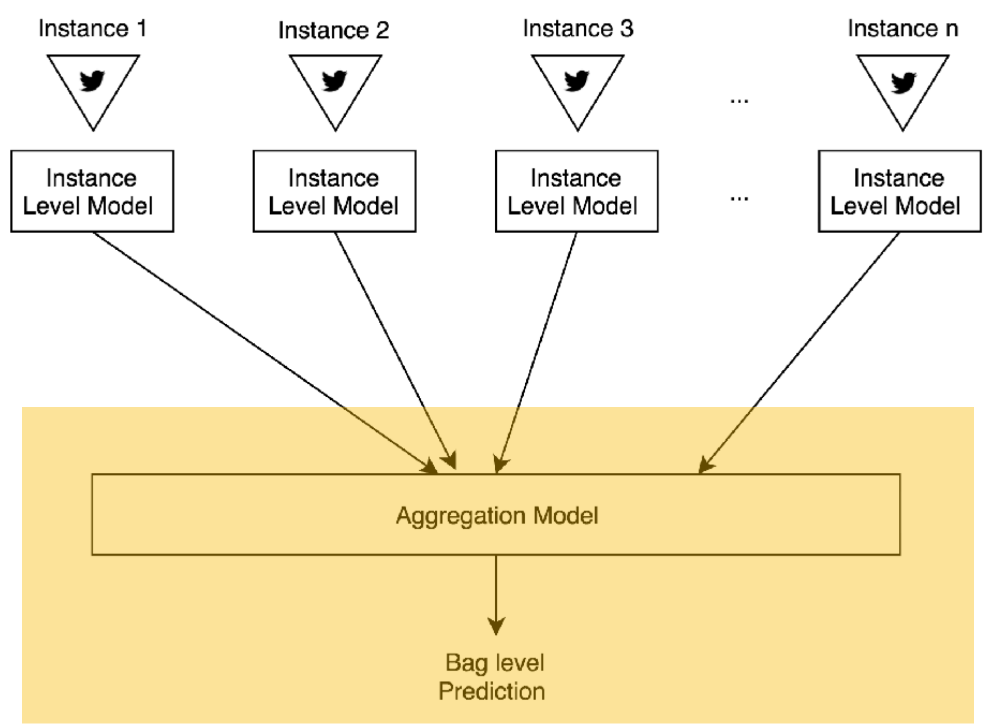


Model Architecture



Aggregation Scheme





04

Training and Losses

Loss Function

$$\mathcal{L}(\hat{y}, y) = -\frac{1}{N}\sum_{i}^{N}[y_i log\hat{y}_i + (1 - y_i)log(1 - \hat{y}_i)]$$

Backpropagation

Adam Optimizer

Results

Chosen parameters

Exploring the Hyper-parameter space

Running Time

Accuracy

F-measure

Case Studies

Chosen Parameters

Preprocessing Choices :	
Vocabulary Size	**5000 words**
Number of words per tweet	**20 words**
Treatment of URLs, @mentions and #hashtags	**url, $hashtag$, $mention$**
Number of tweets per bag	**10**
Embedding Size	**32**
Dense Nodes	**100**
Dropout	**25%**
Batch Size	**256**
Epochs	**200 with early stopping**
Learning Rate	**0.0001**

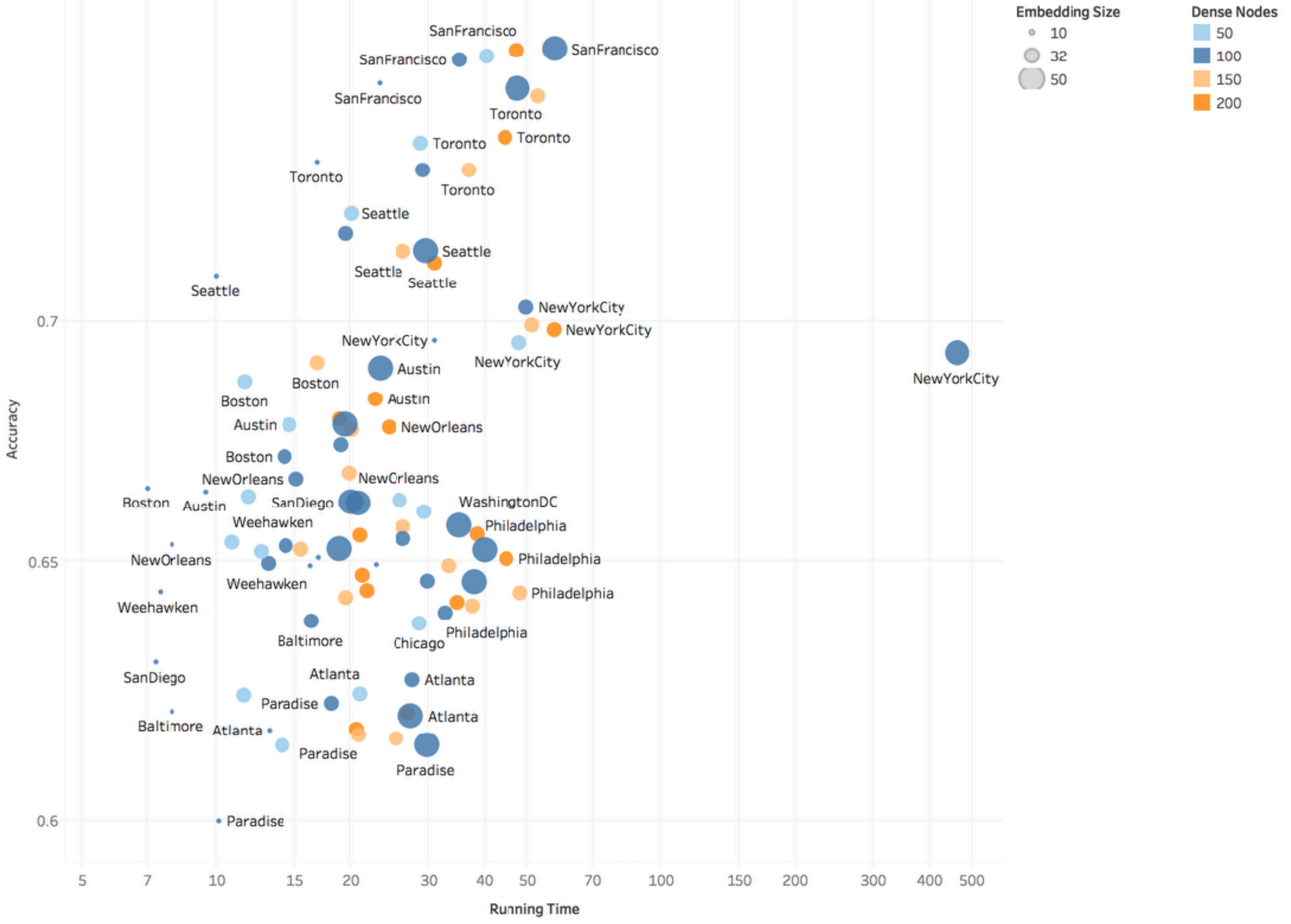
Hyper-parameter space

Preprocessing Choices :	
Vocabulary Size	**5000 words**
Number of words per tweet	**20 words**
Treatment of URLs, @mentions and #hashtags	**url, $hashtag$, $mention$**
Number of tweets per bag	10
Embedding Size	**32**
Dense Nodes	**100**
Dropout	**25%**
Batch Size	**256**
Epochs	**200 with early stopping**
Learning Rate	0.0001

Hyper-parameter space

Preprocessing Choices :	
Vocabulary Size	**5000 words**
Number of words per tweet	**20 words**
Treatment of URLs, @mentions and #hastags	**url, $hashtag$, $mention$**
Number of tweets per bag	**10**
Embedding Size	**10 vs 32 vs 50**
Dense Nodes	**50 vs 100 vs 200**
Dropout	**25%**
Batch Size	**256**
Epochs	**200 with early stopping**
Learning Rate	**0.0001**

Hyper-parameter space



Running Time

City	MISVM[1]	SIL[1]	GICF[2]	milNN[2]
Atlanta	38,799	3,031	106	**38**
Austin	21,304	3,072	81	**34**
Baltimore	27,069	3,030	69	**35**
Boston	19,166	2,724	60	**28**
Chicago	22,935	2,866	288	**45**
New Orleans	33,043	2,826	76	**52**
New York City	10,096	3,448	822	**85**
Paradise	12,891	2,825	149	**40**
Philadelphia	14,453	3,320	212	**52**
San Diego	22,547	3,108	67	**57**
San Francisco	12,562	3,883	222	**64**
Seattle	17,088	3,738	143	**41**
Toronto	17,861	3,847	131	**48**
Washington,D.C.	21,026	2,790	193	**42**
Weehawken	9,460	1,894	63	**29**

1: Argo Research Cluster - Processor: 64 Core AMD Opteron; Memory: 512 GB(50 used); **2**: MacBook Pro (2016)- Processor: 2.9 GHz Intel Core i7 ; Memory: 16 GB 2133 MHz LPDDR3 . All times in **seconds**.

Running Time

4x faster than GICF
70x faster than SIL

City	MISVM[1]	SIL[1]	GICF[2]	milNN[2]
Atlanta	38,799	3,031	106	**38**
Austin	21,304	3,072	81	**34**
Baltimore	27,069	3,030	69	**35**
Boston	19,166	2,724	60	**28**
Chicago	22,935	2,866	288	**45**
New Orleans	33,043	2,826	76	**52**
New York City	10,096	3,448	822	**85**
Paradise	12,891	2,825	149	**40**
Philadelphia	14,453	3,320	212	**52**
San Diego	22,547	3,108	67	**57**
San Francisco	12,562	3,883	222	**64**
Seattle	17,088	3,738	143	**41**
Toronto	17,861	3,847	131	**48**
Washington,D.C.	21,026	2,790	193	**42**
Weehawken	9,460	1,894	63	**29**

1: Argo Research Cluster - Processor: 64 Core AMD Opteron; Memory: 512 GB(50 used); **2**: MacBook Pro (2016)- Processor: 2.9 GHz Intel Core i7 ; Memory: 16 GB 2133 MHz LPDDR3 . All times in **seconds**.

Running Time

4x faster than GICF
70x faster than SIL

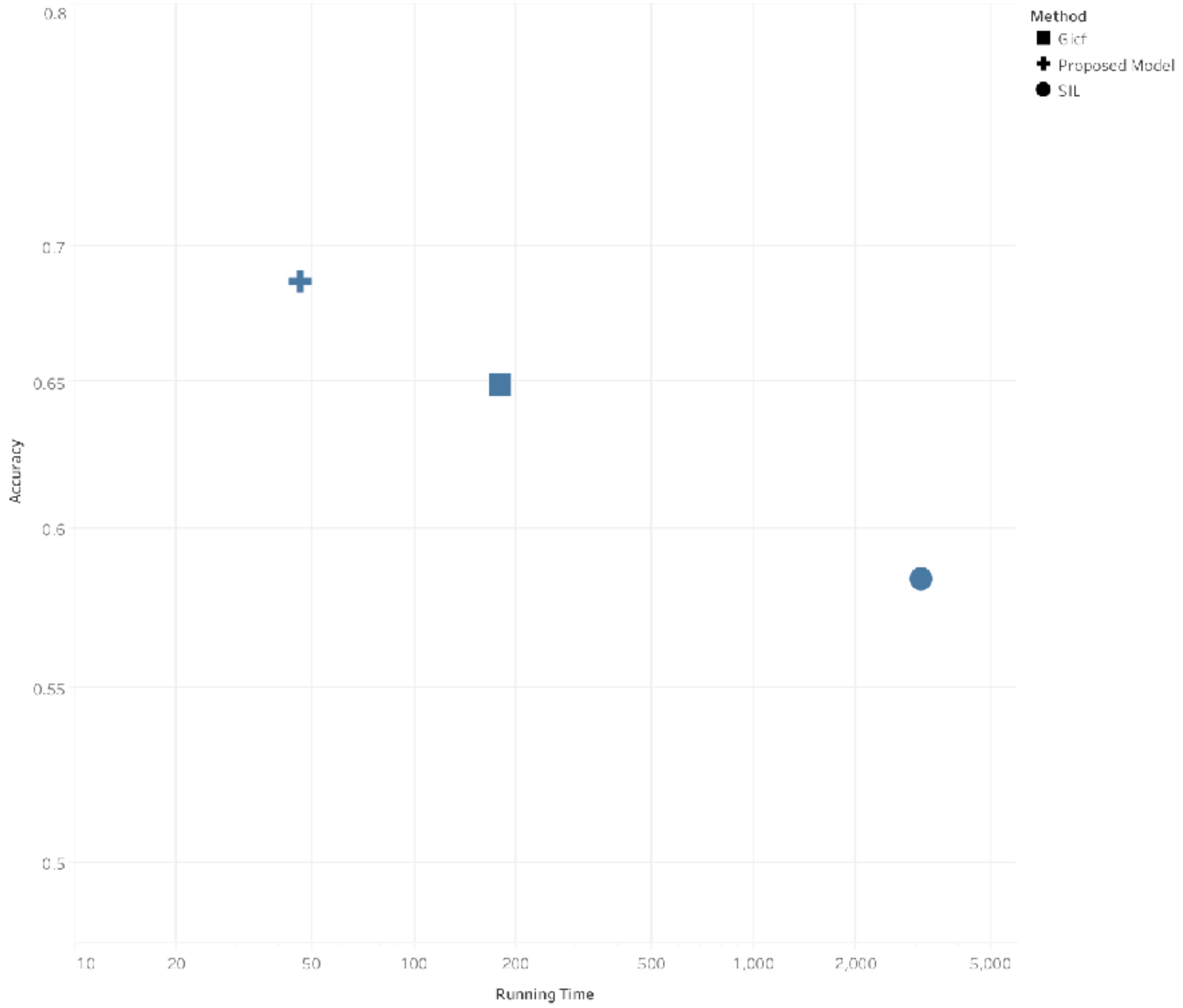
City	MISVM[1]	SIL[1]	GICF[2]	milNN[2]
Atlanta	38,799	3,031	106	**38**
Austin	21,304	3,072	81	**34**
Baltimore	27,069	3,030	69	**35**
Boston	19,166	2,724	60	**28**
Chicago	22,935	2,866	288	**45**
New Orleans	33,043	2,826	76	**52**
New York City	10,096	3,448	822	**85**
Paradise	12,891	2,825	149	**40**
Philadelphia	14,453	3,320	212	**52**
San Diego	22,547	3,108	67	**57**
San Francisco	12,562	3,883	222	**64**
Seattle	17,088	3,738	143	**41**
Toronto	17,861	3,847	131	**48**
Washington,D.C.	21,026	2,790	193	**42**
Weehawken	9,460	1,894	63	**29**

1: Argo Research Cluster - Processor: 64 Core AMD Opteron; Memory: 512 GB(50 used); **2**: MacBook Pro (2016)- Processor: 2.9 GHz Intel Core i7 ; Memory: 16 GB 2133 MHz LPDDR3 . All times in **seconds**.

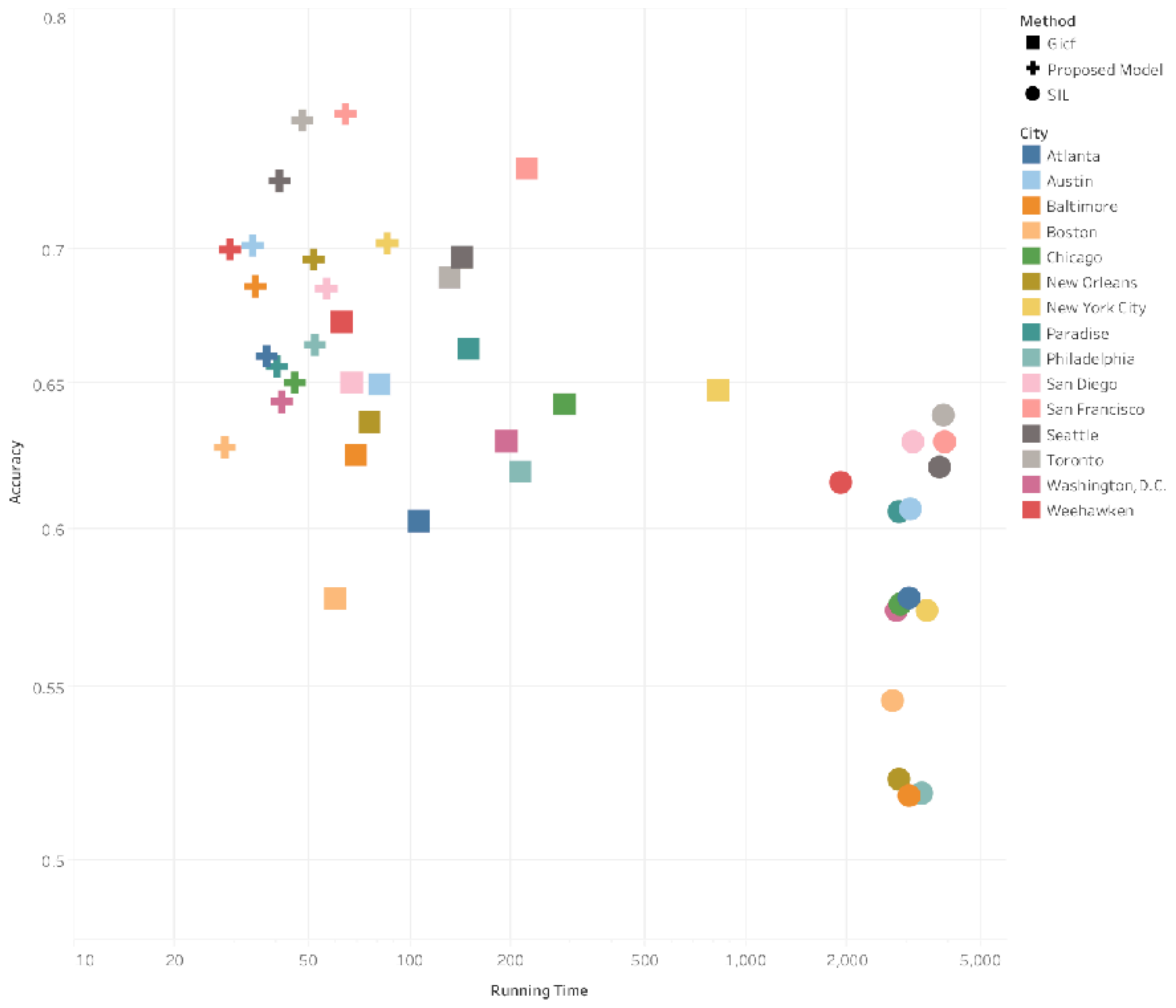
Accuracy

City	MISVM	SIL	GICF	milNN
Atlanta	0.4990	0.5780	0.6025	**0.6602**
Austin	0.4970	0.6070	0.6501	**0.7015**
Baltimore	0.4990	0.5180	0.6248	**0.6858**
Boston	0.5000	0.5460	0.5774	**0.6276**
Chicago	0.5000	0.5760	0.6429	**0.6502**
New Orleans	0.5000	0.5230	0.6365	**0.6962**
New York City	0.5000	0.5740	0.6476	**0.7024**
Paradise	0.4980	0.6060	**0.6629**	0.6565
Philadelphia	0.4960	0.5190	0.6195	**0.6644**
San Diego	0.5000	0.6300	0.6504	**0.6850**
San Francisco	0.5000	0.6300	0.7322	**0.7542**
Seattle	0.5000	0.6210	0.6970	**0.7269**
Toronto	0.4990	0.6390	0.6895	**0.7520**
Washington,D.C.	0.5000	0.5740	0.6298	**0.6437**
Weehawken	0.5000	0.6160	0.6727	**0.7000**

Accuracy vs Running Time



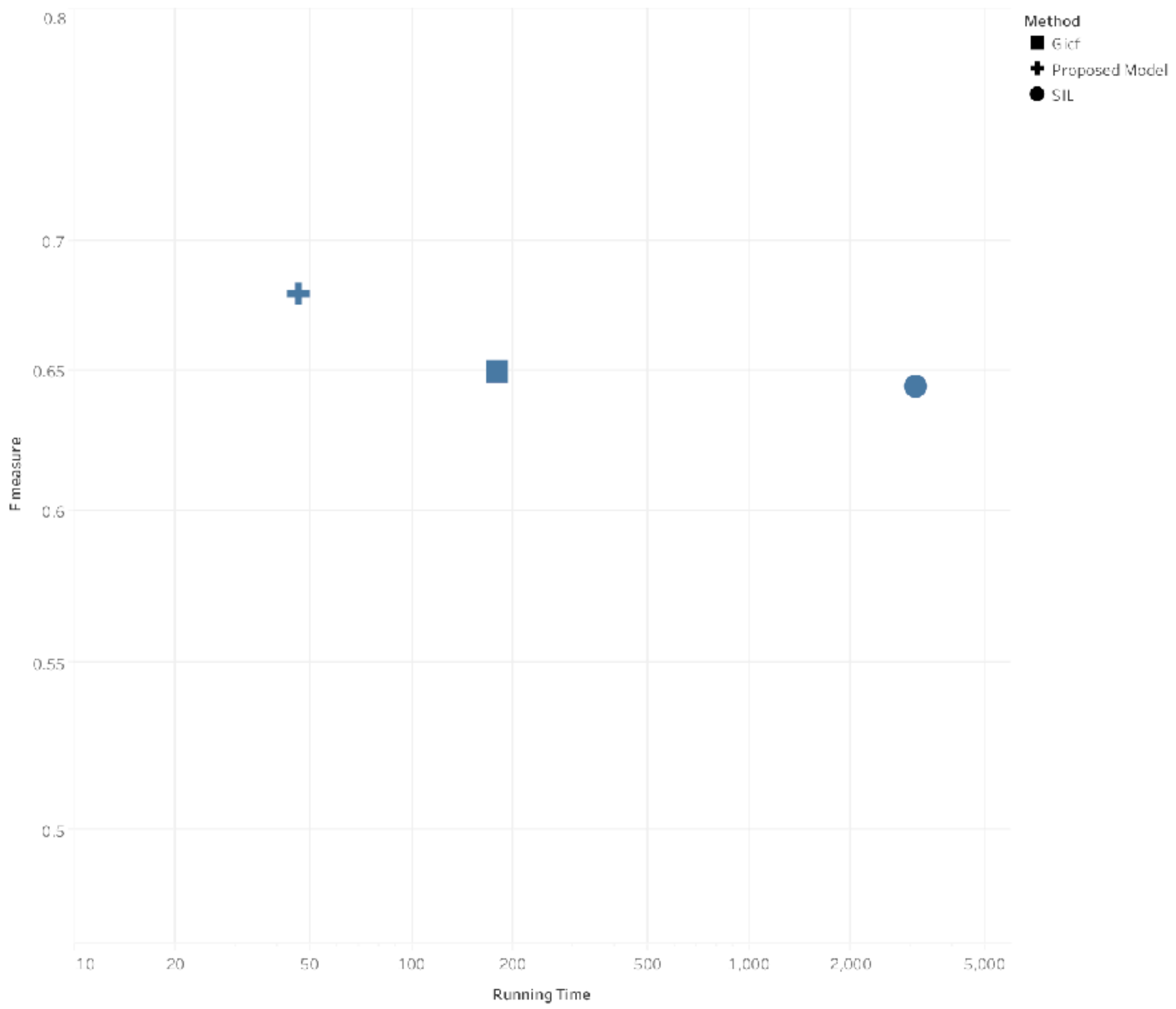
Accuracy vs Running Time by City



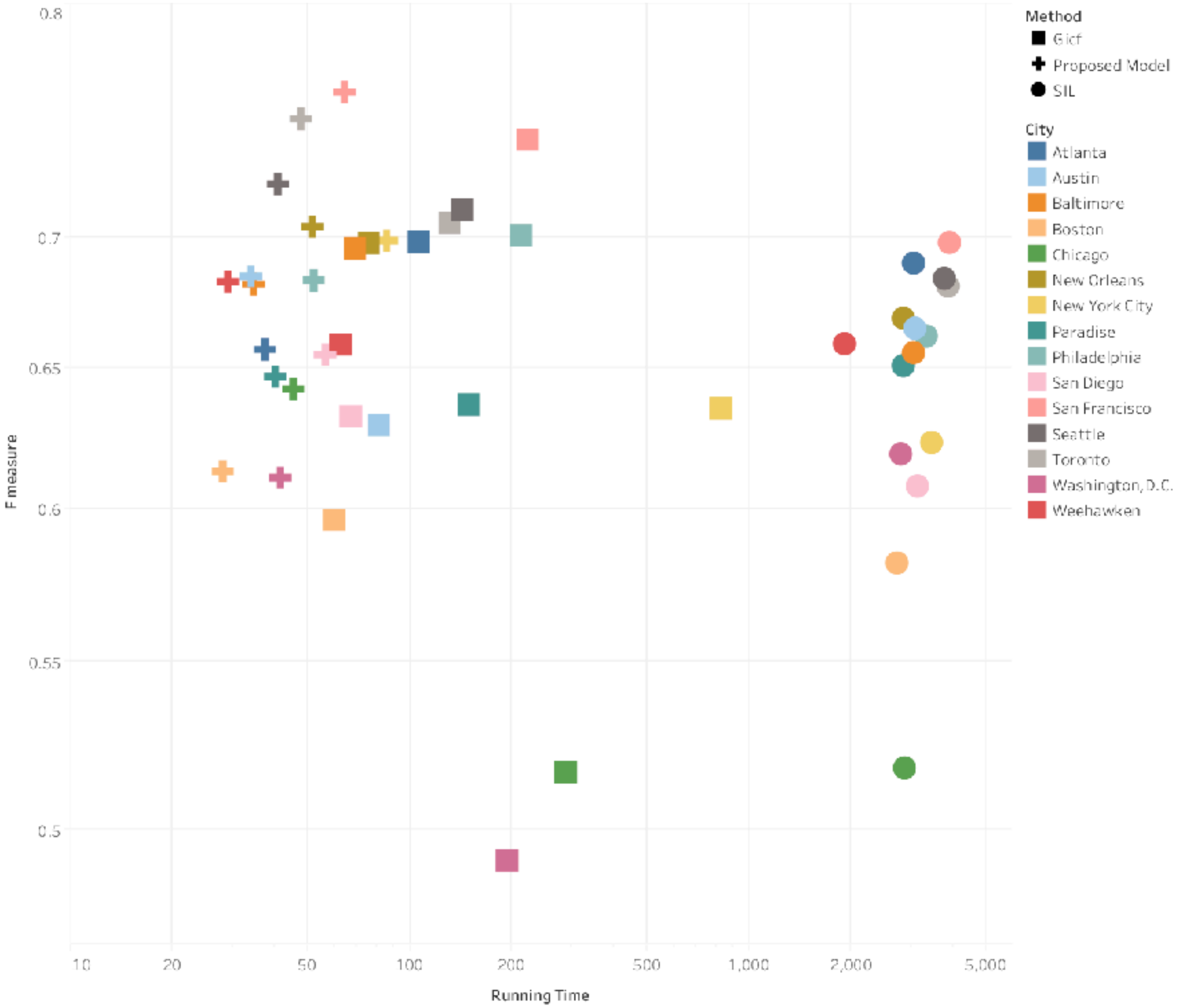
F- measure

City	MISVM	SIL	GICF	milNN
Atlanta	0.0000	0.6900	**0.6982**	0.6568
Austin	0.0000	0.6650	0.6291	**0.6848**
Baltimore	0.0000	0.6560	**0.6957**	0.6816
Boston	0.0000	0.5820	0.5960	**0.6130**
Chicago	0.0000	0.5180	0.5163	**0.6420**
New Orleans	0.0000	0.6690	0.6976	**0.7041**
New York City	0.0000	0.6230	0.6349	**0.6988**
Paradise	0.0000	**0.6510**	0.6365	0.6468
Philadelphia	0.0000	0.6620	**0.7006**	0.6830
San Diego	0.0000	0.6080	0.6325	**0.6548**
San Francisco	0.0000	0.6980	0.7398	**0.7603**
Seattle	0.0000	0.6840	0.7112	**0.7215**
Toronto	0.0000	0.6810	0.7056	**0.7485**
Washington,D.C.	0.0000	**0.6190**	0.4910	0.6108
Weehawken	0.0000	0.6590	0.6588	**0.6827**

F-measure vs Running Time



F-measure vs Running Time by City



Case Studies


NEW
YORK
BROOKLYN
BROADWAY
AV
STREET
CITY
PARK
ST
WILLIAMSBURG
MANHATTAN
AVENUE
BRIDGE
SQUARE
ISLAND
CENTRAL
WEST
AIRPORT
STATEN
STATION
NY
NYC
LEXINGTON
AMERICA
MADISON
JERSEY
ASTORIA
MTA
HUDSON
RIVER
UNION
PLAIN
NEWARK
RD
INTERNATIONAL
ROCKAWAY
PRINCE
LONG
EAST
GARDEN
WHITE
HOUSTON
UTICA
HARLEM
JFK
BLVD
GRAND
METRO
BRONX
PENN
QUEEN
COLUMBUS
LAFAYETTE
NASSAU
LONDON
SUBWAY
SOUTH
WASHINGTON
LAGUARDIA
GREENWICH
AMSTERDAM
ATLANTIC
JAMAICA


SAN
FRANCISCO
INTERNATIONAL
BRIDGE
MISSION
GATE
ST
CA
CASTRO
BLVD
CENTER
VEGAS
DOLORES
AVE
BAY
STATION
AMERICA
BEACH
CITY
MONTGOMERY
CORTLAND
CLARA
CALIFORNIA
LA
SQUARE
NEW
OAKLAND
STREET
AIRPORT
PARK
GOLDEN
SANTA
BART
OCEAN
MARINA
SHOPPING
WESTLAKE
VALLEY
SANFRANCISCO
ALCATRAZ
MALFOY
CAFE
BERKELEY
JOSE
PALO
SACRAMENTO
LONDON
SOUTH
BUENA
BROADWAY
DENVER
SFO
FULTON
ALTO
LAKE
LOUNGE
TAQUERIA
WASHINGTON
MANOR
BRANNAN
HARRISON
PLEASANTON
CHRONICLE

New York City - User 1

1.0 - " I'm at Fashion's Night Out **NYC** (NYC, New York) w/ 158 others http://t.co/r2XDXEp "

0.99496824 - " I'm at **Brooklyn Bridge** w/ 2 others [pic]: http://t.co/ry1BAJZo "

0.43607074 - " Piece of crap airline! Thanks for losing my bag! @united @UnitedAirlines #united #unitedairlines "

0.014939007 - "I'm at Sfuzzi (2533 Mckinney Ave, Routh St, **Dallas**) w/ 7 others http://t.co/BnxbYtSr "

0.0078560486 - " I'm at Public House (400 N State St, at Kinzie, **Chicago**) w/ 7 others http://t.co/q9tiqW2m "

New York City - User 2

0.99559683 - "Beautiful view... good food... great music... romantic husband = perfect evening (@ Le Kaveka Restaurant & Bungalows) http://t.co/sE4EjSGV '

0.99972242 - " A lovely fall brunch with @Accarrino (@ Anthony David's w/ @accarrino) [pic]: http://t.co/4wzMBut6 "

0.99978334 - " I'm at **Lincoln Tunnel** (New York City) w/ 3 others http://t.co/EY5B7SB5"

0.49836314 - "I just became the mayor of Hilton Moorea: Toatea Crepes & Bar on @foursquare! http://t.co/jopZTuuE"

0.2133007 - "Beautiful breakfast overlooking the **ocean** (@ Hilton Moorea: Arii Vahine Restaurant) [pic]: http://t.co/277WG6QF"

0.11722157 - "Standing in line to return what we bought last night. Efficiency! **Walmart** is out of cash. Waiting 15 min for refund. http://t.co/VIXay1dv"

Not New York City - User 3 (Arlington,NY)

0.99989402 - "Back in New Yawk Citay (@ Grand Central Terminal w/ 28 others) http://t.co/2CBNSMtJ"

0.99894804 - "Back to Vassar on the 2:45 Metro North... Snow fall was pretty while it lasted (@ Grand Central Terminal) [pic]: http://t.co/MRE8JqMk"

0.50644404 - "@paradisetaylor and I on date night ï\x98\x8a (@ Regal Columbiana Grande Stadium 14 for The Twilight Saga: Breaking Dawn -...)http://t.co/54ywGeq9"

0.018137755 - "I just ousted @aashim_91 as the mayor of College Center - Vassar College on @foursquare! http://t.co/qSljibJA"

0.3946189 - "I just became the mayor of Matthew's bean on @foursquare! http://t.co/jLE2je5g "

San Francisco - User 1

0.9999975- " I'm at Alcatraz (Alcatraz Island, **San Francisco** Bay, **San Francisco**) w/ 6 others http://t.co/47YWmX9p "

0.9999856 - " I'm at Chinatown Gate (500 Bush St, at Grant Ave, **San Francisco**) http://t.co/rb49RnFa "

0.5055542 - " @matthewharkin @phillo haha, now I'm worried ")

0.025480814 - " I'm at Tiffany & Co. (210 N Rodeo Dr., **Beverly Hills**) http://t.co/YppqS7ix ")

San Francisco - User 2

0.99910492 - "Can't wait for @BankSimple, @usbank is such a joke from a **technology** / ease-of-use perspective."

0.54251802 - "My whole morning **has** been devoted to banking. Not done yet. Living the life."

0.3358801 - "My whole morning **had** been devoted to banking. Not done yet. Living the life."

San Francisco - User 3

0.97942388 - "Engineers love free food! #IDF2011 http://t.co/rThzEeKO http://t.co/oyYqSmYo"

0.99747145 - "@hashimwaheed the left one is USB serial into **Mac**, the other is normal **iPhone** USB into **Mac**"

0.9944582 - "**iOS**5 beta expires today! "limited-edition b7b" redsn0w lets you sync data+ pics: OSX http://t.co/EbVEGO0t Win http://t.co/vC5PK2Eg"

0.0067595979 - "@alexeheath my host at **Apple** surprised me with that visitor's name tag...I had expected it to be my real name :)"

Not San Francisco - User 4 (Arabi,Louisiana)

0.16304019 - "Follow the OG triple OG @thad4mayor to ensure that he don't steal ur wallet when he see you in the streets...<>jtfo"

0.23261635 - "@jbdachamp u show me no luv :("

0.011714808 - "somethins gotta give"

0.10918618- "@Cree_Oh_Lay_CO how ya been?"

0.00020607341 -"@jbdachamp and u won't lol",

0.0094076423 - "@jbdachamp I was MIA 4 a min due 2 **technical** issues but now I'm baaaaack lol"

0.010172283 - "da best part is that the downs dont last always"

0.20761815 - "I luv fridays :)"

Conclusion

Conclusion

Fast and Scalable

Easily trained

General

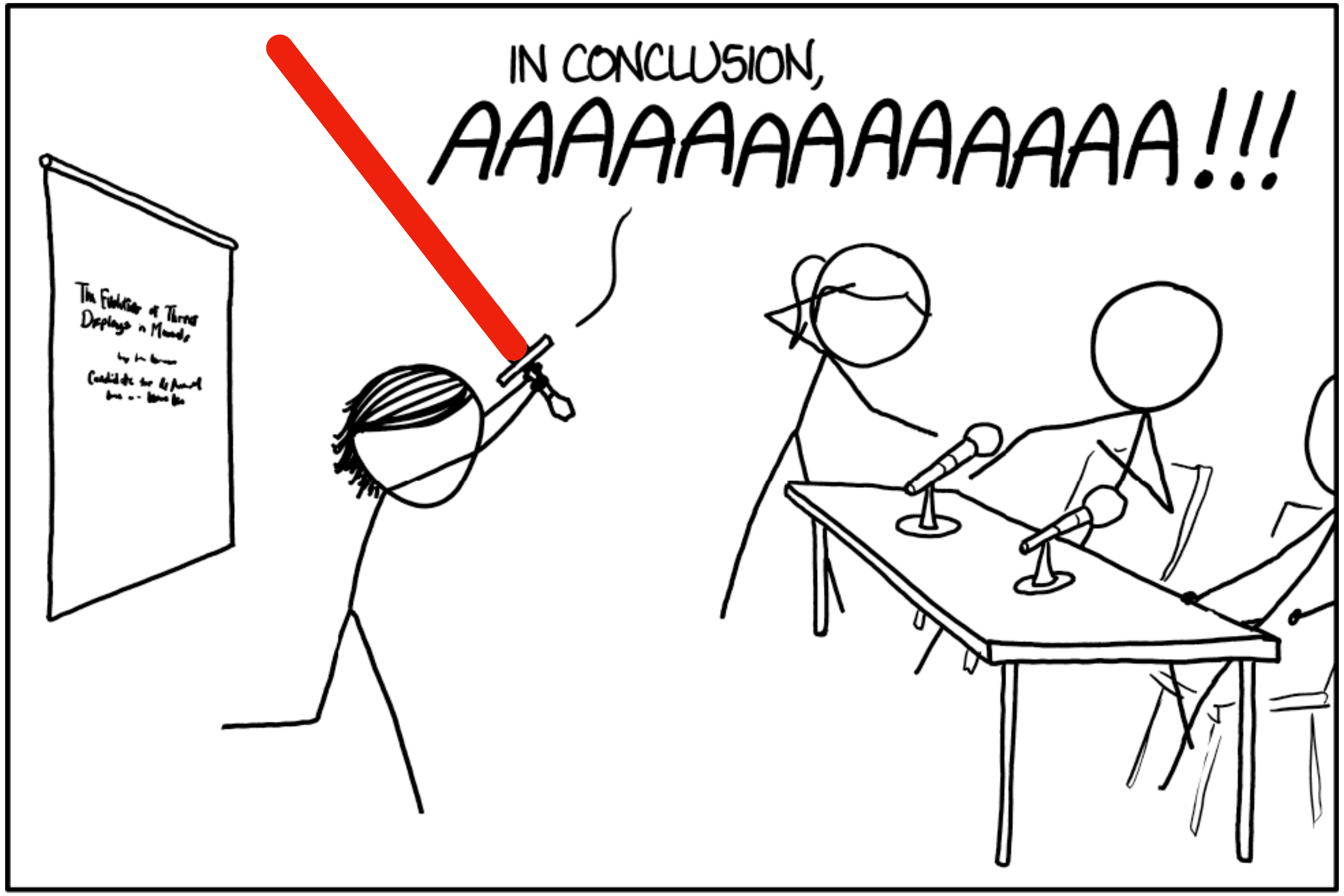
Relaxed assumptions

No more feature engineering/ kernels

Capable of high level structure detection

Not a bag of words model anymore

Conclusion



THE BEST THESIS DEFENSE IS A GOOD THESIS OFFENSE.

Future Directions

TRANSFER LEARNING
Use word vectors that come out of twitter based word vector research

MULTIPLE CLASSIFICATION
Consider all cities at once

INSTANCE LEVEL MODELS
CNNs or RNNs to incorporate higher complexity and variable input size

AGGREGATION FUNCTIONS
eg. Incorporate structure within bags

OTHER APPLICATIONS
Any MIL setup can benefit from this general formulation eg. Bio-engineering, Vision, Sentiment

Questions?

References

[1] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Perez, “Solving the multiple instance problem with axis-parallel rectangles,” *Artificial Intelligence*, vol. 89, no. 1-2, pp. 31–71, 1997. [Online]. Available: http://lis.csail.mit.edu/pubs/tlp/multiple-instance- aij.pdf

[2] S. Andrews, T. Hofmann, and I. Tsochantaridis, “Multiple instance learn- ing with generalized support vector machines,” in *Eighteenth National Conference on Artificial Intelligence*. Menlo Park, CA, USA: American Association for Artificial Intelligence, 2002, pp. 943–944. [Online]. Available: http://dl.acm.org/citation.cfm?id=777092.777234

[3] T. Grtner, P. A. Flach, A. Kowalczyk, and A. J. Smola, “Multi-instance kernels,” in In Proc. *19th International Conf. on Machine Learning.* Morgan Kaufmann, 2002, pp. 179–186

[4] N. Weidmann, E. Frank, and B. Pfahringer, “A two-level learning method for generalized multi-instance problems,” in In Proceedings of the *Fourteenth European Conference on Machine Learning*, 2003.

[5] S. Ray and M. Craven, “Supervised versus multiple instance learning: An empirical comparison,” in Proceedings of the *22nd International Conference on Machine Learning*, ser. ICML ’05. New York, NY, USA: ACM, 2005, pp. 697–704. [Online]. Available: http://doi.acm.org/10.1145/1102351.1102439

[6] Y.-F. Li, J. T. Kwok, I. W. Tsang, and Z.-H. Zhou, A Convex Method for Locating Regions of Interest with Multi-instance Learning. Berlin, Heidelberg: *Springer Berlin Heidelberg*, 2009, pp. 15–30.

[7] G. Liu, J. Wu, and Z.-H. Zhou, “Key instance detection in multi-instance learning,” in Proceedings of the *Asian Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, S. C. H. Hoi and W. Buntine, Eds., vol. 25. Singapore Management University, Singapore: PMLR, 04–06 Nov 2012, pp. 253–268. [Online]. Available: http://proceedings.mlr.press/v25/liu12b.html

[8] D. Kotzias, M. Denil, N. de Freitas, and P. Smyth, “From group to individual labels using deep features,” in Proceedings of the *21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’15. New York, NY, USA: ACM, 2015, pp. 597–606. [Online]. Available: http://doi.acm.org/10.1145/2783258.2783380

[9] A. Woodruff and C. Plaunt, “Gipsy: Geo-referenced information processing system,” vol. 45, 05 1994.

[10] E. Amitay, N. Har’El, R. Sivan, and A. Soffer, “Web-a-where: Geotagging web content,” in Proceedings of the *27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR ’04. New York, NY, USA: ACM, 2004, pp. 273–280. [Online]. Available: http://doi.acm.org/10.1145/1008992.1009040

[11] Z. Cheng, J. Caverlee, and K. Lee, “You are where you tweet: a content-based approach to geo-locating twitter users,” in *CIKM*, 2010.

[12] J. Liu and D. Inkpen, “Estimating user location in social media with stacked denoising auto-encoders,” in *VS@HLT-NAACL*, 2015

[13] A. Rahimi, T. Cohn, and T. Baldwin, “A neural model for user geolocation and lexical dialectology,” CoRR, vol. abs/1704.04008, 2017. [Online]. Available: http://arxiv.org/abs/1704.04008

[14] LeCun, Y., Bengio, Y. and Hinton, G. Deep learning. *Nature* 521, 7553 (2015), 436-444.[

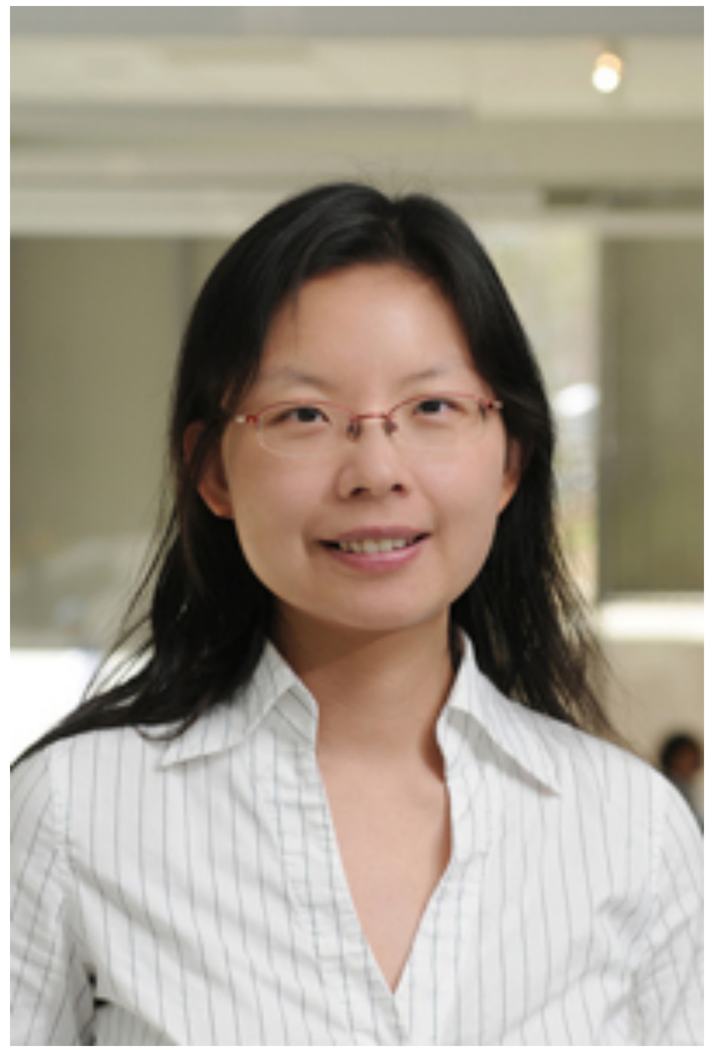
[15] A. Kolmogorov, "On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables", Proceedings of the *USSR Academy of Sciences*, vol. 108, pp. pp. 179–182, 1956.

[16] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[17] "xkcd: Thesis Defense", *Xkcd.com*. [Online]. Available: https://xkcd.com/1403/. [Accessed: 15- Jan- 2018].

Committee Members

Dr. Huzefa Rangwala



Dr. Jessica Lin

Dr. Aditya Johri

Fin.