



Augmenting Intelligence



Recomendaciones Laborales
GCBA - TyP

DESCRIPCIÓN DEL PROBLEMA

Se dispone de información acerca de los ciudadanos interesados en oportunidades laborales, e información de oportunidades laborales las cuales son publicadas por los empleadores que las ofrecen. En base a esto, GCBA está diseñando un algoritmo que permita recomendar las oportunidades laborales más acordes a los ciudadanos interesados.

ESQUEMA GENERAL DE TRABAJO

- 1) Identificación y análisis de entidades y atributos con potencial informativo para solución del problema
- 2) Análisis cuantitativo de datos con potencial informativo para la solución del problema
- 3) Comparación sintáctica entre valores de atributos con potencial para conectar vecinos y oportunidades laborales que puedan ser de su interés.
- 4) Propuesta de diseño de algoritmo de recomendación

1) Identificación y análisis de entidades y atributos con potencial informativo para solución del problema

Entidades	Curriculum Vitae					
		Experiencia Laboral	Formacion Academica		Conocimientos Aptitudes	
				Tipo de Formacion	Nivel de conocimiento/ Aptitudes	Tipo de conocimientos / Aptitudes
atributos	nivel educativo	puesto	descripcion	descripcion	descripcion	descripcion

Entidades	Registro Laboral Formal
	Cargos
atributos	descripcion

1) Identificación y análisis de entidades y atributos con potencial informativo para solución del problema

Entidades	Capacitacion				
					Aptitudes
atributos	descrip_normalizada	descripcion	categoria front	categoria back	descripcion

Entidades	Oportunidad Laboral										
										Oportunidad Laboral - Idioma Requerido	Oportunidad Laboral - Areas de Interes
atributos	descripcion	tipo de puesto	grado de estudio	edad minima	edad maxima	turno de trabajo	estado	genero requerido	experiencia requerida	idioma requerido	areas de interes

1) Identificación y análisis de entidades y atributos con potencial informativo para solución del problema

Entidades	Postulaciones			
atributos	vecino_id	oportunidad_ laboral_id	consiguio trabajo?	estado de la entrevista

2) Análisis cuantitativo de datos con potencial informativo para la solución del problema

Entidades relevantes y sus relaciones	Valor Absoluto	Valor Porcentual
#vecinos	1,120,027	100%
#vecinos con algún registro laboral formal	356,096	31.79%
#vecinos con alguna capacitación	438,662	39.17%
#descripciones de oportunidades laborales (OLs) disponibles	4,578	100%
#OLs con estado abierto	557	12.17%
#descripciones distintas de OLs con tipo de puesto null	1,424	31%
#descripciones distintas de OLs con tipo de puesto distinto de null	3,154	69%
#descripciones de capacitaciones distintas	1,812	100%
#descripciones distintas de cargos en registro laboral formal	370	100%
#descripciones distintas de puestos en experiencia laboral	55,327	100%

2) Análisis cuantitativo de datos con potencial informativo para la solución del problema

Entidades relevantes y sus relaciones	Valor Absoluto	Valor Porcentual	
#vecinos distintos con algún cv	84,392	7.53%	100%
#vecinos distintos con algún cv y que hicieron alguna capacitación	19	0.00%	0.02%
#vecinos distintos con algún cv y que no hicieron ninguna capacitación	84,373	7.53%	99.98%
#vecinos distintos con algún cv que tuvieron algún trabajo formal pero no hicieron ninguna capacitación	40,864	3.65%	48.42%
#vecinos distintos con algún cv que tuvieron algún trabajo formal pero no hicieron ninguna capacitación	43,509	3.88%	51.56%
#vecinos con algún cv que tuvieron al menos un trabajo formal e hicieron alguna capacitación	5	0.00%	0.01%
#vecinos con algún cv que no tuvieron ningún trabajo formal e hicieron alguna capacitación	14	0.00%	0.02%

2) Análisis cuantitativo de datos con potencial informativo para la solución del problema

Conclusiones preliminares

- 1) La cantidad de vecinos con algún CV es una fracción pequeña en relación al total de vecinos, pero igualmente considerable en términos absolutos.
- 2) La cantidad de vecinos con algún CV que hicieron capacitaciones es muy pequeña, por lo cual esta información no permitirá conectar directamente los vecinos con CV con las oportunidades laborales.
- 3) La cantidad de vecinos con CV que tuvieron algún trabajo es muy similar a la cantidad de vecinos con CV que no tuvieron ningún trabajo formal.

3) Comparación sintáctica entre valores de atributos con potencial para conectar vecinos con oportunidades laborales que puedan ser de su interés.

Resultados Capacitación vs Oportunidad Laboral	Valor Absoluto	Valor Porcentual
Cantidad total de pares de descripciones de oportunidades laborales y descripciones de capacitaciones que pueden formarse	8,295,336	100%
Cantidad de pares de descripciones de oportunidades laborales y descripciones de capacitaciones con similitud mayor a 0.9	2,526	0.03%
Cantidad de descripciones de oportunidad laboral distintas	4,578	100%
Cantidad de descripciones de oportunidades laborales distintas que tiene similitud mayor a 0.9 con alguna descripción de capacitación	1,339	29.25%
Cantidad de descripciones de capacitación distintas	1,812	100%
Cantidad de descripciones de capacitación distintas que tienen similitud mayor a 0.9 con alguna descripción de oportunidad laboral	217	11.98%

3) Comparación sintáctica entre valores de atributos con potencial para conectar vecinos con oportunidades laborales que puedan ser de su interés.

Resultados Registro Laboral Formal vs Oportunidad Laboral	Valor Absoluto	Valor Porcentual
Cantidad total de pares de descripciones de oportunidades laborales y descripciones de registro laboral formal que pueden formarse	1,693,860	100%
Cantidad de pares de descripciones de oportunidades laborales y descripciones de registro laboral formal con similitud mayor a 0.9	178	0.01%
Cantidad de descripciones de oportunidad laboral distintas	4,578	100%
Cantidad de descripciones de oportunidades laborales distintas que tiene similitud mayor a 0.9 con alguna descripción de cargo en registro laboral formal	139	3.04%
Cantidad de descripciones de cargo distintas en registro laboral formal	370	100%
Cantidad de descripciones de cargo distintas en registro laboral formal que tienen similitud mayor a 0.9 con alguna descripción de oportunidad laboral	31	8.38%

3) Comparación sintáctica entre valores de atributos con potencial para conectar vecinos con oportunidades laborales que puedan ser de su interés.

Resultados Experiencia Laboral vs. Oportunidad Laboral	Valor Absoluto	Valor Porcentual
Cantidad total de pares de descripciones de oportunidades laborales y descripciones de puesto en experiencia laboral	253,287,006	100%
Cantidad de pares de descripciones de oportunidades laborales y descripciones de puesto en experiencia laboral con similitud mayor a 0.9	193,582	0.08%
Cantidad de descripciones de oportunidad laboral distintas	4,578	100%
Cantidad de descripciones de oportunidades laborales distintas que tiene similitud mayor a 0.9 con alguna descripción de puesto en experiencia laboral	3,803	83.07%
Cantidad de descripciones de puesto distintas en experiencia laboral	55,327	100%
Cantidad de descripciones de puesto distintas en experiencia laboral que tienen similitud mayor a 0.9 con alguna descripción de oportunidad laboral	17,767	32.11%

3) Comparación sintáctica entre valores de atributos con potencial para conectar vecinos con oportunidades laborales que puedan ser de su interés.

Conclusiones preliminares

- 1) Existen conexiones sintácticas prometedores entre las descripciones de las oportunidades laborales y las descripciones de la experiencia laboral de los vecinos. Esta información tiene mucho potencial para conectar en forma directa vecinos con oportunidades laborales.
- 2) Por otro lado, si bien las conexiones sintácticas entre la descripción de capacitaciones y la descripción de las oportunidades laborales no son tan altas como el caso anterior, igualmente puede aportarnos valor informativo. En este caso, debido a que no hay muchos vecinos con CV que hayan hecho capacitaciones, esta información sólo podría usarse como features de las oportunidades laborales, de manera que sean útiles en estrategias de filtrado basado en contenido.
- 3) Las conexiones sintácticas entre la descripción de registro laboral formal y la descripción de las oportunidades laborales no son muy prometedoras.

4) Propuesta de diseño de algoritmo de recomendación

- Del análisis preliminar realizado, podemos observar que existen relaciones sintácticas entre atributos con potencial para conectar ciudadanos con oportunidades laborales de interés.
- El hecho que existan estas conexiones sintácticas es buen indicio, porque es la manera más simple y directa de vincular los datos y constituye un primer paso para continuar el análisis con otros atributos. A su vez, utilizando técnicas más complejas se puede indagar sobre los vínculos semánticos que pudieran existir.
- Para abordar el problema, proponemos que se utilice una combinación de los enfoques clásicos de recomendación que utilizan filtrado colaborativo y basado en contenido, de manera que pueda mitigarse la dificultad de los escenarios donde se dispone de pocas interacciones por parte de los usuarios.

4) Propuesta de diseño de algoritmo de recomendación

- Para este caso, podemos interpretar una interacción como la postulación de un vecino a una oportunidad laboral. De esta forma, podemos construir una matriz de interacciones donde las filas representan los vecinos, las columnas representan las oportunidades laborales y las celdas representan las postulaciones de los vecinos a las oportunidades laborales.

	Oportunidad Laboral 1	...	Oportunidad Laboral n
vecino 1	1	...	1
...
vecino m	1	...	0

4) Propuesta de diseño de algoritmo de recomendación

- Asimismo, cada vecino puede tener atributos (*features*) que permiten describir y determinar su perfil (edad, trabajos previos, cursos, estudios, etc), lo cual es información que el algoritmo puede utilizar para recomendar las oportunidades laborales a los vecinos “parecidos”.
- Similarmente, cada oportunidad laboral puede tener atributos (*features*) que permiten describirla (descripción, puesto, requisitos, etc), lo cual es información que el algoritmo puede utilizar para encontrar las oportunidades laborales que se parecen a otras, las cuales se consideran como buenas recomendaciones.
- Estas features pueden incorporarse como parte de la matriz de interacciones.

		Oportunidad Laboral 1	...	Oportunidad Laboral n
		OL feature 1, ...,OL feature k	...	OL feature 1, ...,OL feature k
vecino 1	vecino feature 1, ..., vecino feature j	1	...	1
...
vecino m	vecino feature 1, ..., vecino feature j	1	...	0

4) Propuesta de diseño de algoritmo de recomendación

- Una forma de representar las features de los vecinos y las oportunidades laborales es utilizar un vector numérico de los valores de sus atributos relevantes obtenido a partir de un embedding de palabras pre-entrenado (como los provistos por [OpenAI](#)).
- Teniendo en cuenta esto, puede realizarse una interpretación más amplia de qué significa una interacción en este caso. Esto es, podemos calcular la distancia entre el vector numérico que representa un vecino y el vector numérico que representa una oportunidad laboral. Por ejemplo, si la distancia entre el vector m y el vector $m + n$ es 0.82, entonces éste es el valor de la interacción entre vecino m y oportunidad laboral n .

		Oportunidad Laboral 1	...	Oportunidad Laboral n
		vector numerico m +1	...	vector numerico m+n
vecino 1	vector numerico 1	1	...	1
...
vecino m	vector numerico m	1	...	0.82

4) Propuesta de diseño de algoritmo de recomendación

- Para llevar adelante la implementación del algoritmo, sugerimos la utilización de librerías diseñadas específicamente para la construcción de sistemas de recomendación, como es el caso de [lightfm](https://arxiv.org/abs/1507.08439) (ver <https://arxiv.org/abs/1507.08439>).
- De acuerdo a cómo se construya la matriz de interacciones, pueden llegar a obtenerse mejores o peores resultados. Por ejemplo, una cuestión importante que debe tenerse en cuenta es si se va a considerar toda la historia de interacciones, solo una ventana de X meses, o solo el último mes.
- También debe tenerse en cuenta que esta implementación no puede hacer recomendaciones a usuarios nuevos que aún no hayan sido incorporados al modelo, ni puede recomendar oportunidades laborales nuevas que aún no hayan sido incorporadas al modelo. Esto implica que el modelo debe ser re-entrenado periódicamente.